Boletín de la

SOCIEDAD MATEMÁTICA MEXICANA

Tercera Serie

Volumen 19
Número 2
Octubre de 2013

Contenido

ARTÍCULOS DE INVESTIGACIÓN

Continúa/Continued on back cover

PREFACE

The International Year of Statistics 2013 was an initiative set forth by the International Statistical Institute (ISI) in order to highlight the importance of statistics worldwide and its impact in modern society. The choice of year could be justified by the 300th year anniversary of the publication of *Ars Conjectandi* by Jacobi Bernoulli or the 250th year anniversary of the publication of *An essay towards solving a problem in the doctrine of chances* by Thomas Bayes, two works that have strongly influenced the current directions of statistics and probability and the broad spectrum of fields that interact with them.

However, it takes little effort to find more anniversaries of other very important contributions, e.g. 90 years of the publication by Eggenberger y Polya, 80 years of the axiomatisation of Probability by Andréi Kolmogorov, 40 years of the work by Robert C. Merton presenting the Black and Scholes formula as well as the work by Thomas S. Ferguson introducing the cornerstone model in Bayesian nonparametrics. This makes us think that, no matter the year we pick, we could always find anniversaries of important contributions in statistics and probability, which makes evident their role in almost any field in science.

In Mexico, more than 70 universities, research centres, societies and institutions formed part of the over 2300 entities participating in this celebration. In particular, the Sociedad Matemática Mexicana devotes this special volume to divulge some recent research lines followed by renowned Mexican researchers.

Therefore, as invited editors, we have undertaken the difficult task of deciding whom to invite to form part of this volume, in order to have a representative sample of the current research directions in probability and statistics in Mexico. Some of our selection criteria was based on youth, expertise in the field and national representativeness. It goes without saying that our list of contributors is by no means exhaustive and only constitutes a small sample of our rich and high quality community. Although, nowadays it is difficult to establish the boundaries between statisticians and probabilitists we tried to distribute the contributions in half. It is also worth mentioning that all articles were subject of a strict refereeing by experienced experts in each of the treated fields.

We are fully grateful to the authors of this volume, for their valuable contribution and efforts placed in this volume. We are sure that these articles will serve as inspiration to present and future generations, placing the role of statistics and probability as one of the engines driving our modern society. We are indebted to all referees for their careful work put in revising all articles in the volume. Lastly, on behalf of the Mexican community in statistics and probability, would like to acknowledge Prof. Víctor M. Pérez Abreu C., who apart of being member of the steering committee of the international initiative statistics2013, was the main instigator of this celebration in Mexico.

Daniel Hernández, CIMAT

Ramsés H. Mena, IIMAS-UNAM

Bol. Soc. Mat. Mexicana (3) Vol. 19, 2013

ON THE DISTRIBUTION OF EXPLOSION TIME OF STOCHASTIC DIFFERENTIAL EQUATIONS

JORGE A. LEÓN, LILIANA PERALTA HERNÁNDEZ, AND JOSÉ VILLA-MORALES

ABSTRACT. In this paper we use the Itô's formula and comparison theorems to study the blow-up in finite time of stochastic differential equations driven by a Brownian motion. In particular, we obtain an extension of Osgood criterion, which can be applied to some nonautonomous stochastic differential equations with additive Wiener integral noise. In most cases we are able to provide with a method to figure out the distribution of the explosion time of the involved equation.

1. Introduction

Consider the stochastic differential equation

$$\begin{aligned} dX_t &= b(X_t)dt + \sigma(X_t)dW_t, \quad t > 0, \\ X_0 &= x_0. \end{aligned} \tag{1.1}$$

Here $b, \sigma : \mathbb{R} \to \mathbb{R}$ are two locally Lipschitz functions, $x_0 \in \mathbb{R}$ and $\{W_t : t \geq 0\}$ is a Brownian motion defined on a complete probability space $(\Omega, \mathcal{F}, P)$.

It is well-known that the solution X of equation (1.1) may explode in finite time. That is, $|X_t|$ goes to infinite as t approaches to a stopping time that could be finite with positive probability, which is called the explosion time of equation (1.1) (see McKean [12]). The Feller test is an important tool of the stochastic calculus to know if there is blow-up in finite time for (1.1) (see, for example, Karatzas and Shreve [10]). The reader can consult de Pablo et al. [5] (and references therein) for applications of blow-up.

In the case that b is non-decreasing and positive, and $\sigma \equiv 1$, Feller test is equivalent to Osgood criterion [14], as it is proven in León and Villa [11]. It means, the solution of (1.1) explodes in finite time if and only if $\int_{x_0}^{\infty}(1/b(s))ds < \infty$. Also, when $\sigma \equiv 0$ and $b > 0$, Osgood [14] has stated that explosion time is finite if and only if $\int_{x_0}^{\infty}(1/b(s))ds < \infty$. In this case, the explosion time is equals to this integral.

Unfortunately, the distribution of the explosion time of equation (1.1) is not easy to calculate. One way to do it is using linear second-order ordinary differential equations. Indeed, Feller [7] has pointed out the Laplace transformation of this distribution is a bounded solution to some related ordinary differential equations (see Section 5.2 below for a generalization of this result). Also some numerical schemes have been analyzed in order to approximate the time of explosion (consult Dávila et al. [4]). In this paper, in Section 5.1, we also obtain the

2010 *Mathematics Subject Classification*: Primary 45R05, 60H10; Secondary 49K20.

Keywords and phrases: iterated logarithm theorem for martingales, Itô's formula, comparison theorems for integral and stochastic differential equations, Osgood criterion, partial differential equations of second order, time of explosion.

partial differential equation that has the distribution of the explosion time as a bounded solution.

Now consider the nonautonomous stochastic differential equation

$$\begin{aligned} dX_t &= b(t,X_t)dt+\sigma(t,X_t)dW_t, \quad t>0, \\ X_0 &= x_0. \end{aligned} \tag{1.2}$$

For this equation, Feller test and Osgood criterion are not useful anymore, but, in the case that σ is independent of x, we are still able to associate the Laplace transformation of the distribution of the explosion time of (1.2) with a partial differential equation as Theorem (5.10) below establishes.

The main purpose of this paper is to deal with some extensions of Osgood criterion for some equations of the form (1.2). For instance, Lemma (3.2) provides a better understanding of Theorem 2.1 in [3], or if, in (1.2), σ is independent of x, we obtain an extension of Osgood criterion by means of the law of iterated logarithm and comparison theorems. It is worth mentioning that versions of these important tools have been used to analyze global solutions of integral equations as it is done by Constantin [3], or to obtain an extension of Osgood criterion to integral equations with additive noise and with $0<b(t,x)=b(x)$ non-decreasing (see León and Villa [11]).

The paper is organized as follows. Our comparison theorem for integral equations is introduced in Section 3. Some extensions of Osgood criterion are given in Sections 2, 3 and 4. Finally, the relation between partial differential equations and finite blow-up is considered in Section 5.

2. Osgood criterion for some stochastic differential equation with diffusion coefficient

Let $\sigma:\mathbb{R}\to\mathbb{R}$ and $h:\mathbb{R}\to\mathbb{R}$ be a differentiable function and a continuous function, respectively. We consider the stochastic differential equation

$$X_t^\xi=\xi+\frac{1}{2}\int_0^t \sigma(X_s^\xi)\sigma'(X_s^\xi)h^2(s)ds+\int_0^t \sigma(X_s^\xi)h(s)dW_s, \quad t\geq 0, \tag{2.1}$$

where $\xi\in\mathbb{R}$. Here and in what follows, $W=\{W_t:t\geq 0\}$ is a Brownian motion.

Now we assume that there are $-\infty\leq x_1<x_2\leq\infty$ such that $\sigma\neq 0$ on (x_1,x_2). Let $\xi\in(x_1,x_2)$ be fixed and define $\Psi_\xi:(x_1,x_2)\to\mathbb{R}$ as

$$\Psi_\xi(x)=\int_\xi^x \frac{dz}{\sigma(z)}.$$

Set $l_\xi=\Psi_\xi(x_1)\wedge\Psi_\xi(x_2)$, $r_\xi=\Psi_\xi(x_1)\vee\Psi_\xi(x_2)$ and $Y_t=\int_0^t h(s)dW_s$, $t\geq 0$.

The following result is our first extension of Osgood criterion.

THEOREM (2.2). *Let $\tau_\xi=\inf\{t\geq 0: Y_t\notin(l_\xi,r_\xi)\}$. Then, the process $X_t^\xi=\Psi_\xi^{-1}(Y_t)$, $0\leq t<\tau_\xi$ is a solution of equation (2.1).*

REMARK (2.3). *In this case, τ_ξ is called the explosion time of the solution to equation (2.1).*

Proof. Applying Itô's formula with $f(x)=\Psi_\xi^{-1}(x)$, $x\in(l_\xi,r_\xi)$ we have

$$f(Y_{t\wedge\tau_\xi^k})-f(0)=\frac{1}{2}\int_0^{t\wedge\tau_\xi^k} f''(Y_s)h^2(s)ds+\int_0^{t\wedge\tau_\xi^k} f'(Y_s)h(s)dW_s, \tag{2.4}$$

where

$$\tau_\xi^k = \inf\{t > 0 : Y_t \notin (l_\xi + k^{-1}, r_\xi - k^{-1})\}.$$

Letting $k \to \infty$ in (2.4) we get the result holds. □

An immediate consequence of Theorem (2.2) is the following:

COROLLARY (2.5). *Let $\int_0^\infty h^2(s)ds = \infty$. Then the solution of equation (2.1) explodes in finite time if and only if either $l_\xi > -\infty$, or $r_\xi < \infty$. Moreover, if l_ξ and r_ξ are two real numbers, then*

$$P(\tau_\xi \in dt) = \sum_{k=-\infty}^{\infty} (-1)^k \frac{r_\xi + k(r_\xi - l_\xi)}{\sqrt{2\pi}(H(t))^{3/2}} \exp\left(-\frac{(r_\xi + k(r_\xi - l_\xi))^2}{2H(t)}\right) dt,$$

with $H(t) = \int_0^t (h(s))^2 ds$.

Proof. It is well-known that there is a Brownian motion $B = \{B_t : t \geq 0\}$ such that $Y_t = B_{H(t)}$, $t \geq 0$, (see, for instance, Durrett [6]). Let $\tilde{\tau}_\xi = \inf\{t > 0 : B_t \notin (l_\xi, r_\xi)\}$. Then, it is easy to show that $P(\tau_\xi \leq t) = P(\tilde{\tau}_\xi \leq H(t))$. Consequently, the proof follows from Borodin and Salminen [1] (page 212). □

REMARK (2.6). *Suppose that, for example, $\sigma > 0$, $\Psi_\xi(x_1) = -\infty$ and $\Psi_\xi(x_2) < \infty$. Then, as an immediate consequence of the proof of Corollary (2.5), we get that $\tau_\xi = \inf\{t > 0 : \int_0^t h(s)dW_s = \Psi_\xi(x_2)\}$ and*

$$P(\tau_\xi \leq t) = \Phi\left(\frac{\Psi_\xi(x_2)}{\sqrt{H(t)}}\right), \tag{2.7}$$

where

$$\Phi(x) = \frac{2}{\sqrt{2\pi}} \int_x^\infty e^{-z^2/2} dz.$$

Observe that we get a similar result when σ is negative, or the involved interval has the form (l_ξ, ∞).

Now we illustrate this remark with two examples.

EXAMPLE (2.8). *Let $\sigma(x) = |x|^\alpha$, $x \in \mathbb{R}$, $\alpha > 1$ and $\xi \in \mathbb{R}$. Then*

$$\Psi_\xi(x) = \begin{cases} \frac{1}{1-\alpha}(|x|^{1-\alpha} - |\xi|^{1-\alpha}), & \xi > 0,\ x \geq 0, \\ \frac{1}{1-\alpha}(|\xi|^{1-\alpha} - |x|^{1-\alpha}), & \xi < 0,\ x \leq 0. \end{cases}$$

Hence,

$$\Psi_\xi(-\infty) = \frac{|\xi|^{1-\alpha}}{1-\alpha} \quad \text{and} \quad \Psi_\xi(0) = \infty, \quad \text{for } \xi < 0,$$

and

$$\Psi_\xi(\infty) = \frac{|\xi|^{1-\alpha}}{\alpha - 1} \quad \text{and} \quad \Psi_\xi(0) = -\infty, \quad \text{for } \xi > 0.$$

Therefore, there is explosion in finite time and

$$P(\tau_\xi \leq t) = \Phi\left(\frac{|\xi|^{1-\alpha}}{(\alpha - 1)\sqrt{H(t)}}\right).$$

EXAMPLE (2.9). *Let* $\sigma(x) = e^{\alpha x}$, $x \in \mathbb{R}$, $\alpha \neq 0$ *and* $\xi \in \mathbb{R}$. *Then*

$$\Psi_\xi(x) = \frac{1}{\alpha}(e^{-\alpha\xi} - e^{-\alpha x}),$$

$$\Psi_\xi(-\infty) = \begin{cases} -\infty, & \alpha > 0, \\ \frac{1}{\alpha}e^{-\alpha\xi}, & \alpha < 0, \end{cases} \quad \text{and} \quad \Psi_\xi(\infty) = \begin{cases} \frac{1}{\alpha}e^{-\alpha\xi}, & \alpha > 0, \\ \infty, & \alpha < 0. \end{cases}$$

Thus we deduce that there is explosion on the left for $\alpha < 0$, *there is explosion on the right for* $\alpha > 0$ *and*

$$P(\tau_\xi \leq t) = \Phi\left(\frac{e^{-\alpha\xi}}{|\alpha|\sqrt{H(t)}}\right).$$

3. An extension of Osgood criterion for integral equations

In this section we generalize recent results obtained in [2] and [11]. Now we study the following nonautonomous integral equation

$$X_t^\xi = \xi + \int_0^t a(s)b(X_s^\xi)ds + g(t), \quad t \geq 0. \tag{3.1}$$

The explosion time T_ξ^X of this equation is defined as $T_\xi^X = \inf\{t \geq 0 : X_t^\xi \notin \mathbb{R}\}$. In the remaining of this paper we will need the following conditions:

H1:: $a : (0,\infty) \to (0,\infty)$ is a continuous function such that

$$\lim_{t\to\infty} \int_t^{t+\eta} a(s)ds > 0, \quad \text{for some } \eta > 0.$$

H2:: $b : \mathbb{R} \to [0,\infty)$ is a continuous function such that there exist $-\infty \leq l < \infty$ and $-\infty < r < \infty$ satisfying that $b > 0$ and locally Lipschitz on (l,∞), and $b : [r,\infty) \to (0,\infty)$ is non-decreasing.

H3:: $g : [0,\infty) \to \mathbb{R}$ is a continuous function such that

$$\limsup_{t\to\infty}\left(\inf_{0\leq h\leq\tilde{\eta}} g(t+h)\right) = \infty, \quad \text{for some } \tilde{\eta} > 0.$$

Henceforth we utilize the convention

$$A_t(x) = \int_t^x a(z)dz, \quad t \geq 0 \text{ and } x \in (t,\infty),$$

and

$$B_\xi(x) = \int_\xi^x \frac{dz}{b(z)}, \quad x \in (l,\infty).$$

We begin with the following generalization of Osgood criterion.

LEMMA (3.2). *Let* ***H1*** *and* ***H2*** *be satisfied and* $x_0 > l$. *Consider the ordinary differential equation*

$$\begin{aligned} \frac{dy(t)}{dt} &= a(t)b(y(t)), \quad t > t_0, \\ y(t_0) &= x_0. \end{aligned} \tag{3.3}$$

a) *Assume that* $B_{x_0}(\infty) \geq A_{t_0}(\infty)$, *then*

$$y(t) = B_{x_0}^{-1}(A_{t_0}(t)), \quad t \geq t_0.$$

b) *If* $B_{x_0}(\infty) < A_{t_0}(\infty)$, *then there is blow up in finite time and the time of explosion* $T_{x_0}^y$ *is equal to* $A_{t_0}^{-1}(B_{x_0}(\infty))$.

REMARK (3.4). *Observe that equation (3.3) (resp. equation (3.1)) has a unique solution for $x_0 > l$ (resp. for $\xi > l$) that may explode in finite time because of Hypotheses **H1** and **H2** (resp. **H1**-**H3**). This fact will be used in the proof of Theorem (3.6) below without mentioning.*

Proof. From (3.3) we see that

$$\int_{t_0}^{t} \frac{y'(s)}{b(y(s))} ds = \int_{t_0}^{t} a(s) ds.$$

The change of variable $z = y(s)$ yields $B_{x_0}(y(t)) = A_{t_0}(t)$.

Now we deal with Statement a). If $B_{x_0}(\infty) \geq A_{t_0}(\infty)$, then $B_{x_0}(\infty) > A_{t_0}(t)$, for all $t > t_0$. Therefore $y(t) = B_{x_0}^{-1}(A_{t_0}(t))$, $t > t_0$ is well-defined.

Finally we consider Statement b). In this case we have that $B_{x_0}^{-1}(A_{t_0}(t))$ is only defined for $t < A_{t_0}^{-1}(B_{x_0}(\infty)) < \infty$. □

Also we are going to need the following elementary comparison result.

LEMMA (3.5). *Let $x_0 > r$ and $T > t_0$. Assume that **H1** and **H2** are satisfied, and that $u, v : [t_0, T] \to \mathbb{R}$ are two continuous functions.*

a) *Suppose that u and v are such that*

$$\begin{aligned} v(t) &> x_0 + \int_{t_0}^{t} a(s) b(v(s)) ds, \quad t \in [t_0, T], \\ u(t) &= x_0 + \int_{t_0}^{t} a(s) b(u(s)) ds, \quad t \in [t_0, T]. \end{aligned}$$

Then $v(t) \geq u(t)$, for all $t \in [t_0, T]$.

b) *If*

$$\begin{aligned} r < v(t) &< x_0 + \int_{t_0}^{t} a(s) b(v(s)) ds, \quad t \in [t_0, T], \\ u(t) &= x_0 + \int_{t_0}^{t} a(s) b(u(s)) ds, \quad t \in [t_0, T]. \end{aligned}$$

Then $v(t) \leq u(t)$, for all $t \in [t_0, T]$.

Proof. We first deal with Statement a). Let $N = \{t \geq t_0 : b(u(s)) \leq b(v(s)),\ s \in [t_0, t]\}$. Since $t_0 \in N$, then the continuity of v and u, together with the fact that b is non-decreasing on (r, ∞), leads us to show that $\tilde{T} = \sup N > t_0$. If $\tilde{T} < T$ then

$$v(\tilde{T}) - u(\tilde{T}) > \int_{t_0}^{\tilde{T}} a(s)[b(v(s)) - b(u(s))] ds \geq 0,$$

which is impossible due to the definition of $\tilde{T}$.

Finally, we proceed similarly to prove that b) is also true and to finish the proof. □

THEOREM (3.6). *Let $\xi \in \mathbb{R}$. Assume **H1**-**H3**. Then the explosion time T_ξ^X of the solution X^ξ of (3.1) is finite if and only if*

$$\int_r^\infty \frac{ds}{b(s)} < \infty. \tag{3.7}$$

Proof. Suppose that $T_\xi^X < \infty$. Since g is continuous, then

$$\int_0^t a(s)b(X_s^\xi)ds \begin{cases} < \infty, & t < T_\xi^X, \\ = \infty, & t = T_\xi^X. \end{cases}$$

Hence, there is $t_0 \in (0, T_\xi^X)$ such that

$$\xi + \int_0^{t_0} a(s)b(X_s^\xi)ds + \inf_{s\in[0,T_\xi^X]} g(s) > r,$$

and consequently $X_t^\xi > r$ for $t \in [t_0, T_\xi^X]$.

Now set

$$M = \sup\{|g(t)| : 0 \le t \le T_\xi^X\} + \xi + \int_0^{t_0} a(s)b(X_s^\xi)ds.$$

This yields

$$X_t^\xi < M + 1 + \int_{t_0}^t a(s)b(X_s^\xi)ds, \quad t \in [t_0, T_\xi^X].$$

On the other hand, we consider the integral equation

$$u(t) = (M+1) + \int_{t_0}^t a(s)b(u(s))ds, \quad t \ge t_0.$$

Because $M > r$, Lemmas (3.2) and (3.5) give $T_{M+1}^u = A_{t_0}^{-1}(B_{M+1}(\infty)) \le T_\xi^X < \infty$. Whence

$$\int_{M+1}^\infty \frac{ds}{b(s)} < \infty.$$

The continuity and positivity of b in $[r,\infty)$ implies (3.7).

Reciprocally, suppose that X^ξ does not explodes in finite time. From Hypotheses **H1** and **H3**, we can find a sequence $\{t_n : n \in \mathbb{N}\}$ such that $t_n \uparrow \infty$ and

$$r + 1 < \xi + \inf_{0\le h\le\tilde\eta} g(t_n + h) \uparrow \infty, \text{ as } n \to \infty.$$

Observe that

$$X_{t+t_n}^\xi > \xi + \inf_{0\le h\le\tilde\eta} g(t_n + h) - 1 + \int_0^t a(s + t_n)b(X_{s+t_n}^\xi)ds, \quad t \in [0, \tilde\eta].$$

Now consider the integral equation

$$u(t) = \xi + \inf_{0\le h\le\tilde\eta} g(t_n + h) - 1 + \int_0^t a(s + t_n)b(u(s))ds, \quad t \in [0, \tilde\eta].$$

Therefore Lemmas (3.2) and (3.5) yield

$$\int_{\xi+\inf_{0\le h\le\tilde\eta} g(t_n+h)-1}^\infty \frac{ds}{b(s)} > \int_{t_n}^{t_n+\tilde\eta} a(s)ds.$$

Whence **H1** implies $\int_r^\infty \frac{ds}{b(s)} = \infty$. □

We finish this section with the following result for bounded noise.

PROPOSITION (3.8). *Assume Hypotheses **H1** and **H2**. Also assume that g in equation (3.1) is a bounded function and that $\xi + \inf_{s\ge0} g(s) > r$. Then, we have the following statements:*

a) $\int_r^\infty (1/b(s))ds = \infty$ *implies that the solution of equation (3.1) does not explode in finite time.*

b) $\int_r^\infty (1/b(s))ds < \infty$ *yields that the solution of equation (3.1) blows up in finite time and*

$$T_\xi^X \in (A_0^{-1}(B_{\xi+\sup_{s\geq 0} g(s)}(\infty)), A_0^{-1}(B_{\xi+\inf_{s\geq 0} g(s)}(\infty))).$$

Proof. Let $\varepsilon > 0$ be such that $\xi + \inf_{s\geq 0} g(s) > r + \varepsilon$. Set

$$Z_t^\xi = \xi + \sup_{s\geq 0} g(s) + \varepsilon + \int_0^t a(s)b(Z_s^\xi)ds$$

and

$$Y_t^\xi = \xi + \inf_{s\geq 0} g(s) - \varepsilon + \int_0^t a(s)b(Y_s^\xi)ds.$$

By Lemma (3.5) we have,

$$Y_t^\xi < X_t^\xi < Z_t^\xi, \quad t < T^Z_{\xi+\sup_{s\geq 0} g(s)+\varepsilon}.$$

Letting $\varepsilon \downarrow 0$ the proof is an immediate consequence of Lemma (3.2), and Hypotheses **H1** and **H2**. □

4. Stochastic differential equation with additive Wiener integral noise

In this section we study equation (3.1) when the noise g is a Wiener integral. More precisely, here we study the stochastic differential equation

$$X_t^\xi = \xi + \int_0^t a(s)b(X_s^\xi)ds + I_t, \tag{4.1}$$

where $I_t = \int_0^t f(s)dW_s$ and $f : [0,\infty) \to \mathbb{R}$ is a square-integrable function on $[0,M]$, for any $M > 0$.

In the remaining of this section we utilize the following assumption:

H4:: $\int_0^\infty f^2(s)ds = \infty$ and

$$\sum_{n=M}^{\infty} \frac{1}{\Upsilon^p(n)} \left(\int_n^{n+2} f^2(s)ds\right)^{p/2} < \infty, \tag{4.2}$$

for some $M, p > 0$, where

$$\Upsilon(t) = \sqrt{2\left(\int_0^t f^2(s)ds\right) \log\log\left(e^e \vee \int_0^t f^2(s)ds\right)}.$$

REMARK (4.3). *Observe that (4.2) holds if, for example,*

$$t \mapsto \left(\int_0^{t+2} f^2(s)ds\right)\left(\int_0^t f^2(s)ds\right)^{-1} - 1,$$

is a decreasing function in $L^p([M,\infty))$ for some $M, p > 0$.

On the other hand, as a consequence of iterated logarithm theorem for locally square integrable martingales, we can now state the following:

LEMMA (4.4). *Under the fact that $\int_0^\infty f^2(s)ds = \infty$, we have*

$$\limsup_{t\to\infty} \frac{I_t}{\Upsilon(t)} = 1 \quad \text{with probability one.} \tag{4.5}$$

Proof. The result is Theorem (1.1) in Qing Gao [9]. □

The following theorem is the main result of this section.

THEOREM (4.6). *Assume Hypotheses **H1**, **H2** and **H4**. Then the stochastic differential equation (4.1) blows up in finite time with probability 1 if and only if* $\int_r^\infty \frac{ds}{b(s)} < \infty$.

Proof. We first observe that, by Theorem (3.6), we only need to show that the paths of I satisfy Hypothesis **H3** almost surely.

Burkholder-Davis-Gundy inequality (see, for instance, Theorem 3.5.1 in [6]) yields

$$E\left[\left(\sup_{s,t\in[n,n+2]}|I_t-I_s|\right)^p\right] \le c_p\left(\int_n^{n+2} f^2(s)ds\right)^{p/2},$$

where c_p is a constant depending only on p. Then, by (4.2),

$$E\left[\sum_{n=M}^{\infty}\left(\sup_{s,t\in[n,n+2]}\frac{|I_t-I_s|}{\Upsilon(n)}\right)^p\right] \le c_p\sum_{n=M}^{\infty}\frac{1}{\Upsilon^p(n)}\left(\int_n^{n+2} f^2(s)ds\right)^{p/2} < \infty.$$

Therefore, it is enough to prove that $I(\omega)$ satisfies **H3** for $\omega\in\Omega$ for which there exists $n_0\in\mathbb{N}$ such that

$$\sup_{s,t\in[n,n+2]}\frac{|I_t(\omega)-I_s(\omega)|}{\Upsilon(n)} \le \frac{1}{4}, \quad \text{for } n\ge n_0$$

and (4.5) is satisfied. Hence, we can find a sequence $\{t_n : n\in\mathbb{N}\}$ such that $t_n > n$ and

$$\frac{I_{t_n}(\omega)}{\Upsilon(t_n)} \ge \frac{1}{2} \quad \text{for all } n\in\mathbb{N}.$$

Finally, using the properties established in this proof, we are able to write, for $n\ge n_0$,

$$\begin{aligned}
\inf_{s\in[t_n,t_n+1]} I_s(\omega) &= I_{t_n}(\omega) + \inf_{s\in[t_n,t_n+1]}\big(I_s(\omega)-I_{t_n}(\omega)\big)\\
&\ge I_{t_n}(\omega) + \inf_{s\in[t_n,t_n+1]}(-|I_s(\omega)-I_{t_n}(\omega)|)\\
&\ge I_{t_n}(\omega) - \left(\sup_{s,t\in[[t_n],[t_n]+2]}\frac{|I_s(\omega)-I_t(\omega)|}{\Upsilon([t_n])}\right)\Upsilon([t_n])\\
&\ge \frac{1}{2}\Upsilon(t_n) - \frac{1}{4}\Upsilon([t_n]) \ge \frac{1}{4}\Upsilon(t_n)\to\infty,
\end{aligned}$$

as $n\to\infty$, where $[t]$ is the integer part of t and, in the last inequality, we have used that Υ is a non-decreasing function. □

Now, in order to state a consequence of Theorem (4.6), we consider the equation

$$Y_t = \xi + \int_0^t \tilde{b}(s,Y_s)ds + I_t, \quad t\ge 0. \tag{4.7}$$

Here, for each $T>0$, the function $\tilde{b}:[0,\infty)\times\mathbb{R}\to[0,\infty)$ is locally Lipschitz (uniformly on $s\in[0,T]$), $b(\cdot,x)$ is continuous, for $x\in\mathbb{R}$, and I satisfy Hypothesis **H4** with f continuous. Remember that, in this case, equation (4.7) has a unique solution that may explode in finite time.

COROLLARY (4.8). *Let a and b satisfy Conditions **H1** and **H2**, respectively. Assume that $\xi\in\mathbb{R}$, b is locally Lipschitz, $\int_r^\infty(1/b(x))dx<\infty$ (resp. $\int_r^\infty(1/b(x))dx=\infty$) and $a(s)b(x)\le\tilde{b}(s,x)$ (resp. $\tilde{b}(s,x)\le a(s)b(x)$), $(s,x)\in[0,\infty)\times\mathbb{R}$. Then, the solution to equation (4.7) explodes (resp. does not explode) in finite time.*

Proof. We only consider the case that $\int_r^\infty (1/b(x))dx = \infty$ and $\tilde{b}(s,x) \le a(s)b(x)$, since the proof is similar for the other one.

Let X^ξ and Y be the solutions of equations (4.1) and (4.7), respectively. Then, from Milian [13] (Theorem 2), we get

$$Y_t \le X_t^\xi, \quad t \ge 0.$$

Thus, by Theorem (4.6), the solution Y of equation (4.7) cannot explode in finite time because it cannot go to $-\infty$ in finite time since $\tilde{b}$ is $\mathbb{R}_+$-valued and I has continuous paths and, consequently, bounded paths on compact intervals of $[0,\infty)$. Therefore the proof is complete. □

EXAMPLE (4.9). *Take*

$$\begin{aligned} a(x) &= x^\alpha, \ x \in (0,\infty), \\ b(x) &= 8x^2 - 36x + 48, \ x \in \mathbb{R}, \\ f(x) &= x^\beta, \ x \in (0,\infty), \ \beta > -\frac{1}{2}. \end{aligned}$$

Hence

$$\lim_{t\to\infty} \int_t^{t+1} x^\alpha dx = \begin{cases} +\infty, & \alpha > 0, \\ 1, & \alpha = 0, \\ 0, & \alpha < 0, \end{cases}$$

and

$$\frac{(t+2)^{2\beta+1}}{t^{2\beta+1}} - 1 = \left(1 + \frac{2}{t}\right)^{2\beta+1} - 1 \le C\frac{1}{t}.$$

The last function belongs to $L^p([1,\infty])$, for any $p > 1$. Thus f satisfied (4.2) due to Remark 4.3.

On the other hand, it is clear that $\int_\xi^\infty \frac{dx}{8x^2-36x+48} < \infty$, $\xi > 0$. Then

$$X_t^\xi = \xi + \int_0^t s^\alpha (8(X_s^\xi)^2 - 36(X_s^\xi) + 48)ds + \int_0^t s^\beta dW_s,$$

explodes in finite time when $\alpha \ge 0$. Notice that b is not necessarily increasing as in [11] or [2]. Moreover, we can improve Theorem (4.6) in some particular cases, see [15].

EXAMPLE (4.10). *The function $Y_t \equiv 1$ is solution to*

$$Y_t = 1 + \int_0^t (Y_s)^2 ds - t, \quad t \ge 0.$$

*Although $\int_1^\infty (1/s^2)ds < \infty$, Y does not blow-up in finite time because $g(t) = -t$, $t \ge 0$, does not satisfies Hypothesis **H3**.*

Also notice that $f(t) = \exp(\exp(t))$, $t \ge 0$, does not satisfies (4.2). We intuitively understand that in this case the noise is to strong and we have also blow up in finite time, for any initial condition. We have a contrary effect as in Example (4.10).

PROPOSITION (4.11). *Let f and I be defined as in equation (4.1). Suppose **H1**, **H2** and $\int_0^\infty f^2(s)ds < \infty$ are satisfied. Then I is bounded with probability one and, under the assumption $\xi + \inf_{s\ge 0} I_s > r$, the stochastic differential equation (4.1) blows up in finite time if and only if $B_r(\infty) < \infty$.*

REMARK (4.12). *Observe that $\xi + \inf_{s\ge 0} I_s$ depends on ω.*

Proof. The result follows from [6] (Lemma 3.4.7 and Theorem 3.4.9), and Proposition (3.8). □

5. An approach to obtain the distribution of the explosion time of a stochastic differential equation

Now we study some stochastic differential equations of the form

$$X_t^\xi = \xi + \int_0^t b(s, X_s^\xi)ds + \int_0^t \sigma(s, X_s^\xi)dW_s, \quad t \geq 0. \tag{5.1}$$

Namely, we propose a method to figure out the distribution of the explosion time τ_ξ of X^ξ. Intuitively, τ_ξ is a stopping time such that (5.1) has a solution up to this stopping time and $\limsup_{t\uparrow\tau_\xi} |X_t| = \infty$.

(5.1) Autonomous case. This section is devoted to deal with the stochastic differential equation

$$X_t^\xi = \xi + \int_0^t b(X_s^\xi)ds + \int_0^t \sigma(X_s^\xi)dW_s, \quad t \geq 0,$$

with $b, \sigma \in C^1(\mathbb{R})$. In this case, McKean [12] has shown that $X^\xi_{(\tau_\xi)-} \in \{-\infty, \infty\}$ on $[\tau_\xi < \infty]$. So, henceforth, we can utilize the convention

$$\tau_\xi^+ = \inf\{t > 0 : X_t^\xi = \infty\} \quad \text{and} \quad \tau_\xi^- = \inf\{t > 0 : X_t^\xi = -\infty\}.$$

THEOREM (5.2). *Consider a bounded function $u : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ that satisfies the following boundary value problem:*

$$\frac{\partial u}{\partial t}(t,x) = \frac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2}(t,x) + b(x)\frac{\partial u}{\partial x}(t,x), \quad t > 0 \text{ and } x \in \mathbb{R}, \tag{5.3}$$

$$u(0,x) = 0, \quad \text{for all } x \in \mathbb{R}. \tag{5.4}$$

a) *Assume that $u(t,\infty) = u(t,-\infty) = 1$. Then $P(\tau_\xi \leq t) = u(t,\xi)$.*
b) *$u(t,\infty) = 1$ and $u(t,-\infty) = 0$ implies that $P(\tau_\xi^+ \leq t) = u(t,\xi)$.*
c) *If $u(t,\infty) = 0$ and $u(t,-\infty) = 1$, we have $P(\tau_\xi^- \leq t) = u(t,\xi)$.*

Remarks

1) *Maximum principle provides with conditions on b and σ such that the solution of equation (5.3) is bounded (see Friedman [8]).*
2) *It is quite interesting to observe that (5.3) is related to transition density of process X^ξ, or related to the fundamental solution of the associated Cauchy problem (see [8]). On the other hand, (5.4) and the conditions in Statement a)-c) are intuitively clear. In fact, (5.4) establishes that if we begin at a real point ($\xi \in \mathbb{R}$), then we need some time to get blow-up. And other conditions mean that if we begin at cementery state ($\pm\infty$), then the time to blow-up is less than any time.*
3) *Observe that $P(\tau^\xi \leq t) = P(\tau_+^\xi \leq t) + P(\tau_-^\xi \leq t)$ and that, for example in Statement a), we have $P(\tau^\xi < \infty) = u(\infty, \xi)$.*
4) *If X^ξ does not explodes in finite time, then equation (5.3)–(5.4) has not a bounded solution satisfying conditions established in either Statement a), b), or c).*

Proof. Using Itô's formula on $0 \le s < t$ and that u is solution to (5.3) we obtain

$$u(t-(s\wedge\tau_\xi^m), X^\xi_{s\wedge\tau_\xi^m}) = u(t,\xi) + \int_0^{s\wedge\tau_\xi^m} \frac{\partial u}{\partial x}(t-r, X_r^\xi)\sigma(X_r^\xi)dW_r,$$

where $\tau_\xi^m = \inf\{t>0 : |X_t^\xi| > m\}$. Since u is bounded, then the above stochastic integral is a martingale. Therefore

$$u(t,\xi) = E\left[u(t-(s\wedge\tau_\xi^m), X^\xi_{s\wedge\tau_\xi^m})\right].$$

Letting $s \uparrow t$, then continuity of X^ξ and the boundedness of u, together with the dominated convergence theorem, allow us to write

$$\begin{aligned} u(t,\xi) &= E\left[u(t-(t\wedge\tau_\xi^m), X^\xi_{t\wedge\tau_\xi^m})\right] \\ &= E\left[u(t-(t\wedge\tau_\xi^m), X^\xi_{t\wedge\tau_\xi^m}), \tau_\xi \le t\right] \\ &\quad + E\left[u(t-(t\wedge\tau_m^\xi), X^\xi_{t\wedge\tau_m^\xi}), \tau_\xi > t\right]. \end{aligned}$$

Taking $m \to \infty$,

$$u(t,\xi) = E\left[u(t-\tau_\xi, X^\xi_{\tau_\xi}), \tau_\xi \le t\right] + E\left[u(0, X_t^\xi), \tau_\xi > t\right]. \tag{5.5}$$

Now we consider Statement a),

$$u(t,\xi) = E\left[u(t-\tau_\xi, X^\xi_{\tau_\xi}), \tau_\xi \le t\right] = P(\tau_\xi \le t).$$

Statement b) is proven as follows. From equality (5.5) we get

$$\begin{aligned} u(t,\xi) &= E\left[u(t-\tau_\xi, X^\xi_{\tau_\xi}), \tau_\xi \le t, \tau_\xi^+ < \tau_\xi^-\right] \\ &\quad + E\left[u(t-\tau_\xi, X^\xi_{\tau_\xi}), \tau_\xi \le t, \tau_\xi^- < \tau_\xi^+\right] \\ &= P(\tau_\xi^+ \le t). \end{aligned}$$

Finally, Statement c) is proven similarly. So the proof is complete. □

EXAMPLE (5.6). a) *In Example 2.8, with $h \equiv 1$, we have*

$$\bar{u}(t,\xi) = \begin{cases} \Phi\left(\frac{|\xi|^{1-\alpha}}{(\alpha-1)\sqrt{t}}\right), & \xi > 0, \\ 0, & \xi \le 0, \end{cases}$$

and

$$\underline{u}(t,\xi) = \begin{cases} 0, & \xi > 0, \\ \Phi\left(\frac{|\xi|^{1-\alpha}}{(\alpha-1)\sqrt{t}}\right), & \xi \le 0, \end{cases}$$

satisfy Statements b) and c) of Theorem (5.2), respectively. In particular, if $\xi > 0$, then $\bar{u}(\infty,\xi) = 1$, therefore we have a positive blow-up. This phenomenon is explained by Milan [13] (Theorem 1) due to the involved solution being a nonnegative process when $\xi > 0$.

b) *For $\beta > 0$, the partial differential equation*

$$\begin{aligned} \frac{\partial u}{\partial t}(t,x) &= \frac{a^2}{2}e^{2\beta x}\frac{\partial^2 u}{\partial x^2}(t,x) + \beta a^2 e^{2\beta x}\frac{\partial u}{\partial x}(t,x), \\ u(0,x) &= 0, \quad \forall x \in \mathbb{R}, \end{aligned}$$

has solution,

$$u(t,x) = \exp\left(-\frac{e^{-2\beta x}}{2(a\beta)^2 t}\right).$$

Since $u(t,\infty) = 1$ and $u(t,-\infty) = 0$, then the distribution of explosion time to the stochastic differential equation

$$X_t^x = x + \int_0^t \beta a^2 e^{2\beta X_s^x} ds + \int_0^t a e^{\beta X_s^x} dW_s,$$

is given by

$$P(\tau_\xi \le t) = \exp\left(-\frac{e^{-2\beta x}}{2(a\beta)^2 t}\right)$$

because of $P(\tau_\xi^+ < \infty) = 1$.

REMARK (5.7). *It is not difficult to see that Examples 2.8 and 2.9 are solution of the corresponding partial differential equations (PDEs), then we conjecture that the distribution of the explosion time is the solution of such a PDE. If this is true, then we have the following criterion of explosion: There is explosion in finite time if and only if the corresponding PDE has a bounded solution. Moreover, this criterion could be applied in more dimensions and for non autonomous processes (see [8]).*

(5.2) Laplace transform of the distribution of the explosion time. Finally, in this subsection we indicate how we could calculate the Laplace transformation of the distribution of the explosion time τ_ξ of the solution to equation (4.7). It means, we assume that the equation

$$X_t^\xi = \xi + \int_0^t b(s, X_s^\xi) ds + I_t, \quad t \ge 0, \tag{5.8}$$

has a unique solution that may blow-up in finite time, where b takes values in $\mathbb{R}_+$ and $I_t = \int_0^t f(s) dW_s$. Note that if $\omega \in \Omega$ is such that $\tau_\xi(\omega) < \infty$, then $X_t^\xi > \xi + \inf_{0 \le s \le \tau_\xi(\omega)} I_t(\omega)$, $t \le \tau_\xi(\omega)$, and consequently $\int_0^{\tau_\xi(\omega)} b(s, X_s^\xi) ds = \infty$. Thus, $X_{\tau_\xi -}^\xi = \infty$, on $[\tau_\xi < \infty]$, and $\tau_\xi = \tau_\xi^+$.

We begin with an auxiliary result.

LEMMA (5.9). *Let $\lambda > 0$ and τ_ξ the explosion time of the solution of equation (5.8). Then*

$$E\left(e^{-\lambda \tau_\xi}\right) = \lambda \int_0^\infty P(\tau_\xi \le u) e^{-\lambda u} du.$$

Proof. Let us denote the distribution of τ_ξ by F_{τ_ξ}. Fubini theorem leads to justify

$$\begin{aligned} E\left(e^{-\lambda \tau_\xi}\right) &= \int_{(0,\infty)} e^{-\lambda s} F_{\tau_\xi}(ds) = \lambda \int_{(0,\infty)} \left(\int_s^\infty e^{-\lambda u} du\right) F_{\tau_\xi}(ds) \\ &= \lambda \int_0^\infty \left(\int_{(0,u]} F_{\tau_\xi}(ds)\right) e^{-\lambda u} du = \lambda \int_0^\infty F_{\tau_\xi}(u) e^{-\lambda u} du. \end{aligned}$$

Consequently, the proof is complete. □

Now we can state the main result of this subsection.

THEOREM (5.10). *Consider $\lambda > 0$, the explosion time τ_ξ of the solution of (5.8) and a bounded function $u : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ that is a solution of the partial differential equation*

$$\begin{aligned} -\frac{\partial u}{\partial t}(t,x) &= \frac{1}{2} f^2(t) \frac{\partial^2 u}{\partial x^2}(t,x) + b(t,x) \frac{\partial u}{\partial x}(t,x) - \lambda u(t,x), \ t > 0 \text{ and } x \in \mathbb{R}, \\ u(t,\infty) &= 1. \end{aligned}$$

Then, $\lambda \int_0^\infty P(\tau_\xi \le u) e^{-\lambda u} du = u(0,\xi)$.

Proof. As in the proof of Theorem (5.2) Itô's formula gives

$$u(0,\xi) = E\left(u(t \wedge \tau_\xi^m, X^\xi_{t \wedge \tau_\xi^m}) \exp(-\lambda(t \wedge \tau_\xi^m)) \right).$$

We can now easily complete the proof of this result by combining Lemma (5.9), the arguments used in the last part of the proof of Theorem (5.2) and the fact that $X^\xi_{\tau_\xi^m} \to \infty$ as $m \to \infty$ on $[\tau_\xi < \infty]$. Indeed we first take $m \uparrow \infty$, and then $t \uparrow \infty$. □

REMARK (5.11). *In some cases we have the converse of Theorem (5.10). For example, consider the stochastic differential equation*

$$X_t^\xi = \xi + \int_0^t (g(X_s^\xi) + a - b) ds + cW_t, \ \ t \ge 0,$$

where $a, b, c \in \mathbb{R}$, $a > b$ and $g : \mathbb{R} \to [0,\infty)$. Then the associated ordinary differential equation is

$$\begin{aligned} \frac{c^2}{2} w''(x) + (g(x) + a - b) w'(x) - \lambda w(x) &= 0, \ \ t > 0, \\ w(\infty) &= 1. \end{aligned} \tag{5.12}$$

Therefore, if X^ξ explodes in finite time then (5.12) has a bounded solution, in fact it is the Laplace transform of the explosion time (see [7]). Then the solution of

$$\begin{aligned} -\frac{\partial u}{\partial t}(t,x) &= \frac{c^2}{2} \frac{\partial^2 u}{\partial x^2}(t,x) + (g(x - bt) + a) \frac{\partial u}{\partial x}(t,x) - \lambda u(t,x), \ \ t > 0, \ x \in \mathbb{R}, \\ u(t,\infty) &= 1. \end{aligned}$$

is given by $u(t,x) = w(x - bt)$.

Acknowledgements

Professor Villa-Morales was partially supported by grant 118294 of CONACyT and grant PIM13-3N of UAA. The authors Thanks Universidad Autónoma de Aguascalientes and Cinvestav-IPN for their hospitality and economical support. Other authors were also partially supported by a CONACyT grant.

Received April 18, 2013

Final version received August 17, 2013

JORGE A. LEÓN
DEPARTAMENTO DE CONTROL AUTOMÁTICO
CINVESTAV-IPN
APARTADO POSTAL 14–740
07000 MÉXICO D.F., MEXICO
jleon@ctrl.cinvestav.mx

LILIANA PERALTA HERNÁNDEZ
DEPARTAMENTO DE CONTROL AUTOMÁTICO
CINVESTAV-IPN
APARTADO POSTAL 14–740
07000 MÉXICO D.F., MEXICO
lylyaanaa@gmail.com

JOSÉ VILLA-MORALES
DEPARTAMENTO DE MATEMÁTICAS Y FÍSICA
UNIVERSIDAD AUTÓNOMA DE AGUASCALIENTES
AVENIDA UNIVERSIDAD 940, CIUDAD UNIVERSITARIA
20131 AGUASCALIENTES, AGS., MEXICO
jvilla@correo.uaa.mx

REFERENCES

[1] A.N. BORODIN AND P. SALMINEN, Handbook of Brownian Motion-Facts and Formulae, Second Edition, Birkhäuser, 2002.
[2] M.J. CEBALLOS-LIRA, J.E. MACÍAS-DÍAZ, AND J. VILLA, *A generalization of Osgood's test and a comparison criterion for integral equations with noise*, Electronic Journal of Differential Equations **2011**, no. 05, (2011), 1–8.
[3] A. CONSTANTIN, *Global existence of solutions for perturbed differential equations*, Ann. Mat. Pura Appl. **CLXVIII**, no. IV, (1995), 237–299.
[4] J. DÁVILA, J.F. BONDER, J.D. ROSSI, P. GROISMAN, AND M. SUED, *Numerical analysis of stochastic differential equations with explosions*, Stoch. Anal. Appl. **23**, no.4, (2005), 809–825.
[5] A. DE PABLO, R. FERREIRA, F. QUIRÓS, AND J.L. VÁZQUEZ, *Blow-up. El problema matemático de explosión para ecuaciones y sistemas de reacción-difusión*, Bol. Soc. Esp. Mat. Apl. **32**, (2005), 75–111.
[6] R. DURRETT, Stochastic Calculus: A Practical Introduction, CRC Press, 1996.
[7] W. FELLER, *Diffusion processes in one dimension*, Trans. Amer. Math. Soc. **77**, (1954), 1–31.
[8] A. FRIEDMAN, Partial Differential Equations of Parabolic Type, Dover, 1964.
[9] FU QING GAO, *Laws of the Iterated Logarithm for Locally Square Integrable Martingales*, Acta Mathematica Sinica, English Series **25**, no. 2, (2009), 209–222.
[10] I. KARATZAS AND S.E. SHREVE, Brownian Motion and Stochastic Calculus, Second Edition, Springer-Verlag, 1991.
[11] J.A. LEÓN AND J. VILLA, *An Osgood criterion for integral equations with applications to stochastic differential equations with an additive noise*, Statistics & Probability Letters **81**, no. 4, (2011), 470–477.
[12] H.P. MCKEAN, Stochastic Integrals, Academic Press, (1969).
[13] A. MILIAN, *Stochastic Viability and comparison theorem*, Colloquium Mathematicum **68**, no. 2, (1995), 297–316.
[14] W.F. OSGOOD, *Beweis der Existenz einer Lösung der Differentialgleichung* $dy/dx = f(x,y)$ *ohne Hinzunahme der Cauchy-Lipschitz'schen Bedigung Monatsh*, Math. Phys. (Vienna) **9**, no. 1, (1898), 331–345.
[15] J. VILLA, *Un ejemplo de explosión en ecuaciones diferenciales estocásticas con ruido aditivo*, Aportaciones Matemáticas SMM, Comunicaciones **44**, (2011), 187–194.

Bol. Soc. Mat. Mexicana (3) Vol. 19, 2013

LIMIT THEOREMS AND INFINITE DIVISIBILITY IN NON-COMMUTATIVE PROBABILITY

OCTAVIO ARIZMENDI AND CARLOS VARGAS

ABSTRACT. We introduce in a unified way the basic theory on additive limit theorems and infinite divisibility for the natural notions of stochastic independence in non-commutative probability: tensor (classical), Boolean, free and monotone. We also present specialized, recent, results on free infinite divisibility.

1. Introduction

The concept of stochastic independence is fundamental in probability theory. During the 80's, Hudson and Parthasarathy [35] and Voiculescu [61] introduced quantum (or non-commutative) probability spaces, where new notions of independence may occur. The idea traces back to von Neumann (1932), who, was aiming at the mathematical foundation for the statistical questions in quantum mechanics. In an attempt to tackle key problems on operator algebras, Voiculescu launched the theory of free probability, which is based on the concept of free independence. This field started to draw more attention a few years later, when natural realizations of free independence and practical applications of free probability were found in several areas of mathematics, such as random matrix theory [64], representation theory [21] and combinatorics [54].

Limit theorems play a central role in classical probability. In particular, the class of infinitely divisible measures arises when looking at some of these limit theorems. The concept of infinite divisibility of probability distributions was introduced in 1929 by Bruno de Finetti [29]. The theory was soon enriched in the 30's by Kolmogorov [38], Lévy [40, 41] and Khintchine [37], culminating with the Lévy-Khintchine representation, which gives a complete characterization of infinitely divisible measures.

Starting in the 60's, the interest in the theory increased due to its practical applications (mainly to waiting time theory and modeling). These applications required explicit examples of infinite divisible distributions and hence motivated the search for efficient criteria to test the infinite divisibility of given measures. During the subsequent three decades, a variety of methods were developed and the infinite divisibility of a large number of new classes of distributions (such as mixtures, GGC, Thorin class, etc.) was established (see [23, 30, 59]).

Similar limit theorems in free probability and the concept of free infinite divisibility were considered since the beginning and a Lévy-Khintchine type characterization of free infinitely divisible measures was given by Bercovici and Voiculescu in [19]. Later, the notions of Boolean [58] and monotone independence [44] were introduced and the corresponding Lévy-Khintchine representations were found.

2010 *Mathematics Subject Classification*: 46L54, 60E07.
Keywords and phrases: non-commutative probability, infinite divisibilty, free probability.

On the other hand, Speicher's combinatorial approach to free probability [54, 48] (which was later extended to the Boolean [58] and monotone [33] cases) provided new characterizations of infinitely divisible measures in terms of cumulants. However, until recent, explicit examples were quite rare.

In the last years, with the introduction of non-classical stochastic processes [22], there has been an increasing interest in finding explicit examples and developing criteria to effectively test non-classical infinite divisibility.

The purpose of this note is to survey about limit theorems and infinite divisibility with respect to these non-classical notions of independence. We start with the basic theory and present in a unified way the fundamental limit theorems and characterizations of infinitely divisible distributions. In the last section, we specialize on the free case, presenting very recent results and examples.

The selection of material included in this survey is determined partly by the personal interests of the authors.

2. Non-Commutative Probability Spaces

Hudson and Parthasarathy [35] and Voiculescu [61] introduced quantum (or non-commutative) generalizations of probability spaces inspired by the algebraic approach to probability theory: All the information about the joint distributions of compactly supported random variables $X_1,\dots,X_n$ can be retrieved from evaluating their mixed moments (i.e. the values of $\mathbb{E}(X_{i_1}\cdots X_{i_k}), k \geq 1, i_j \leq n$). For instance, the notion of independence of the n-tuple $X_1,\dots,X_n$ is equivalent to the condition, that for all $k_1,\dots,k_n \geq 0$,

$$\mathbb{E}(X_1^{k_1}\cdots X_n^{k_n}) = \mathbb{E}(X_1^{k_1})\cdots\mathbb{E}(X_n^{k_n}). \tag{2.1}$$

A probability space can then be understood as a pair $(\mathcal{A},\mathbb{E})$ where $\mathcal{A}$ is a unital, commutative algebra (over $\mathbb{C}$) of random variables and $\mathbb{E}:\mathcal{A}\to\mathbb{C}$ is a unital linear functional.

The idea now is to consider a pair $(\mathcal{A},\tau)$, consisting of a non-commutative algebra $\mathcal{A}$ with unit $1_{\mathcal{A}}$ and a linear functional $\tau:\mathcal{A}\to\mathbb{C}$. Elements $a_1,\dots,a_k\in\mathcal{A}$ are called (non-commutative) random variables and the values $\tau(a_{i_1},\dots,a_{i_k}), k \geq 1, 1\leq i_1,\dots,i_k\leq n$ are the (mixed) moments of the n-tuple $a_1,\dots,a_k$.

Let us consider the richer structure where $\mathcal{A}$ is a C^*-algebra and τ a state, (i.e a unital, positive linear functional). Then for every normal element $a\in\mathcal{A}$, (i.e. $a^*a = aa^*$), there exist a unique (compactly supported) probability measure μ_a on $\mathbb{C}$ which encodes all the information about the mixed moments of $\{a,a^*\}$:

$$\int_{\mathbb{C}} z^n\bar{z}^m\,d\mu_a(z) = \tau(a^n(a^*)^m), \quad \forall n,m\geq 0.$$

In particular, the distribution μ_a of self-adjoint random variables $a=a^*$ is supported on the real line. We will concentrate mainly in such elements.

When we talk about the distribution of non-normal elements, we refer to the collection of mixed moments of $\{a,a^*\}$, which in general can not be encoded in a probability measure. We should point out that every element in a commutative probability space is normal, and hence these generic algebraic distributions can only be realized in non-commutative probability spaces.

The importance of such algebraic distributions will be more evident when we consider the mixed moments of several non-commutative random variables. In

the same way that the factorization of mixed moments encodes the independence relation of the considered variables, there are new factorizations which encode interesting relations between non-commutative random variables.

(2.1) Five Notions of Independence. If we fix the individual distributions of two elements $a, b \in \mathcal{A}$, their joint distribution (mixed moments) can be quite arbitrary, unless some notion of independence is assumed to hold between a and b. Several classification works by A. Ben Ghorbal and M Schürman[16], Muraki [**?**] and Speicher[55], have discussed the essential properties of classical independence which should be present in a non-commutative notion of independence. Under the most general assumptions considered in these works, there are only five notions of independence: tensor (classical), free, Boolean, monotone and anti-monotone. Each type of independence can be thought as a rule for calculating mixed moments, in the spirit of eq. (2.1). Let $(A_n)_{n\geq 1}$ be a sequence of subalgebras of $\mathcal{A}$.

We say that $(A_n)_{n\geq 1}$ are *tensor-independent* iff, for any $k \geq 1$ and any k-tuple $a_1, \ldots, a_k \in \mathcal{A}$, such that $a_i \in A_{j(i)}$, $1 \leq i \leq k$,

$$\tau(a_1 a_2 \cdots a_k) = \prod_{s \in I} \tau\left(\prod_{j(i)=s} a_i\right), \tag{2.2}$$

where $I = \{j(i) | i \leq k\} \subset \mathbb{N}$ and the a_i are multiplied in the same order as they appear in the left hand side.

The subalgebras $(A_n)_{n\geq 1}$ are *Boolean-independent* iff

$$\tau(a_1 a_2 \cdots a_k) = \tau(a_1)\tau(a_2)\cdots\tau(a_k) \tag{2.3}$$

whenever $k \geq 1$, $a_1, \ldots, a_k \in \mathcal{A}$, are such that $a_i \in A_{j(i)}$, $1 \leq i \leq k$ and $j(i) \neq j(i+1)$.

If we write $\bar{a} := a - \tau(a)$ for $a \in \mathcal{A}$, we say that $(A_n)_{n\geq 1}$ are *free* iff

$$\tau(\bar{a}_1 \bar{a}_2 \cdots \bar{a}_k) = 0, \tag{2.4}$$

whenever $k \geq 1$, $a_1, \ldots, a_k \in \mathcal{A}$, are such that $a_i \in A_{j(i)}$, $1 \leq i \leq k$ and $j(i) \neq j(i+1)$.

The linearly ordered subalgebras $A_1 < A_2 < \ldots$ are *monotone-independent* iff the following rule holds for any $k \geq 1$ and any k-tuple $a_1, \ldots a_k \in \mathcal{A}$, such that $a_i \in A_{j(i)}$, $1 \leq i \leq k$ and $j(i) \neq j(i+1)$:

$$\tau(a_1 \cdots a_{m-1} a_m a_{m+1} \cdots a_k) = \tau(a_1 \cdots a_{m-1} a_{m+1} \cdots a_k)\tau(a_m) \tag{2.5}$$

whenever $j(m) = \max\{j(s) | 1 \leq s \leq k\}$ (the *anti-monotone independence* is obtained by replacing max by min).

For any notion of independence, we say that $a_1, \ldots a_n, \in \mathcal{A}$ are independent if so are the (not necessarily unital) algebras $\langle a_1 \rangle, \ldots, \langle a_n \rangle$ ($\langle a_i \rangle$ is the algebra generated by a_i). With the exception of tensor independence, these new notions are only meaningful in non-commutative spaces: If $a, b \in \mathcal{A}$ commute and are Boolean, free, or monotone independent, then either a or b is a multiple of the identity, essentially. If $\mathcal{A}$ is commutative, tensor independence reduces to the usual classical independence.

A nice facet of non-commutative probability is that many objects in mathematics may be seen as random variables, a prominent example being matrices, and random versions of them. Surprisingly, free independence describes the collective behavior of large random matrices. In his seminal work [64], Voiculescu introduced a new conceptual way of dealing with random matrices, starting with

a multivariate version of Wigner's semicircle law [66]: One may consider a non-commutative probability space of $n \times n$ random matrices (with $\tau := n^{-1}\mathbb{E} \circ Tr$). The mixed moments of (entrywise) independent Gaussian matrices $X_1, \ldots, X_k$ converge as $(n \to \infty)$ to the corresponding mixed moments of free semicircular non-commutative random variables $s_1, \ldots, s_k$. In other words, Gaussian random matrices are asymptotically free semicirculars (see Section 3). As the theory of free probability developed, broader applications to Random Matrices have been found.

In theory, if $a, b \in \mathcal{A}$ are self-adjoint and satisfy an independence relation, we may compute the mixed moments of a and b. These are enough to describe the distribution of the self-adjoint element $a+b$. However, in practice, one uses either analytical or combinatorial means to explicitly compute μ_{a+b} in terms of μ_a and μ_b. Understanding and developing these analytic and combinatorial properties of distributions of non-commutative variables is of eminent importance for progress in the field.

(2.2) Non-classical Convolutions. Recall that the classical convolution of two probability measures μ_1, μ_2 on $\mathbb{R}$ is defined as the probability measure $\mu_1 * \mu_2$ on $\mathbb{R}$ such that $C_{\mu_1 * \mu_2}(t) = C_{\mu_1}(t) + C_{\mu_2}(t), t \in \mathbb{R}$, where $C_\mu(t) = \log \hat{\mu}(t)$ with $\hat{\mu}(t)$ the characteristic function of μ. Classical convolution corresponds to the sum of tensor independent random variables: $\mu_a * \mu_b = \mu_{a+b}$, for a and b independent random variables. The (classical) cumulants are the coefficients $c_n = c_n(\mu)$ in the series expansion (whenever it exists)

$$C_\mu(t) = \sum_{n=1}^{\infty} \frac{c_n}{n!}(it)^n.$$

Similar convolutions and related transforms exist for the free, Boolean and monotone theories.

We will denote by $\mathcal{M}, \mathcal{M}^+$, respectively, the set of Borel probability measures on $\mathbb{R}$ and $\mathbb{R}_+ := [0, \infty)$. The complex upper and lower half-planes are respectively denoted by $\mathbb{C}^+$ and $\mathbb{C}^-$. Let $G_\mu(z) = \int_{\mathbb{R}} \frac{\mu(dx)}{z-x}$ $(z \in \mathbb{C}^+)$ be the Cauchy transform of $\mu \in \mathcal{M}$.

Free convolution was defined in [61] for compactly supported probability measures and later extended in [43] for the case of finite variance, and in [19] for the general unbounded case. Let $G_\mu(z)$ be the Cauchy transform of $\mu \in \mathcal{M}$ and $F_\mu(z)$ its reciprocal $\frac{1}{G_\mu(z)}$. It was proved in Bercovici and Voiculescu [19] that there are positive numbers η and M such that F_μ has a right inverse F_μ^{-1} defined on the region $\Gamma_{\eta,M} := \{z \in \mathbb{C}^+; |Re(z)| < \eta Im(z)\}$.

The Voiculescu transform of μ is defined by $\phi_\mu(z) = F_\mu^{-1}(z) - z$, on any region of the form $\Gamma_{\eta,M}$ where F_μ^{-1} is defined.

The *free additive convolution* of two probability measures $\mu_1, \mu_2 \in \mathcal{M}$ is the probability measure $\mu_1 \boxplus \mu_2$ on $\mathbb{R}$ such that

$$\phi_{\mu_1 \boxplus \mu_2}(z) = \phi_{\mu_1}(z) + \phi_{\mu_2}(z), \quad \text{for } z \in \Gamma_{\eta_1,M_1} \cap \Gamma_{\eta_2,M_2}.$$

Free additive convolution corresponds to the sum of free random variables: $\mu_a \boxplus \mu_b = \mu_{a+b}$, for a and b free random variables. The free cumulants [54] are the

coefficients $\kappa_n = \kappa_n(\mu)$ in the series expansion (whenever it exists)

$$\phi_\mu(z) = \sum_{n=1}^{\infty} \kappa_n z^{1-n}. \tag{2.6}$$

The *free multiplicative convolution* $\mu_1 \boxtimes \mu_2$ of probability measures $\mu_1, \mu_2 \in \mathcal{M}$, one of them in $\mathcal{M}^+$, say $\mu_1 \in \mathcal{M}^+$, is defined as the distribution of $\mu_{X_1^{1/2} X_2 X_1^{1/2}}$ where $X_1 \geq 0, X_2$ are free, self-adjoint elements such that $\mu_{X_i} = \mu_i$, (see [19].)

The *Boolean convolution* [58] of two probability measures $\mu_1, \mu_2 \in \mathcal{M}$ is defined as the probability measure $\mu_1 \uplus \mu_2$ on $\mathbb{R}$ such the transform $K_\mu(z) = z - F_\mu(z)$, (usually called *self-energy*), satisfies

$$K_{\mu_1 \uplus \mu_2}(z) = K_{\mu_1}(z) + K_{\mu_2}(z), \; z \in \mathbb{C}^+.$$

Boolean convolution corresponds to the sum of Boolean-independent random variables. Boolean cumulants are defined as the coefficients $r_n = r_n(\mu)$ in the series (whenever it exists)

$$K_\mu(z) = \sum_{n=1}^{\infty} r_n z^{1-n}. \tag{2.7}$$

The monotone convolution was defined in [44] and extended to unbounded measures in [28]. The *monotone convolution* of two probability measures $\mu_1, \mu_2 \in \mathcal{M}$ is defined as the probability measure $\mu_1 \rhd \mu_2$ on $\mathbb{R}$ such that

$$F_{\mu_1 \rhd \mu_2}(z) = F_{\mu_1}(F_{\mu_2}(z)), \; z \in \mathbb{C}^+.$$

Monotone convolution corresponds to the sum of monotone independent random variables. Recently, Hasebe and Saigo [33] defined a notion of monotone cumulants $(h_n)_{n \geq 1}$ which satisfy that $h_n(\mu^{\rhd k}) = k h_n(\mu)$.

(2.3) Combinatorial Theory of Cumulants. Speicher's combinatorial treatment of free probability [54] gives a nice description of free convolutions using cumulants. These ideas were extended by Speicher and Wourodi [58] for the Boolean case and Hasebe and Saigo [33] for the monotone case.

A *partition* π is an equivalence relation on the set $[n] := \{1, 2, \ldots, n\}$, which is decomposed into equivalence classes $V_1, \ldots, V_r$, called blocks. We write $\pi = \{V_1, \ldots, V_r\}$ and $a \sim_\pi b$ for $a, b \in V_i$.

A partition π is called *non-crossing* iff $a \sim_\pi c, b \sim_\pi d \Rightarrow b \sim_\pi c$ for all $1 \leq a < b < c < d \leq n$.

A non-crossing partition is called *interval* partition if, $a \sim_\pi b \Rightarrow a \sim_\pi b - 1$ for all $1 \leq a < b \leq n$. We denote by $\mathcal{P}(n) \supset NC(n) \supset \mathcal{I}(n)$ the sets of partitions, non-crossing partitions and interval partitions of $[n]$, respectively.

A *monotone* partition (π, λ) is a non-crossing partition, together with a linear order λ of the blocks $\{V_1, \ldots, V_r\}$ of π, which respects the nesting structure of π. More precisely, if $1 \leq a < b < c \leq n$ are such that $a, c \in V_i$, $b \in V_j$ and $i \neq j$, then $\lambda(V_j) < \lambda(V_j)$.

For a random variable $a \in \mathcal{A}$, its classical, free, Boolean and monotone *cumulants* $(c_n^a)_{n \geq 1}$, $(\kappa_n^a)_{n \geq 1}$, $(r_n^a)_{n \geq 1}$, $(h_n^a)_{n \geq 1}$, satisfy the moment-cumulant formulas

$$\tau(a^n) = \sum_{\pi \in \mathcal{P}(n)} c_\pi^a = \sum_{\pi \in NC(n)} \kappa_\pi^a = \sum_{\pi \in \mathcal{I}(n)} r_\pi^a = \sum_{(\pi, \lambda) \in \mathcal{M}(n)} \frac{h_\pi^a}{|\pi|!}, \tag{2.8}$$

where, for a sequence of complex numbers $(f_n)_{n\geq 1}$ and a partition $\pi = \{V_1, \ldots, V_i\}$, we define $f_\pi := f_{|V_1|} \cdots f_{|V_i|}$, where $|\pi|$ is the number of blocks of the partition π.

3. Central Limit Theorems

For a classical random variable X with all moments, mean 0 and variance 1, let us denote by $S_n^*(X) = (X_1 + X_2 + \cdots + X_n)/\sqrt{n}$ the normalized sum of n independent copies of X. The so-called Central Limit Theorem states that $S_n^*(X)$ converges, as $n \to \infty$, to the standard normal distribution $\mathcal{N}(0,1)$. Similar results hold for other notions of independence.

We briefly sketch the standard proof of these central limit theorems. The method is based on showing the convergence of *all* the moments. For each additive convolution $\circledast \in \{*, \boxplus, \uplus, \rhd\}$, let $S_n^{\circledast}(X)$ be the normalized sum of n $\circledast$-independent copies of X. We fix n and expand the k-th moment of $S_n^{\circledast}(X)$:

$$\tau((S_n^{\circledast}(X))^k) = n^{-k/2} \sum_{i_1,\ldots,i_k=1}^{n} \tau(X_{i_1}, \ldots, X_{i_k}). \tag{3.1}$$

Since all X_i have the same distribution, $\tau(X_{i_1}, \ldots, X_{i_k})$ will depend only on which indices are equal and which are not (and on the order of the involved indices for $\circledast = \rhd$). It is useful to associate a tuple $i_1, \ldots, i_k$ with the partition π defined by $r \sim_\pi s \Leftrightarrow i_r = i_s$. For $\circledast = \rhd$, we consider the linear order λ of the blocks of π which is induced by the order of the indices. We write $\tau_\pi(X) := \tau(X_{i_1}, \ldots, X_{i_k})$ for $\circledast \in \{*, \uplus, \boxplus\}$ and $\tau_{(\pi,\lambda)}(X) := \tau(X_{i_1}, \ldots, X_{i_k})$ for $\circledast = \rhd$. We sum first over all tuples $i_1, \ldots, i_k$ with the same associated partition, obtaining, for $\circledast \in \{*, \uplus, \boxplus\}$:

$$\tau((S_n^{\circledast}(X))^k) = \frac{1}{n^{k/2}} \sum_{\pi \in \mathcal{P}(k)} \sum_{\substack{1 \leq i_1, \ldots, i_k \leq n \\ \{i_1, \ldots, i_k\} \sim \pi}} \tau_\pi(X) = \sum_{\pi \in \mathcal{P}(k)} \frac{\tau_\pi(X) n!}{(n - |\pi|)! n^{k/2}},$$

and similarly

$$\tau((S_n^{\rhd}(X))^k) = \sum_{(\pi,\lambda) \in \mathcal{P}_\leq(k)} \frac{\tau_{(\pi,\lambda)}(X) n!}{|\pi|! (n - |\pi|)! n^{k/2}}.$$

Since $\tau(X) = 0$ it is possible to deduce from the factorization rules of each independence, that $\tau_\pi(X) = \tau_{(\pi,\lambda)}(X) = 0$ whenever π has as block of size 1. On the other hand, the factor $n^{-k/2}$ forces that at least $k/2$ different indices should appear for the contribution of π (or (π, λ)) not to vanish as $n \to \infty$.

By combining these two conditions, we only need to consider pairings. Depending on the notion of independence, additional conditions on the blocks of π (or (π, λ)) have to be fulfilled for $\tau_\pi(X)$ or $\tau_{(\pi,\lambda)}(X)$ to be non-zero.

For $\circledast = *$ no further condition needs to be imposed. In the Boolean case the repeated indices need to be consecutive and hence only the interval pairing counts. For free random variables one can show that only non-crossing pairings matter. For the monotone case we get only the monotone pairings. Since $\tau(X^2) = 1$, we have, for all the non-vanishing pairings π (or ordered pairings (π, λ), resp.), that actually $\tau_\pi(X) = 1$ ($\tau_{(\pi,\lambda)}(X) = 1$, resp.).

After enumerating the valid partitions and noticing that $n!(n^{k/2}(n - |\pi|)!)^{-1} \to 1$ as $n \to \infty$, we obtain that all the odd moments vanish asymptotically and the

asymptotics of the even moments are given by

$$\tau((S_n^*(X))^{2k}) \to \frac{(2k)!}{(2^k)k!}, \qquad \tau((S_n^{\uplus}(X))^{2k}) \to 1,$$
$$\tau((S_n^{\boxplus}(X))^{2k}) \to \frac{(2k)!}{(k!)^2(k+1)}, \qquad \tau((S_n^{\triangleright}(X))^{2k}) \to \frac{(2k)!}{(k!)^2 2^k},$$

which we can recognize as the moments of known probability measures.

Thus, the free central limit theorem [24, 61] states that the normalized sum of free copies of X converges weakly to the standard semicircle distribution $\mathbf{w}$, with density

$$\mathbf{w}(dx) = \frac{1}{2\pi}\sqrt{4-x^2}, \quad x \in [-2,2].$$

Similarly, the distributions appearing in the Boolean [58] and monotone [44] CLT's are, respectively, the symmetric Bernoulli distribution, $\mathbf{b} := 1/2\delta_{-1} + 1/2\delta_1$, and the arcsine distribution $\mathbf{a}$, with density

$$\frac{1}{\pi\sqrt{2-x^2}} \quad x \in [-\sqrt{2}, \sqrt{2}].$$

Once we established these fundamental limit theorems, we would like to point out generalizations in various directions.

(3.1) Operator-valued Central Limit Theorems. All the statements for the different notions of independence that we have presented so far seem to be parallel realms of classical probability. However, there are operator-valued generalizations of free, Boolean and monotone independence with their corresponding operator-valued central limit theorems [56, 63]. This branch of non-commutative probability is completely novel, as no such operator-valued generalization of classical independence exists. Operator-valued free probability has found important applications to random matrix theory and wireless communications [17, 34, 52, 53, 57].

(3.2) Berry-Essen Estimates. In classical probability, the Berry-Essen theorem gives a quantitative bound on the speed of convergence to the central limit: If X is a centered random variable with variance 1, then the distance between $S_n^*(X)$ and a standard normal random variable Y can be estimated in terms of the Kolmogorov distance Δ by

$$\Delta(S_n^*(X), Y) \leq C\frac{|\tau(X^3)|}{\sqrt{n}},$$

where C is a fixed constant.

A free analogue of the Berry-Esseen theorem was obtained by Chistyakov and Götze [27]: If X is a centered random variable with variance 1 and finite fourth moment, then the Kolmogorov distance between $S_n^{\boxplus}(X)$ and a standard semicircular random variable s is bounded by

$$\Delta(S_n^{\boxplus}(X), s) \leq c\frac{|\tau(X^3)| + \sqrt{\tau(X^4)}}{\sqrt{n}},$$

where c is a fixed constant. A weaker version of this result was obtained independently by Kargin [36] under assumption on X being bounded.

Recently, Mai and Speicher [42] obtained Berry-Essen estimates for multivariate and operator-valued versions of the free central limit theorem.

The Berry Esseen type estimates are still open for the Boolean and monotone case. It is clear for the Boolean case that one should use a different distance: since the limiting measure is atomic, its Kolmogorov distance to any absolutely continuous measure will be 1/4. For the monotone case, one can construct examples where the order of convergence $n^{-1/2}$ is not achieved.

(3.3) Superconvergence to the Central Limit. A qualitative generalization of the free central limit theorem was proved by Bercovici and Voiculescu [20]. It turns out that the convergence to the free central limit is of a much stronger nature than the classical one. They referred to this phenomenon as *superconvergence*.

THEOREM (3.2) ([20]). *Let X be a bounded random variable, and let μ_k be the distribution of $S_k^{\boxplus}(X)$.*

1. *There is some $N > 0$ such that, for all $k \geq N$, the distribution μ_k is absolutely continuous w.r.t. the Lebesgue measure.*
2. *For $k \geq N$, the densities $d\mu_k/dx$ converge uniformly on $\mathbb{R}$ to the density of the semicircle law.*
3. *If a_k and b_k are respectively, the lower and upper edges of the support of μ_k, then $a_k \to -2$ and $b_k \to 2$ as $k \to \infty$.*

Later, Wang [67] proved that (1) and (2) also hold when μ_k has unbounded support and, furthermore, he showed the L^p-convergence to the semicircular density for $p > 1/2$.

The property (3) from Theorem (3.2) also holds for the Boolean and monotone CLT's with rate of convergence $n^{-1/2}$. However, one can find examples where (1) and (2) fail.

THEOREM (3.3). [11] *Let X be bounded, centered, with variance* 1.

1. *Let $[a_k, b_k]$ be the support of the random variable $S_k^{\uplus}(X)$. Then $a_k \to -1$, $b_k \to 1$ as $k \to \infty$. Moreover, $\|S_k^{\uplus}(X)\| < 1 + \|X\|/\sqrt{k}$.*
2. *Let $[a_k, b_k]$ be the support of the random variable $S_k^{\rhd}(X)$. Then $a_k \to -\sqrt{2}$, $b_k \to \sqrt{2}$ as $k \to \infty$. Moreover, $\|S_k^{\rhd}(X)\| < \sqrt{2} + 2\|X\|/\sqrt{k}$.*

4. Infinite Divisibility

Let $\mu \in \in \mathcal{M}$ and denote by $\circledast$ one of the additive convolutions of probability measures on $\mathbb{R}$, namely, $\circledast \in \{*, \uplus, \boxplus, \rhd\}$. We say that μ is $\circledast$-infinitely divisible if, for all $n \in \mathbb{N}$, there exists a probability measure μ_n such that

$$\mu = \underbrace{\mu_n \circledast \mu_n \circledast \cdots \circledast \mu_n}_{n \text{ times}}.$$

We will denote by $ID(\circledast)$ the set of $\circledast$-infinitely divisible measures.

Just as in the classical case, infinitely divisible distributions play a very important role in the context of limit theorems in non-commutative probability. For example, it is easily seen that the Central Limits of Section 2 belong to $ID(\circledast)$. The following theorem (due to Lévy and Khintchine [37, 41] for the classical case, Bercovici and Pata [18] for the free and Boolean cases, and Anshelevich and Williams [2] for the monotone one) shows how the sums of independent identically

distributed non-commutative random variables give rise to infinitely divisible distributions.

THEOREM (4.1). *Let* $\circledast \in \{*, \uplus, \boxplus, \rhd\}$. *A probability measure* μ *is* $\circledast$*-infinitely divisible if and only if the is a sequence of probability measures* $\{\mu_n\}_{n\in N}$*, and a sequence of positive integers* $k_1 < k_2 < \cdots$ *such that the sequence*

$$\underbrace{\mu_n \circledast \mu_n \circledast \cdots \circledast \mu_n}_{k_n} \to \mu,$$

in distribution.

We shall point out that the more general theorem for triangular infinitesimal arrays where the measures are not assumed to be identically distributed is also true for free and Boolean convolutions. However, this is not true for the monotone convolution.

(4.1) Lévy Khintchine Representations. Recall that a probability measure μ is infinitely divisible in the classical sense if and only if its classical cumulant transform $\log\widehat{\mu}$ has the Lévy-Khintchine representation

$$\log\widehat{\mu}(u) = i\gamma u - \int_{\mathbb{R}} (e^{iut} - 1 - \frac{iut}{1+t^2})\frac{1+t^2}{t^2}\sigma(dt), \quad u \in \mathbb{R}. \tag{4.2}$$

where $\gamma \in \mathbb{R}$ and σ is a fine measure. If this representation exists, the pair (γ, σ) is determined in a unique way and is called the (classical) generating pair of μ. In this case we denote μ by $\rho_*^{\gamma,\sigma}$

Bercovici and Voiculescu [19] proved that the Voiculescu transform admits an analogous Lévy-Khintchine representation: a probability measure μ is $\boxplus$-infinitely divisible if and only if there exists a finite measure σ on $\mathbb{R}$ and a real constant γ such that

$$\phi_\mu(z) = \gamma + \int_{\mathbb{R}} \frac{1+tz}{z-t}\sigma(dt), \qquad z \in \mathbb{C}^+. \tag{4.3}$$

The pair (γ, σ) is called the *free* generating pair of μ and we denote μ by $\rho_{\boxplus}^{\gamma,\sigma}$.

For the Boolean case things are simpler. As shown by Speicher and Wourodi [58], any probability measure is $\uplus$-infinitely divisible and there is also a Boolean Lévy-Khintchine representation. Indeed, it follows by the Nevanlinna-Pick theory that for any probability measure μ there exists a real constant γ and a finite measure σ on $\mathbb{R}$, such that

$$K_\mu(z) = \gamma + \int_{\mathbb{R}} \frac{1+tz}{z-t}\sigma(dt), \qquad z \in \mathbb{C}^+. \tag{4.4}$$

The pair (γ, σ) is called the *Boolean* generating pair of μ and we denote μ by $\rho_{\uplus}^{\gamma,\sigma}$.

A characterization of $\rhd$-infinitely divisible measures was done by Muraki [46] and Belinschi [13]. A probability measure μ belongs to $ID(\rhd)$ if and only there exists a composition semigroup of reciprocal Cauchy transforms $F_{s+t} = F_s \circ F_t = F_t \circ F_s$ and $F_1 = F_\mu$. In this case the map $t \mapsto F_t(z)$ is differentiable for each fixed z in $\mathbb{R}$ and we define the mapping A_μ on $\mathbb{C}^+$ by

$$A_\mu(z) = \left.\frac{dF_t(z)}{dt}\right|_{t=0}, \qquad z \in \mathbb{C}^+.$$

For mappings of this form there exists $\gamma \in \mathbb{R}$ and a finite measure σ, such that

$$A_\mu(z) = -\gamma - \int_{\mathbb{R}} \frac{1+tz}{z-t}\,\sigma(dt). \tag{4.5}$$

This is the Lévy-Khintchine formula for monotone convolution and in this case we denote μ by $\rho_{\triangleright}^{\gamma,\sigma}$. The monotone cumulants h_n are the coefficients in the series

$$-A_\mu(z) = \sum_{n=1}^{\infty} h_n z^{1-n}. \tag{4.6}$$

(4.2) Cumulant Criteria. For measures with compact support (or, more generally, for measures which are determined by moments), working with cumulants turns out to be very useful to rule out measures which are not infinitely divisible.

The main criteria is the conditionally positive definiteness of cumulants for infinitely divisible measures, similar to the classical case. Let μ be a measure and denote by $\{c_n\}_{n\geq 1}, \{\kappa_n\}_{n\geq 1}, \{r_n\}_{n\geq 1}, \{h_n\}_{n\geq 1}$, respectively, its corresponding sequences of classical, free, Boolean and monotone cumulants.

Recall that a sequence $\{a_n\}_{n\geq 1}$ is conditionally positive definite if for every $n \geq 1$ and $\alpha_i \in \mathbb{C}, i = 1,\ldots,n$

$$\sum_{i,j=1}^{n} \alpha_i \overline{\alpha_j} a_{i+j} \geq 0.$$

THEOREM (4.7). *Let $\circledast \in \{*, \boxplus, \uplus, \triangleright\}$ be any of the additive convolutions. Let μ be a probability measure on $\mathbb{R}$ determined by moments. Then μ is $\circledast$-infinitely divisible if and only if the sequence of cumulants with respect to the convolution $\circledast$ is conditionally positive definite.*

(4.3) Bercovici-Pata Bijections. As mentioned at the beginning of this section, a correspondence between limit distributions arising from classical, Boolean, free and monotone convolutions is due to the works of Bercovici and Pata [18] and Anshelevich and Williams [2].

THEOREM (4.8) ([2, 18]). *Fix a finite positive Borel measure σ on $\mathbb{R}$, a real number γ, a sequence of probability measures $\{\mu_n\}_{n\in N}$, and a sequence of positive integers $k_1 < k_2 < \cdots$ The following assertions are equivalent:*

1. *The sequence $\underbrace{\mu_n * \mu_n * \cdots * \mu_n}_{k_n}$ converges weakly to $\rho_*^{\gamma,\sigma}$;*
2. *The sequence $\underbrace{\mu_n \boxplus \mu_n \boxplus \cdots \boxplus \mu_n}_{k_n}$ converges weakly to $\rho_{\boxplus}^{\gamma,\sigma}$;*
3. *The sequence $\underbrace{\mu_n \uplus \mu_n \uplus \cdots \uplus \mu_n}_{k_n}$ converges weakly to $\rho_{\uplus}^{\gamma,\sigma}$;*
4. *The sequence $\underbrace{\mu_n \triangleright \mu_n \triangleright \cdots \triangleright \mu_n}_{k_n}$ converges weakly to $\rho_{\triangleright}^{\gamma,\sigma}$;*
5. *The measures*

$$k_n \frac{x^2}{x^2+1} d\mu_n(x) \to \sigma$$

weakly, and

$$\lim_{n\uparrow\infty} k_n \int_{\mathbb{R}} \frac{x}{x^2+1} d\mu_n(x) = \gamma.$$

The last theorem gives a correspondence between the limit measures known as the **Bercovici-Pata bijection**. This can be stated concretely from the Lévy-Khintchine representations of the various infinitely divisible measures. Indeed, the different Lévy pairs are obtained by the limits in (5) above.

Definition (4.9).

1. The (classical-to-free) Bercovici-Pata bijection $\Lambda : ID(*) \to ID(\boxplus)$ is defined by the application $\rho_*^{\gamma,\sigma} \mapsto \rho_{\boxplus}^{\gamma,\sigma}$.
2. The (Boolean-to-free) Bercovici-Pata bijection $\mathbb{B} : \mathcal{M} \to ID(\boxplus)$ is defined by the application $\rho_{\uplus}^{\gamma,\sigma} \mapsto \rho_{\boxplus}^{\gamma,\sigma}$.
3. The (classical-to-monotone) Bercovici-Pata bijection $\Lambda^{\triangleright} : ID(*) \to ID(\triangleright)$ is defined by the application $\rho_*^{\gamma,\sigma} \mapsto \rho_{\triangleright}^{\gamma,\sigma}$.

The weak continuity of Λ and Λ^{-1} was proved in [12]. The weak continuity of $\mathbb{B}$ and $\mathbb{B}^{-1}$ follows from the continuity of the free and Boolean convolution powers since $\mathbb{B}(\mu) = (\mu^{\boxplus 2})^{\uplus 1/2}$. Finally the weak continuity of $\Lambda^{\triangleright}$ was proved in Hasebe [31]. In summary, the arrows the following commutative diagram are weakly continuous.

$$\begin{array}{ccccc} & & ID(\uplus) = \mathcal{M} & & \\ & \swarrow^{\mathbb{B}} & \downarrow & \searrow & \\ ID(\boxplus) & \xrightarrow{\Lambda^{-1}} & ID(*) & \xrightarrow{\Lambda^{\triangleright}} & ID(\triangleright) \end{array}$$

REMARK (4.10). *If follows from the Lévy-Khintchine representations that the Boolean cumulants of μ are free cumulants of its image under the Boolean Bercovici-Pata bijection $\mathbb{B}$, namely, $r_n(\mu) = k_n(\mathbb{B}(\mu))$. Similarly, $c_n(\mu) = k_n(\Lambda(\mu))$ and $c_n(\mu) = h_n(\Lambda^{\triangleright}(\mu))$.*

(4.4) Convergence of the 4th-moment. In a seminal paper, Nualart and Peccati [50] proved a convergence criterion for normalized multiple integrals in a fixed chaos with respect to the classical Brownian motion to the normal distribution. More precisely, let $(W_t)_{t>0}$ be a standard Brownian motion. For every square-integrable function f on $\mathbb{R}_+^m$ we denote by $I_m^W(f)$ the m-th multiple Wiener-Ito stochastic integral of f with respect to W.

THEOREM (4.11) ([50]). *Let $\{X_n = I_m^W(f_n)\}_{n>0}$ be a sequence of multiple Wiener-Ito integrals in a fixed m-chaos. Then the following are equivalent*

1. $E[X_n^4] \to 3$,
2. $\mu_{X_n} \to \mathcal{N}(0,1)$.

As pointed out in the previous section, the standard proof for the convergence to any of these "Gaussian" distributions consists of showing the convergence of *all* moments. Thus, Theorem (4.11), gives a drastic simplification for the moment method for the case of multiple integrals.

Recently, it was proved by Kemp et al. [39] that the Nualart-Peccati 4th moment criterion holds also for the free Brownian motion $(S_t)_{t>0}$ and its multiple Wigner integrals $I_m^S(f)$.

Motivated by these results, Arizmendi [3] showed that, when restricted to infinitely divisible measures, convergence to the "Gaussian" distributions can be ensured by the convergence of the second and fourth moments.

THEOREM (4.12).

1. *Let* $\{\mu_n = \mu_{X_n}\}_{n>0}$ *be a sequence of probability measures with variance* 1 *and mean zero such that* $\mu_n \in ID(*)$ *and suppose that* $E[X_n^4] \to 3$ *then* $\mu_{X_n} \to \mathcal{N}(0,1)$.
2. *Let* $\{\mu_n = \mu_{X_n}\}_{n>0}$ *be a sequence of probability measures with variance* 1 *and mean zero such that* $\mu_n \in ID(\boxplus)$ *and suppose that* $E[X_n^4] \to 2$. *then* $\mu_{X_n} \to \mathbf{w}$.
3. *Let* $\{\mu_n = \mu_{X_n}\}_{n>0}$*be a sequence of probability measures with variance* 1 *and mean zero such that* $\mu_n \in ID(\triangleright)$ *and suppose that* $E[X_n^4] \to 3/2$, *then* $\mu_{X_n} \to \mathbf{a}$.

The proof is based in the following simple lemma (which is the Boolean version of the theorem above) and on the continuity of Bercovici-Pata bijections.

LEMMA (4.13). *Let* X_n *be random variables with variance* 1 *and mean* 0, *if* $E(X_n^4) \to 1$ *then* $\mu_{X_n} \to \mathbf{b}$.

5. Some Recent Results in Free Infinite Divisibility

Freeness and free infinite divisibility is, by far, much more explored than their Boolean and monotone counterparts. In this section we want to describe recent results and examples of freely infinitely divisible (FID, for short) distributions. The selection of the topics, strongly motivated by the interests of the authors, is focused on results which have no analog in the classical world. In particular, the free counterparts of important classes of classical infinitely divisible distributions, such as the compound Poisson, stable, self-decomposable, Type G, Generalized Gamma Convolutions and Meixner laws, obtained by the Bercovici-Pata bijection, will not be treated here. For more information on such classes, the interested reader is advised to look at the original work [18], but also [12], where detailed properties of the Bercovici-Pata bijection are given, and [4, 51, 60], among others.

(5.1) Polynomials in free random variables and free subordinators. An important subclass of $ID(*)$ is the class of infinitely divisible measures with support on $\mathbb{R}_+$. This class plays an important role in the theory of stochastic processes since it corresponds to the one dimensional marginal distributions of real-valued Lévy processes with non-decreasing sample paths, known as *subordinators*. Any classically infinitely divisible distribution supported on $\mathbb{R}_+$ satisfies that μ^{*t} is concentrated on $\mathbb{R}_+$ for all time $t > 0$. In contrast, we can easily find an FID distribution with support on $\mathbb{R}_+$ such that $\mu^{\boxplus t}$ is not supported on $\mathbb{R}_+$ for all times $t > 0$ (e.g. the semicircle distribution with mean 2 and variance 1). This observation lead to consider the class of free regular measures. We say that a measure $\mu \in ID(\boxplus)$ is *free regular* if $\mu^{\boxplus t}$ is supported on $\mathbb{R}_+$ for all $t \geq 0$. A measure μ is free regular if and only if $\mu = \Lambda(\nu)$ for some $\nu \in ID(*) \cap \mathcal{M}^+$. This class has very nice closure properties.

THEOREM (5.1) ([9]). *Let* μ, ν *be free regular measures and let* σ *be FID. Then the following properties hold.*

1. $\mu \boxtimes \nu$ *is free regular.*
2. $\mu \boxtimes \sigma$ *is FID.*

Furthermore, the class of FID behaves well under squaring.

THEOREM (5.2) ([9]). *Let X be a self-adjoint even element (i.e. μ_X is symmetric). If X is FID, then X^2 is free regular (FID, in particular).*

We note here that there is no classical analog of the previous theorems see [9]. The free commutator [49] also preserves FID.

THEOREM (5.3) ([9]). *Let X and Y be free and self-adjoint elements, and suppose that $\mu_1 := \mu_X$ and $\mu_2 := \mu_Y$ are FID. Then the distribution of the free commutator $\mu_1 \square \mu_2 := \mu_{i(XY-YX)}$ is also FID.*

We have seen that if a_1, a_2, a_3 are free even FID random variables, then $i(a_i a_j - a_j a_i)$, $a_i a_j + a_j a_i$, a_i^2 and $a_j a_i^2 a_j$ are also FID. Combining these results one can easily see that the following polynomials are FID: $a_1^2 + a_2^2 + a_2 a_1 + a_2 a_1$, $i(a_1 a_2^2 - a_2 a_2^2)$, $a_1^4 + a_2^4 - a_2^2 a_1^2 - a_2^2 a_1^2$, $a_1 a_2^2 a_1 + a_2 a_1^2 a_2 + a_1 a_2 a_1 a_2 + a_2 a_1 a_2 a_1$, $a_1 a_2^2 a_1 + a_2 a_1^2 a_2 - a_1 a_2 a_1 a_2 - a_2 a_1 a_2 a_1$, $a_1 a_2 a_3 + a_2 a_1 a_3 + a_3 a_1 a_2 + a_3 a_2 a_1$, etc. Therefore, it is natural to ask which polynomials preserve FID.

(5.2) Free infinite divisibility of Gaussian distribution. In [10], the authors considered the free infinite divisibility (also FID, for short) of the ultraspherical family $\mathbf{u}_n$ with density

$$c_n (4 - t^2)^{n-1/2} \mathbf{1}_{[-2,2]}.$$

where c_n is a normalizing constant.

This family contains all the Gaussian distributions with respect to the 5 fundamental notions of independence in non-commutative probability (the normal law $\mathbf{g}$ appears in the limit $n \to \infty$ if $\mathbf{u}_n$, properly normalized).

Arizmendi and Perez-Abreu [10] proved using kurtosis arguments that $\mathbf{u}_n$ is not FID for $n < 1$ and from cumulant criteria conjectured that $\mathbf{u}_n$ is FID for all $n \in [1, +\infty)$.

Since the class $ID(\boxplus)$ is closed under weak convergence, one important consequence of the FID of this family is that the normal distribution $\mathbf{g}$ is *both* classically and freely infinitely divisible.

Following these remarks and considerations of general Brownian motions, the FID of the normal distribution $\mathbf{g}$ was proved by Belinschi et al. [14]. However, they used another family to approximate the normal law.

THEOREM (5.4). [14] *Let μ_c denote the Askey-Wimp-Kerov distribution with density $\mu_c(dx) = \dfrac{1}{2\pi\Gamma(c+1)} \dfrac{1}{|D_{-c}(ix)|} 1_{\mathbb{R}}(x)dx$, where the function $D_{-c}(z)$ for $c > 0$ is defined by*

$$D_{-c}(z) = \frac{e^{-\frac{z^2}{4}}}{\Gamma(c)} \int_0^\infty x^{c-1} e^{-zx - \frac{x^2}{2}} dx,$$

and $D_{-c}(z)$ for $c \in (-1, 0)$ is the analytic continuation of D_{-c} for $c > 0$. For any $c \in (-1, 0]$ the measure μ_c is $\boxplus$-infinitely divisible. In particular, the normal law μ_0 belongs to $ID(\boxplus) \cap ID(*)$.

The normal distribution plays no important role in free probability and thus there is a priori no reason for this distribution to be FID.

The importance of the paper [14] does not only come from proving this surprising result, which was a conjecture of one of the present authors [10]. On one hand, this result has lead to consider the question of how big is the class $ID(\boxplus) \cap ID(*)$.

On the other hand, the techniques used in [14] have been used and further developed [1, 5, 6, 7, 25, 32] and lead to consider the class $\mathcal{UI}$ firstly introduced in Arizmendi and Hasebe [7]. This class has proved to be very useful in proving the $\boxplus$-infinite divisibility of a distribution and will be described in the next part 5.4.

The FID of $\mathbf{u}_n$ was proved for $n \in \mathbb{N}$ by Arizmendi and Belinschi [5] and for general $n \in [1, +\infty)$ in a recent paper by Hasebe [32]. This gives another proof of the fact that $\mathbf{g}$ belongs to $ID(\boxplus)$.

Let us finally mention that other FID families which approximate the Gaussian distribution. The proofs of the FID of these families rely on the ideas of [14].

EXAMPLE (5.5). *The following are families of distributions in $ID(\boxplus)$ which suitably scaled approximate the Gaussian distribution.*

1. [1] *The q-Gaussian distribution*

$$\mathbf{g}_q(dx) = \frac{\sqrt{1-q}}{\pi} \sin\theta(x) \prod_{n=1}^{\infty} (1-q^n)|1-q^n e^{2i\theta(x)}|^2 \, 1_{\left[-\frac{2}{\sqrt{1-q}}, \frac{2}{\sqrt{1-q}}\right]}(x)dx$$

   for $q \in [0,1)$, where $\theta(x)$ is the solution of $x = \frac{2}{\sqrt{1-q}} \cos\theta$, $\theta \in [0,\pi]$.

2. [32] *The Student distribution*

$$\mathbf{t}_n(dx) = \frac{1}{B(\frac{1}{2}, n-\frac{1}{2})} \frac{1}{(1+x^2)^n} 1_{\mathbb{R}}(x)dx, \; n = 1,2,3,\cdots.$$

3. [32] *For $p > 3/2$, the gamma distribution*

$$\gamma_p(dx) = \frac{1}{\Gamma(p)} x^{p-1} e^{-x} 1_{[0,\infty)}(x)dx.$$

(5.3) Distributions which are clasically and freely infinitely divisible. From the Bercovici-Pata bijection one may have the idea that there are two parallel classes: $ID(\boxplus)$ and $ID(*)$ and in principle, apart from trivial examples and fixed points there is no reason why a distribution may belong to both classes.

We have seen already that the normal distribution belongs to $ID(\boxplus) \cap ID(*)$. Recent works [6, 9, 25, 32, 51] have provided more examples of probability measures in this class.

EXAMPLE (5.6) (Measures which are in $ID(*) \cap ID(\boxplus)$).

1. *The Cauchy distribution*

$$\mathbf{c}(dx) = \frac{1}{\pi(1+x^2)} 1_{\mathbb{R}}(x)dx.$$

2. [9] *The chi-square distribution* $\frac{1}{\sqrt{\pi x}} e^{-x} 1_{[0,\infty)}(x)dx$.
3. [32] *The Student distribution* $\mathbf{t}_n$, *n=1,2,3...*
4. [9, 32] *For $n = 1,2,3,\cdots$, the F-distribution with density*

$$f(x) := \frac{1}{B(1/2, n/2)} \frac{1}{(nx)^{1/2}} \left(1 + \frac{x}{n}\right)^{-(1+n)/2}, \quad x > 0.$$

5. [25] *For $0 < t \le \frac{1}{2}$, the symmetric Meixner distributions*

$$\rho_t(dx) := \frac{4^t}{2\pi\Gamma(2t)} |\Gamma(t+ix)|^2 dx, \; x \in \mathbb{R}.$$

6. [25] *The* logistic distribution

$$\mu_2(dx) = \frac{\pi}{2\cosh^2(\pi x)}\,dx,\ x \in \mathbb{R}.$$

7. [7] *The positive Boolean stable law with stability index* $\alpha \in (0, \frac{1}{2}]$

$$\frac{d\mathbf{b}_\alpha}{dx} = \frac{\frac{1}{\pi}\sin(\alpha\pi)x^{\alpha-1}}{x^{2\alpha} + 2\cos(\alpha\pi)x^\alpha + 1},\ x > 0.$$

8. [32] *For* $p \in (0, 1/2] \cup [3/2, \infty)$*, the gamma distribution* γ_p.

A still remaining question is whether there is a general theory on the intersection of free and classical infinite divisibility.

(5.4) The class $\mathcal{UI}$. We begin with an analytical characterization of FID distributions.

THEOREM (5.7). *A Borel probability measure* μ *on the real line is FID if and only if its Voiculescu transform* $\phi_\mu(z)$ *extends to an analytic function* $\phi_\mu : \mathbb{C}^+ \to \mathbb{C}^-$.

In practice, it is quite challenging to decide whether or not a distribution is FID. The class $\mathcal{UI}$ (univalent inverse reciprocal Cauchy transforms) was introduced in [7], following the ideas of [14].

Definition (5.8). A probability measure μ is said to be in class $\mathcal{UI}$ if F_μ is univalent in $\mathbb{C}^+$ and, moreover, F_μ^{-1} has an analytic continuation from $F_\mu(\mathbb{C}^+)$ to $\mathbb{C}^+$ as a univalent function.

The importance of this class is given by the following lemma (implicitly used in [14])

LEMMA (5.9). [7] *If* $\mu \in \mathcal{UI}$ *then* μ *is FID.*

It is easily seen that the class $\mathcal{UI}$ satisfies the following properties.

1. The class $\mathcal{UI}$ is closed with respect to the weak convergence.
2. The class $\mathcal{UI}$ is *not* closed under free convolution.
3. $ID(\boxplus)$ contains strictly $\mathcal{UI}$.

Although the class $\mathcal{UI}$ is a proper subset of $ID(\boxplus)$, most of the known FID distributions belong to $\mathcal{UI}$. In fact, up to now, all the FID distributions presented in the present section are $\mathcal{UI}$.

EXAMPLE (5.10). *The following probability measures belong to* $\mathcal{UI}$.

1. *Wigner's semicircle law*

$$\mathbf{w}(dx) = \frac{1}{2\pi}\sqrt{4 - x^2}\,1_{[-2,2]}(x)\,dx.$$

2. *The free Poisson law (or Marchenko-Pastur law)*

$$\mathbf{m}(dx) = \frac{1}{2\pi}\sqrt{\frac{4-x}{x}}\,1_{[0,4]}(x)\,dx.$$

3. [7, 32] *For* $\frac{1}{2} \le |a| < 1$*, the beta distribution* β_a *with*

$$\beta_a(dx) = \frac{\sin(\pi a)}{\pi a}\left(\frac{1-x}{x}\right)^a 1_{[0,1]}(x)\,dx,$$

for $\frac{1}{2} \le |a| < 1$. $\beta_{\frac{1}{2}}$ *is equal to* $\mathbf{m}$ *up to scaling.*

4. [6] *The Boolean stable law with stability index* α

$$d\mathbf{b}_\alpha^\rho(dx)=\begin{cases}\dfrac{\sin(\pi\rho\alpha)}{\pi}\dfrac{x^{\alpha-1}}{x^{2\alpha}+2x^\alpha\cos(\pi\rho\alpha)+1}dx, & x>0,\\[2ex] \dfrac{\sin(\pi(1-\rho)\alpha)}{\pi}\dfrac{|x|^{\alpha-1}}{|x|^{2\alpha}+2|x|^\alpha\cos(\pi(1-\rho)\alpha)+1}dx, & x<0,\end{cases}$$

for $0<\alpha\le\frac{1}{2}$, $\rho\in[0,1]$.

REMARK (5.11). *For* $\frac{1}{2}\le\alpha\le\frac{2}{3}$ *and* $2-\frac{1}{\alpha}\le\rho\le\frac{1}{\alpha}-1$, *the Boolean stable law* $\mathbf{b}_\alpha^\rho$ *is still freely infinitely divisible, but not in the class* $\mathcal{UI}$ [6].

(5.5) Free divisibility indicator. A remarkable property for the free convolution is the fact that $\mu^{\boxplus t}$ exists as a probability measure, for all $t\ge1$. This contrast with classical probability theory, because the usual convolution μ^{*t} is not necessarily defined even for $t\ge1$, unless μ is $*$-infinitely divisible. Hence, for any measure μ, the following quantity is of interest:

$$\widetilde{\phi}(\mu):=\inf\{t>0:\mu^{\boxplus t}\text{ exists as a probability measure}\}.$$

A probability measure μ is freely infinitely divisible if and only if $\widetilde{\phi}(\mu)=0$.

Belinschi and Nica [15] introduced the semigroup of homomorphisms

$$\mathbb{B}_t(\mu)=\left(\mu^{\boxplus(1+t)}\right)^{\uplus\frac{1}{1+t}}.$$

In addition to the semigroup property $\mathbb{B}_t(\mathbb{B}_s(\mu))=\mathbb{B}_{t+s}(\mu)$, the family $\mathbb{B}_t$ also satisfies:

1. $\mathbb{B}_t(\mu\boxtimes\nu)=\mathbb{B}_t(\mu)\boxtimes\mathbb{B}_t(\nu)$.
2. $\mathbb{B}_1=\mathbb{B}$.

Because of the semigroup property one can introduce the *free divisibility indicator*

$$\phi(\mu):=\sup\{t\ge0:\mu\in\mathbb{B}_t(\mathcal{M})\}.$$

By property (2), $\mathbb{B}_t$ reaches free infinite divisibility at time $t=1$, and $\phi(\mu)\ge1$ if and only if $\mu\in ID(\boxplus)$. Moreover, $\phi(\mu)=1-\widetilde{\phi}(\mu)$ if μ is not freely infinitely divisible. Thus, $\phi(\mu)$, rather than just testing free infinite divisibility, measures, in some sense, how divisible a measure is. The explicit calculation of this indicator is expected to be useful to understand the free convolution and free infinite divisibility.

Belinschi and Nica [15] calculated the divisibility indicator the free poisson ($\phi(\mathbf{m})=1$), the semicircle distribution ($\phi(\mathbf{w})=1$), arcsine distribution ($\phi(\mathbf{a})=1/2$) and Cauchy distribution ($\phi(\mathbf{t}_1)=\infty$). They also showed that for any purely atomic distribution the divisibility indicator equals 0.

More recently, the following relation was proved in Arizmendi and Hasebe [8]

$$\phi(\mu^{\uplus t})=\frac{\phi(\mu)}{t}\text{ for }t>0. \tag{5.12}$$

In particular, as conjectured by Bożejko [26], if μ is freely infinitely divisible then so is $\mu^{\uplus t}$. Furthermore, from (5.12) one may reformulate the definition of the free divisibility indicator as

$$\phi(\mu)=\sup\{t\ge0:\mu^{\uplus t}\text{ is freely infinitely divisible}\}.$$

This has been used by Hasebe in [32] to show that the divisibility indicator equals 1 for the following distributions: the Gaussian distribution, the Student distribution $\mathbf{t}_q$ for $q \in (1,2] \cup \bigcup_{n=1}^{\infty}[2n+1/4, 2n+2]$ and the ultraspherical distributions $\mathbf{u}_p$ for $p \in [1,\infty)$.

Received May 21, 2013

Final version received September 13, 2013

OCTAVIO ARIZMENDI
CIMAT,
MÉXICO
octavius@cimat.mx

CARLOS VARGAS
UNIVERSITÄT DES SAARLANDES,
FR 6.1–MATHEMATIK, POSTFACH 151150,
66041 SAARBRÜCKEN, GERMANY

REFERENCES

[1] M. ANSHELEVICH, S. T. BELINSCHI, M. BOŻEJKO, AND F. LEHNER, *Free infinite divisibility for q–Gaussians*, Math. Res. Lett. **17** no. 5, (2010), 905–916.
[2] M. ANSEHELEVICH AND J. D. WILLIAMS, *Limit theorems for monotonic convolution and the Chernoff product formula*, Int. Math. Res. Not. IMRN (2013), doi:10.1093/imrn/rnt018.
[3] O. ARIZMENDI, *Convergence of the fourth moment and infinite divisibility*, Probab. Math. Statist. **33** no. 2, (2013), 201–212.
[4] O. ARIZMENDI, O. E. BARNDORFF-NIELSEN, AND V. PÉREZ-ABREU, *On free and classical type G distributions*, Special number, Braz. J. Probab. Stat. **24**, (2010), 106–127.
[5] O. ARIZMENDI AND S. BELINSCHI, *Free Infinite Divisibility for Ultrasphericals*, Infin. Dimens. Anal. Quantum. Probab. Relat. Top. 16, 1350001 (2013) [11 pages].
[6] O. ARIZMENDI AND T. HASEBE, *On Boolean Stable laws*, To appear in Proc. Amer. Math. Soc.
[7] O. ARIZMENDI AND T. HASEBE, *On a class of explicit Cauchy-Stieltjes transforms related to monotone stable and free Poisson laws*, Bernoulli **19**, no. 5B, (2013), 2750–2767.
[8] O. ARIZMENDI AND T. HASEBE, *Semigroups related to additive and multiplicative, free and Boolean convolutions*, Studia Math. **215**, no. 2, (2013), 157–185.
[9] O. ARIZMENDI, T. HASEBE, AND N. SAKUMA, *On the Law of free Subodinators*, ALEA, Lat. Am. J. Probab. Math. Stat. **10**, no. 2, (2013), 271–291.
[10] O. ARIZMENDI AND V. PÉREZ-ABREU, *On the non-classical infinite divisibility of power semicircle distributions*, Communications on Stochastic Analysis **4**, (2010), 161–178. In honor of Gopinath Kallianpur.
[11] O. ARIZMENDI AND C. VARGAS, Supeconvergence in non-commutative Central Limit Theorems: Combinatorial Approach, preprint (2013).
[12] O. BARNDORFF-NIELSEN AND S. THORBJONSEN, *Classical and free infinite divisibility and Lévy processes*, Quantum independent increment processes. II, Lecture Notes in Math., 1866, *Springer, Berlin*, (2006), 33–159.
[13] S. T. BELINSCHI, *Complex analysis methods in noncommutative probability*, available in arXiv:math/0602343v1.
[14] S.T. BELINSCHI, M. BOZEJKO, F. LEHNER, AND R. SPEICHER, *The normal distribution is $\boxplus$-infinitely divisible*, Adv. Math. **226**, no. 4, (2011), 3677–3698.
[15] S. T. BELINSCHI AND A. NICA, *On a remarkable semigroup of homomorphisms with respect to free multiplicative convolution*, Indiana Univ. Math. J. **57**, (2008), 1679–1713.
[16] A. BEN GHORBAL AND M SCHÜRMAN, *Non-commutative notions of stochastic independence*, Mathematical Proceedings of the Cambridge Philosophical Society **133**, (2002), 531–561.
[17] F. BENAYCH-GEORGES, *Rectangular random matrices, related free entropy and free Fisher's information*, Journal of Operator Theory **62**, no. 2, (2009), 371–419.

[18] H. BERCOVICI AND V. PATA, *Stable laws and domains of attraction in free probability theory (with an appendix by Philippe Biane)*, Ann. of Math. (2) **149**, no. 3, (1999), 1023–1060.
[19] H. BERCOVICI AND D. VOICULESCU, *Free convolution of measures with unbounded supports*, Indiana Univ. Math. J. **42** (1993), 733–773.
[20] H. BERCOVICI AND D. VOICULESCU, *Superconvergence to the central limit and failure of the Cramér theorem for free variables*, Probab. Theory Relat. Fields **102**, (1995), 215–222.
[21] P. BIANE, *Representations of symmetric groups and free probability*, Advances in Mathematics **138**, (1998), 126–181.
[22] P. BIANE, *Processes with free increments*, Math. Z. **227**, (1998), 143–174.
[23] L. BONDESSON, *Generalized gamma convolutions and related classes of distributions and densities*, Lecture Notes in Statistics **76** Springer, Berlin, (1992).
[24] M. BOŻEJKO, *On $\Lambda(p)$ sets with minimal constant in discrete noncommutative groups*, Proc. Amer. Math. Soc. **51**, (1975), 407–412.
[25] M. BOZEJKO AND T. HASEBE, *On free infinite divisibility for classical Meixner distributions*, Probab. Math. Statist. **33**, no. 2, (2013), 363–375.
[26] M. BOŻEJKO AND J. WYSOCZAŃSKI, *Remarks on t-transformations of measures and convolutions*, Ann. Inst. Henri Poincaré-PR **37**, (2001), 737–761.
[27] G. CHISTYAKOV AND F. GÖTZE, *Limit theorems in free probability theory I*, Ann. Probab. **36**, no. 1, (2008), 54–90.
[28] U. FRANZ AND N. MURAKI, *Markov structure of monotone Lévy processes*, In H. Heyer, T. Hirai, T. Kawazoe, and K. Saito, editors, Infinite Dimensional Harmonic Analysis III, Proceedings of the Third German-Japanese Symposium, University of Tubingen, Germany, 15–20 September 2003, pages 37–57, 2005. math.PR/0401390.
[29] B. DE FINETTI, *Sulle funzioni a incremento aleatorio*, Rend Acai Lincei (6) **10**, (1929), 805–808, 865–868.
[30] W. FELLER, An Introduction to Probability Theory and its Applications, vol.II. John Wiley and Sons, 1971.
[31] T. HASEBE, *Monotone convolution semigroups*, Studia Math. **200**, no. 2, (2010), 175–199.
[32] T. HASEBE, *Free infinite divisibility for beta distributions and related distributions*, arXiv:1305.0924.
[33] T. HASEBE AND H. SAIGO, *The monotone cumulants*, Ann. Inst. Henri Poincaré Probab. Stat. **47** (2011), no. 4, 1160–1170.
[34] W. HELTON, R. RASHIDI FAR, AND R. SPEICHER, *Operator-valued semicircular elements: solving a quadratic matrix equation with positivity constraints*, Int. Math. Res. Not. IMRN no. 22, (2007), Art. ID rnm086, 15.
[35] R. HUDSON AND K. PARTHASARATHY, *Quantum Ito's formula and stochastic evolutions*, Comm. Math. Phys. **93**, no. 3, (1984), 301–323.
[36] V. KARGIN, *Berry-Esseen for free random variables*, J. Theor. Probab. **20**, (2007), 381–395.
[37] A. KHINTCHINE, *Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze*, Rec. Math. Moskou **2**, (1937), 79–117.
[38] A. KOLMOGOROV, *Sulla forma generali di un processo stocastico omogeneo*, Rend Acai Lincei (6) **15**, (1932), 163–168, 325–329, 548–553.
[39] T. KEMP, I. NOURDIN, G. PECCATI, AND R. SPEICHER, *Wigner chaos and the fourth moment*, Ann. Probab. **40**, no. 4, (2012), 1577–1635.
[40] P. LÉVY, *Sur les integrales dont les elements sont des variables aleatoires independantes*, Ann. Ec. Norm. Pisa (2) **3**, (1934), 337?-366.
[41] P. LÉVY, Theorie de l'addition des variables aleatoires, Gauthier Villars, Paris, 1937.
[42] T. MAI AND R. SPEICHER, *Operator-valued and multivariate free Berry-Esseen theorems*, arXiv:1202.2740v1.
[43] H. MAASSEN, *Addition of freely independent random variables*, J. Funct. Anal. **106**, (1992), 409–438.
[44] N. MURAKI, *Monotonic Independence, Monotonic Central Limit Theorem and Monotonic Law of Small Numbers*, Infinite dimensional Analysis, Quantum Probability and Related Topics **4**, (2001), 39–58.
[45] N. MURAKI, *The five independences as natural products*, Infin. Dimen. Anal. Quantum Probab. Relat. Top **6**, (2003), 337–371.
[46] N. MURAKI, Monotonic convolution and monotonic Lévy-Hincin formula, preprint, 2000.

[47] N. MURAKI, *The five independences as quasi-universal products*, Infin. Dimen. Anal. Quant. Probab. Relat. Topics **5**, (2002), 113–134.

[48] A. NICA AND R. SPEICHER, Lectures on the combinatorics of free probability, Cambridge University Press, 2006.

[49] A. NICA AND R. SPEICHER, *Commutators of free random variables*, Duke Math. J. **92**, (1998), 553–592.

[50] D. NUALART AND G. PECCATI, *Central limit theorems for sequences of multiple stochastic integrals*, Ann. Probab. **33**, no. 1, (2005), 177–193.

[51] V. PEREZ-ABREU AND N. SAKUMA, *Free generalized gamma convolutions*, Elect. Comm. in Probab. **13**, (2008) 526–539.

[52] R. RASHIDI FAR, T. ORABY, W. BRYC, AND R. SPEICHER, *On slow-fading MIMO systems with nonseparable correlation*, IEEE Trans. Inf. Theory **54**, (2008), 544–553.

[53] D. SHLYAKHTENKO, *Random Gaussian band matrices and freeness with amalgamation*, Internat. Math. Res. Notices no. 20, (1996), 1013–1025. 1422374 (97j:46070)

[54] R. SPEICHER, *Multiplicative functions on the lattice of non-crossing partitions and free convolution*, Math. Ann. **298**, (1994), 611–628.

[55] R. SPEICHER,*On universal products*, Fields Institute Communications, D. Voiculescu, (ed). **12**, (1997), 257–266.

[56] R. SPEICHER, *Combinatorial theory of the free product with amalgamation and operator-valued free probability theory*, Memmoir of the AMS, **627**, (1998).

[57] R. SPEICHER AND C. VARGAS, *Free deterministic equivalents, rectangular random matrix models and operator-valued free probability*, Random Matrices: Theory and Applications **1**, no. 2, (2012).

[58] R. SPEICHER AND R. WOROUDI, *Boolean convolution, in Free Probability Theory*, Ed. D. Voiculescu, Fields Inst. Commun. **12** (Amer. Math. Soc., 1997), 267–280.

[59] F.W. STEUTEL AND K. VAN HARN, *Infinite Divisibility of Probability Distributions on the Real Line*, Monographs and textbooks in Pure and Applied Mathematics **259**, Marcel-Dekker, New York, 2003.

[60] N. SAITO AND H. YOSHIDA, *The infinite divisibility and orthogonal polynomials with a constant recursion formula in free probability theory*, Probab. Math. Statist. **21**, no. 1, (2001), 159–170.

[61] D. VOICULESCU, *Symmetries of some reduced free product C^*-álgebras*, In: Operator Algebras and their Connections with Topology in Ergodic Theory. Lecture Notes in Mathematics **1132**, Springer Verlag, (1985), 556–588.

[62] D. VOICULESCU, *Addition of certain non-commutative random variables*, J. Funct. Anal. **66**, (1986), 323–346.

[63] D. VOICULESCU, *Operations on certain non-commutative operator-valued random variables*, Astérisque **232**, (1995), 243–275, Recent advances in operator algebras (Orléans, 1992). 1372537 (97b:46081).

[64] D. VOICULESCU, *Limit laws for random matrices and free products*, Invent. Math. **104**, (1991), 201–220.

[65] D. VOICULESCU, K. DYKEMA, AND A. NICA, *Free Random Variables*, CRM Monograph Series **1**, Amer. Math. Soc., Providence, 1992.

[66] E. WIGNER, *Characteristic vectors of bordered matrices with infinite dimensions*, Annals of Mathematics **62**, (1955), 548–564.

[67] J.C. WANG, *Local Limit theorems in free probability*, Ann. Probab. **38**, No.4, (2010), 1492–1506.

Bol. Soc. Mat. Mexicana (3) Vol. 19, 2013

A REVIEW OF CONDITIONAL RARE EVENT SIMULATION FOR TAIL PROBABILITIES OF HEAVY TAILED RANDOM VARIABLES

LEONARDO ROJAS-NANDAYAPA

ABSTRACT. Approximating the tail probability of a sum of heavy-tailed random variables is a difficult problem. In this review we exhibit the challenges of approximating such probabilities and concentrate on a rare event simulation methodology capable of delivering the most reliable results: Conditional Monte Carlo. To provide a better flavor of this topic we further specialize on two algorithms which were specifically designed for tackling this problem: the Asmussen-Binswanger estimator and the Asmussen-Kroese estimator. We extend the applicability of these estimators to the non-independent case and prove their efficiencies.

1. Introduction

The term *rare event* is used to designate all those events whose probabilities are small, yet non-negligible and characterized by the difficulty of its calculation. Often, these rare events are extremely important in applications; for instance, consider the consequences of a natural disaster for an insurance company, or an economic crisis for a financial institution or the sudden arrival of huge number of jobs to a server as it often occurs in a web server. Many of the probability models employed for dealing with these problems contain multiple random variables (not necessarily independent) and the quantities of interest are given in terms of transformations such as sums, products or extremes. In consequence, the explicit calculation of a distribution of interest is often non-trivial and one must rely on approximation methods. Among these, the Monte Carlo method is considered to be one the most reliable, specially in cases where analytical approximations are not available.

In this review we mainly focus on the Monte Carlo method for approximating rare event probabilities, but we also discuss asymptotic approximations; the reason for this is that the implementation of efficient Monte Carlo estimators often requires to draw elements from asymptotic theory. In particular, we specialize on tail probabilities of a sum of random variables

$$\mathbb{P}(S_n > x), \qquad x \to \infty.$$

When the involved random variables are light tailed, the approximation of such probabilities is dealt via Large Deviations theory. Notwithstanding, certain phenomena are better modeled with heavy-tailed distributions. However, the approximation of rare event probabilities in the presence of heavy tails is often more involved and it has been considered a challenging problem among the applied

2010 *Mathematics Subject Classification*: 60E05, 90-04.

Keywords and phrases: rare-event simulation, conditional Monte Carlo, heavy-tails, efficiency, non-independent.

probability community. The reason for this is that most classical methods require that the domain of convergence of the moment generating function contains an open set including the origin —a condition which is not satisfied by heavy-tailed random variables as these are characterized by the non-existence of their moment generating functions for positive values of the argument. Therefore, new methods have been called for tackling this problem, and as a result we have seen in the last fifteen years a very intense research activity devoted to Rare Event Simulation. In this review, we focus on the methodology called Conditional Monte Carlo, which has provided some of the most powerful and efficient estimators so far. Some of our contributions employ this technique [5, 23, 11]. Here we include some results which have not been previously published in peer-review journals. These are the generalizations of the so called Asmussen-Binswanger and Assmusen-Kroese estimators for the case of independent but non-identical random variables. These extensions are accompanied by their corresponding proofs of efficiency.

This paper is structured as follows. In section 2 we provide a discussion on independent heavy-tailed random variables. This theory is now at an advanced level and well-understood. Several alternative definitions for heavy tails are reviewed and their relations and main properties are studied. In particular, we pay attention to the rich class of subexponential distributions and we discuss how its defining property provides a useful insight into the occurrence of large values of a sum —a characteristic behavior known as the principle of the single big jump. Moreover, it has been recognized that the subexponential property goes beyond the independent case and it is now an area of active research. One of the main contributions of this author is in this front. The main result in [6] states that a sum of lognormals possesses the subexponential property even when the involved random variables are correlated via a Gaussian dependence structure.

A general overview of Monte Carlo methods is provided in Section 3 with a particular emphasis in the area known as *Rare Event Simulation*. The notions of *rare event* and *efficient estimator* are formalized here in order to provide the proper framework for analyzing Monte Carlo estimators for rare event probabilities. We discuss the classical tools such as importance sampling, exponential change of measure and conditional Monte Carlo. We discuss briefly the limitations of some standard methods when applied in a heavy-tailed setting. Section 4 is devoted exclusively to the approximation of tail probabilities of sums of heavy-tailed random variables; a recount of available methods is given there followed by a more detailed exposition on a set of estimators based on the Conditional Monte Carlo; these are known as the Asmussen-Binswanger [1] and the Asmussen-Kroese [4]. In particular, we provide extensions to the non-independent case and prove the efficiency of these estimators. We stress the fact that Theorems 4.1–4.3 are original contributions and their efficiency proofs can be found in the Appendix in Section 6. Finally, Section 5 contains some concluding remarks.

2. Heavy Tails

The term *heavy-tailed phenomena* [21], is often used to refer to real world phenomena where record values are characterized by its extreme behavior. Examples of this type of phenomena are abundant in insurance; for instance, consider the two record costliest (adjusted for inflation) hurricanes striking the United States

during the period 1900-2010: Katrina (2005) and Andrew (1992) with damage costs of 105,840 and 45,562 million USD respectively [9]. Both records have extreme values but the most striking feature is that the damage cost of hurricane Katrina more than doubles the damage cost of hurricane Andrew! Further examples occur in Finance and Telecommunications where economic losses or system breakdowns due to large data file sizes or long transmission lengths are of great concern.

The examples above sketch the huge relevance of heavy-tailed phenomena and stress the importance of having the right probabilistic distributions for modeling their behavior. In the rest of this section we will provide alternative definitions which lead to several classes of heavy-tailed distributions and will study their properties. In particular, we will pay attention to the distinctive behavior of convolutions of certain types of heavy-tailed distributions known as the *principle of the single big jump*. We also establish some contrasts with respect to light tailed distributions, which typically comprehend most of the classical models in probability and statistics. Our exposition follows closely [16] but we also draw elements from [15] and [20].

We say that a random variable X has a (right) *heavy-tailed distribution* if

$$\mathbb{E}[\mathrm{e}^{\theta X}] = \infty, \qquad \forall \theta > 0.$$

Foss, Korshunov and Zachary [16] employ the term *exponential moment* to refer to the quantity $\mathbb{E}[\mathrm{e}^{\theta X}]$. Adopting this terminology, we say that X has a heavy-tailed distribution if it fails to have a positive exponential moment. In contrast, if a distribution with unbounded right support has a finite positive exponential moment, then we say that it has a *light-tailed distribution*. Further to this, we can easily verify that light-tailed distributions have moments of every order while a random variable with an infinite moment of any order will necessarily have a heavy-tailed distribution. The converse of the last statement is false in general; the classical example is that of a lognormal random variable which has finite moments of every order but it fails to have a positive exponential moment, and in consequence classifies as a heavy-tailed distribution.

In the definition above it is implicit that a right heavy-tailed distribution should have an unbounded right support. In fact, the defining property of a heavy-tailed distribution is inherently related to the rate of decay of its tail probability $\overline{F}(x) := 1 - F(x)$. Therefore, it is natural to obtain equivalent definitions of a heavy-tailed distribution in terms of its tail probability or its hazard rate function. In particular, we define the *hazard function* $\Lambda(x)$ as

$$\Lambda(x) := -\log \overline{F}(x).$$

Moreover, if the tail probability $\overline{F}$ of a distribution is differentiable, then we define the *hazard rate function* $\lambda(x) := \Lambda'(x)$. Hazard (rate) functions arise in a wide variety of applications in survival analysis and reliability where it is known under alternative names such as *survival* or *failure* (rate) functions. In addition, we say that an arbitrary nonnegative function f is *heavy-tailed* iff

$$\limsup_{x \to \infty} \frac{f(x)}{\mathrm{e}^{-\theta x}} = \infty, \qquad \forall \theta > 0.$$

Thus a function is heavy-tailed if it decays slower than an exponential function. The following theorem (cf. [16, Theorem 2.6]) ties together the properties of the

tail and the hazard functions of a heavy-tailed distribution and provides alternative definitions.

THEOREM (2.1). *Let F be a distribution function with unbounded right support. The following are equivalent:*

1. *F is a heavy-tailed distribution.*
2. *$\overline{F}$ is a heavy-tailed function.*
3. $\liminf_{x\to\infty} \Lambda(x)x^{-1} = 0$.

These definitions provide practical means for testing the *heaviness* of any given distribution. Classical examples of heavy-tailed distributions include the subfamily of regularly varying distributions (including Pareto, Loggamma, Burr), the Weibull distribution with parameter $0 < \lambda < 1$, the Cauchy and the Lognormal distributions. On the other hand, the exponential, gamma and normal random variables are examples of light-tailed distributions. In fact, from the definition of a heavy-tailed function it follows that an exponential transformation of a *light-tailed* random variable might yield a heavy-tailed random variable. In particular, the Pareto, loggamma and lognormal are heavy-tailed distributions obtained from an exponential transformation of the exponential, gamma and normal distributions respectively.

The class of heavy-tailed distributions as defined above is too general for deriving useful properties. However, by adding some regularity conditions one can obtain tractable subclasses of heavy-tailed distributions which posses attractive properties, yet remain general enough. One of such subclasses is that of long tailed distributions. This subclass is denoted $\mathcal{L}$ and defined by the following property. A distribution $F \in \mathcal{L}$ iff

$$\lim_{x\to\infty} \frac{\overline{F}(x+y)}{\overline{F}(x)} = 1, \qquad \forall y \in \mathbb{R}.$$

An analogue definition exists for a general function f which is ultimately positive and posses the property listed above; in such case we say that f is a long-tailed function. The following result [16, Lemma 2.17] provides the connection between heavy and long tailed distributions.

LEMMA (2.2). *If f is a long-tailed function then*

$$\lim_{x\to\infty} \frac{f(x)}{\mathrm{e}^{-\theta x}} = \infty, \qquad \forall \theta \geq 0.$$

In consequence, a long-tailed distribution F is necessarily heavy-tailed, but the converse is not always true. The reference [16] provides a counterexample of a heavy-tailed distribution which fails to be long-tailed. By adding the smoothness condition which defines the class of long-tailed distributions, we gain some useful properties in exchange of some generality. In particular, it is possible to prove that the class of long-tailed functions is closed under linear transformations (mixtures), products, maxima, minima, and convolutions. Moreover, the convolution of a long-tailed distribution with an arbitrary distribution is long-tailed.

A very useful characterization of long-tailed distributions is via their *insensitiveness* with respect to a function h. More precisely, we say that a function f is

h-insensitive [16] iff

$$\sup_{|y|\le h(x)} |f(x+y)-f(x)| = o(f(x)), \qquad x \to \infty,$$

uniformly in $|y| \le h(x)$. If the function f is monotone, then h-insensitivity reduces to having $f(x+h(x)) \sim f(x)$ as $x \to \infty$. Clearly, a long-tailed function is insensitive with respect a constant function. However, this property can be strengthened as shown in the following Lemma:

LEMMA (2.3). *If F is a long-tailed distribution, then there exists a function $h(x) \to \infty$ such that $\overline{F}$ is an h-insensitive function.*

For instance, if $\overline{F}$ is regularly varying then it is $o(x)$-insensitive, while the lognormal is $o(x/logx)$-insensitive and the heavy-tailed Weibull with parameter $\lambda \in (0,1)$ is $o(x^{1-\lambda})$-insensitive.

Next, we discuss briefly the relationship between long-tailed distributions, integrated tails and the mean excess function. If a distribution function is such that $\int_0^\infty \overline{F}(x)\,dx < \infty$, then we can define the *integrated tail distribution of F* as

$$\overline{F}_I(x) := \min\left\{1, \int_x^\infty \overline{F}(t)\,dt\right\}.$$

The *mean excess function* can be defined for a distribution having a finite first moment as

$$e(x) := \mathbb{E}[X - x | X > x].$$

That is, the mean excess function is the expected value of the excess of a random variable over a given x, provided that it has exceeded this threshold value. The mean excess function is related to the integrated tail distribution via the relation $e(x) = \overline{F}_I(x)/\overline{F}(x)$. Moreover, the following Lemma [16, Lemma 2.25] provides a useful characterization of long-tailed distributions in terms of mean-excess functions.

LEMMA (2.4). *The integrated tail distribution F_I is long-tailed iff its associated mean excess function is such that $e(x) \to \infty$.*

In applications, the mean excess function is often used to diagnose the presence of heavy-tails. However, the previous lemma shows that if $e(u) \to \infty$ we can only verify that the integrated tail distribution is long-tailed but we cannot say anything about the *heaviness* of the original distribution F. In fact, one can construct a counterexample of a light tailed distribution whose mean excess function goes to infinity. This is a case of a more general fact which says that if F is an absolutely continuous distribution, then its density f is a heavy-tailed function but the converse is false in general; that is, if a density function f is heavy-tailed, its distribution function is not necessarily heavy-tailed. In consequence, the mean excess function of a heavy-tailed function should increase to infinity as we let the threshold value $x \to \infty$, but it is not an absolutely reliable tool to diagnose a heavy tail because a distribution with mean excess function going to infinity is not necessarily heavy-tailed. Counterexamples for all these cases can be found in [16].

Convolutions of certain regular and nonnegative heavy-tailed distributions have a unique property which set them apart from light-tailed distributions: the *principle of the single big jump*. This property is extremely useful and most of the heavy-tailed distributions used in practice possess it; in fact, the family of

distributions defined by this property forms a proper subclass of long-tailed distributions. To define it we concentrate exclusively on distributions with nonnegative values, but point out that some of the subclasses of heavy-tailed distributions defined below can be generalized to distributions supported over the reals.

We start with the following elementary property of a convolution which holds for all nonnegative distributions F with unbounded right support. Let F^{*n} the n-fold convolution of F and $\overline{F^{*n}}$ its corresponding tail distribution, then

$$\liminf_{x\to\infty}\frac{\overline{F^{*n}}(x)}{\overline{F}(x)}\geq n, \qquad \forall\, n\in\mathbb{N}.$$

The following theorem [16, Theorem 2.12] provides a sufficient condition for the liminf above to be equal to n. In fact, this will provide a very useful insight into the characteristic behavior of the convolution of a heavy-tailed distributions:

THEOREM (2.5). *Let F be a nonnegative heavy-tailed distribution. Then*

$$\liminf_{x\to\infty}\frac{\overline{F^{*n}}(x)}{\overline{F}(x)}=n, \qquad \forall\, n\in\mathbb{N}. \tag{2.6}$$

The corresponding liminf of most light-tailed nonnegative distributions like the exponential and gamma will be infinite. Hence, it is tempting to use the theorem above as an alternative definition of heavy-tails, but it turns out this is not possible as one can construct a light tailed nonnegative distribution for which the liminf of the ratio of convolutions as defined above is equal to n. A counterexample can be found in [16]. However, we can strengthen the condition above to obtain a subclass of long-tailed distributions: *subexponential distributions*, This subclass was originally introduced by Chistyakov in 1964 [13]. We say that a nonnegative distribution F belongs to the class of subexponential distributions, denoted $\mathcal{S}$, if it possesses the *subexponential property*; that is, the tail probability of the n-fold convolution of F is asymptotically equivalent to n times the tail probability $\overline{F}$. More precisely,

$$\lim_{x\to\infty}\frac{\overline{F^{*n}}(x)}{\overline{F}(x)}=n, \qquad \forall\, n\in\mathbb{N}. \tag{2.7}$$

Therefore, a heavy-tailed distribution requires and additional regularity condition to be subexponential. That condition is the existence of the limit (2.7). Moreover, it is possible to prove that that subexponential distributions form a proper subclass of long-tailed distributions. That is, any subexponential distribution is long-tailed but not every long-tailed distribution will necessarily be subexponential. For a counterexample see [16]. Curiously, the name *subexponential* was originally employed to refer to the class of distributions satisfying $\lim_{x\to\infty}\overline{F}(x)e^{\lambda x}<\infty$, but nowadays it is employed in the more restrictive sense described above. Subexponentiality is a property of the tail exclusively; however, it has a very interesting implication for the tail of an $n-$convolution —a characteristic which often goes under the name of the *principle of the single big jump*. Let us start by noting that the distribution of the maximum M_n of n arbitrary i.i.d. random variables (not necessarily subexponential) is given by $F^n(u)$. Hence

$$\mathbb{P}(M_n>x)=1-F^n(x)=1-(1-\overline{F}(x))^n=1-\sum_{k=0}^{n}(-1)^k\binom{n}{k}\overline{F}^{n-k}(x)=\overline{F}(x)\big(n+\mathrm{O}(\overline{F}(x))\big).$$

In consequence, the tail probability of the maximum is asymptotically equivalent to the tail probability of the convolution, namely $\mathbb{P}(M_n > x) \sim \mathbb{P}(S_n > x)$, where $S_n := X_1 + \cdots + X_n$. Since the $X_i's$ are nonnegative, then $\{M_n > x\} \subset \{S_n > x\}$ and the conclusion above can be written in a conditional form as

$$\lim_{u\to\infty} \frac{\mathbb{P}(M_n > u)}{\mathbb{P}(S_n > u)} = \lim_{u\to\infty} \mathbb{P}(M_n > u | S_n > u) = 1.$$

This expression is very appealing as it says that the if the sum becomes large it is only likely due to the contribution of a single random variable. This behavior is completely opposite to that of lighted tails where the only likely way that a sum of i.i.d. random variables becomes large is as a consequence of several moderately large but otherwise proportionally sized contributions of two or more random variables. Hence, it turns out that distributions within the class $\mathcal{S}$ should be appropriate for modeling those phenomena which show some stability through time but eventually are shocked by an extreme event. Subexponential distributions inherit the properties of long-tailed distributions but also possess many of their own; for instance, the class of subexponential distributions is closed under maxima, minima, mixtures, convolutions and random translations. Also, most of the heavy-tailed distributions used in practice are subexponential such as the Pareto, loggamma, Burr, Weibull and Lognormal. Also, in the case of nonnegative distributions, there exists an upper bound for the expression $\overline{F^{*n}}(x)/\overline{F}(x)$. Such bound goes under the name of Kesten's bound [16, Theorem 3.34], and it is described in the following Theorem:

THEOREM (2.8). *Let F be a subexponential distribution. Then, for every $\epsilon > 0$ there exists a constant c such that for all $x \geq 0$ and all $n \geq 1$ it holds that*

$$\frac{\overline{F^{*n}}(x)}{\overline{F}(x)} \leq c(1+\epsilon)^n.$$

Notice that the subexponential property is given for nonnegative, independent and identically distributed random variables, thus it would be desirable to extend this definition to more general sets of random variables and investigate more general conditions under which the principle of the single big jump holds. First we concentrate on distributions supported on the whole real line where it turns out that the defining property (2.7), which from now on we call *subexponential-type property*, is no longer a tail property. For instance, if we consider a distribution supported all over the reals which fulfills the subexponential-type property, then the distribution $F^+(x) := F(x)\mathbb{I}_{x\geq 0}$ will not necessarily be subexponential and the principle of the single big jump does not hold anymore; this argument motivates the following alternative definition. We say that a distribution F is *whole-line subexponential* if F^+ is subexponential. Alternatively, one could obtain an equivalent definition by consider the distribution $G(x) = \mathbb{P}(X \leq x | X \geq 0)$ instead of F^+. The following theorem [16, Lemma 3.4 and Theorem 3.6] summarize two alternative equivalent definitions which provide useful insights into the class of whole-line subexponential distributions.

THEOREM (2.9). *The following assertions are equivalent:*

1. *F is whole-line subexponential.*
2. *F is long-tailed and it possesses the subexponential-type property.*

3. *F is long-tailed and there exists a function $h(x) \to \infty$ for which F is h-insensitive and such that for any two independent random variables $X_1, X_2 \sim F$ it holds that*

$$\mathbb{P}(X_1 + X_2 > x, X_1 > h(x), X_2 > h(x)) = o(F(x)), \qquad x \to \infty.$$

This theorem says that for a distribution to be whole-line subexponential it is not enough to just have the *subexponential-type property* but we also require a long-tail. Clearly, whole-line subexponential distributions form a proper subclass of long-tailed and heavy-tailed distributions. Moreover, the third part of the theorem above shows that the principle of the single big jump holds as it says that it is unlikely to observe a large value of the sum as a consequence of two (or more) random variables taking moderately large values.

The following Lemma generalizes the principle of the single big jump to independent but non-identically distributed random variables [16, Corollary 3.18].

LEMMA (2.10). *Let F be a whole-line subexponential distribution and $F_1, \dots, F_n$ be a collection of distributions such that $\lim_{x\to\infty} \overline{F}(x)/\overline{F}_i(x) = c_i \geq 0$. Then it holds that*

$$\lim_{x\to\infty} \frac{\overline{F_1 * F_2 * \cdots * F_n}(x)}{\overline{F}(x)} = \sum_{i=1}^{n} c_i.$$

If $\lim_{x\to\infty} F_1(x)/F_2(x) = c > 0$, then we say that F_1 and F_2 are *tail equivalent*. It is straightforward to prove that if F_2 is tail equivalent to a long-tailed distribution, then F_2 is long tailed as well.

The last subclass of heavy-tailed distributions that we will discuss are the subfamilies of distributions with regularly varying tails with index α, denoted $\mathcal{R}(\alpha)$, and defined as the family of nonnegative random variables whose tail probability can be written as $\overline{F}(x) = L(x)x^{-\alpha}$ with $x, \alpha > 0$, and $L(x)$ is a *slowly varying function*. That is, $L(x)$ is a measurable function satisfying

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1, \qquad \forall t \in (0, \infty).$$

In particular, it holds that F is regularly varying distribution iff

$$\lim_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = t^{-\alpha}, \qquad \forall t \in (0, \infty).$$

The class $\mathcal{R}_\alpha$ is often understood as those distributions with a tail behavior similar to a power function with exponent α while the slowly varying function $L(x)$ acts as a perturbation factor. This class has been largely studied under the more general theory of *regularly varying functions* [7, cf. 7]. Many authors consider that regularly varying is a synonym of heavy-tails [21]; moreover, this subfamily play fundamental roles in the theory of fluctuations of sums and extremes of independent random variables. The Pareto, Burr, α-stable and loggamma are typical examples of regularly varying distributions. The theory of regularly varying distributions is quite extensive, so we decide to omit most of it but we enunciate a few properties. One of the most remarkable results is Karamata's theorem which is as follows: Let $L \in \mathcal{R}_0$ be bounded in $[x_0, \infty)$ and $\alpha > 1$. Then

$$\int_x^\infty \frac{L(t)}{t^\alpha} dt = \frac{L(x)}{(\alpha - 1)x^{\alpha-1}}(1 + o(1)) \qquad x \to \infty.$$

This result says that the integrated tail of a regularly varying function with index $\alpha > 1$ will be regularly varying with index $\alpha - 1$. Even more, it says that the slowly varying function is preserved after the integration. Using *Karamata's Theorem* it is easy to verify that the *mean excess function* $e(u)$ of a regularly varying goes to infinity as $u \to \infty$. Similarly, all the moments of order large than the index α of a regularly varying distribution are infinite while those of smaller order than α are finite. It is trivially seen that the tail probability decays slower than the exponential and it is also provable that a regularly varying distribution satisfies the characteristic property of subexponential distributions. Hence, the class $\mathcal{R}$ inherits all the properties of the class $\mathcal{S}$.

3. Rare Event Simulation

As discussed in the previous section, most of the heavy-tailed distributions used in practice belong to the class of whole-line subexponential distribution. Therefore, the subexponential type-property can be used to approximate the tail probability of a sum of heavy-tailed random variables. Such approximation is very precise in the asymptotic regions of the tail distribution; however, this approximation can loose some precision for moderately large values. Hence, it is desirable to obtain sharper approximations and a natural choice is to recourse to the Monte Carlo method. The elementary version for calculating the tail probability $p_x := \mathbb{P}(X_1 + \cdots + X_n > x)$ is the so called *Crude Monte Carlo* and consists in simulating R identical copies of the random vector $(X_1, \ldots, X_n)$, say $\{(X_{1,r}, \ldots, X_{n,r}), r = 1, 2, \ldots\}$; calculating the sums $S_r := X_{1,r} + \cdots + X_{n,r}$; defining the (Bernoulli) random variables $W_{r,x} := \mathbb{I}(S_r > x)$ and returning the arithmetic average

$$\widehat{p}_{x,R} := \frac{1}{R} \sum_{r=1}^{R} W_{r,x}.$$

The law of large numbers implies that for a fixed x, the sequence of random variables $\widehat{p}_{x,R}$ converges to p_x as $R \to \infty$. Moreover, since the random variables $\{W_r : r = 1, 2, \ldots\}$ have bounded variance, the Central Limit Theorem implies that a measure for the random error is the *margin of error* of the Crude Monte Carlo estimator

$$\mathrm{me}(\widehat{p}_{x,R}) := \sqrt{\frac{p_x(1 - p_x)}{R}}.$$

This formula exhibits the natural trade-off between precision and computational effort that is required for obtaining an estimate. While in theory we could attain any desirable level of precision by simply increasing the number of replications, it is not so uncommon to end up with very long running times which make unfeasible to attain a certain desired precision. In fact, the crude version of the Monte Carlo is fated to deliver poor approximations when used to estimate *rare event probabilities*. More precisely, we say that an indexed family of events $\{A_x : x \in \mathbb{R}\}$ is a sequence of rare events if $p_x := \mathbb{P}(A_x) \to 0$ as $x \to \infty$. Crude Monte Carlo is considered to deliver poor estimates for rare event probabilities because the asymptotic order of the margin of error is larger than the asymptotic order of the probability of interest as the events becomes rarer. For instance, the *relative error*

of the Crude Monte Carlo estimator goes to infinity as the event becomes rarer:

$$\lim_{x\to\infty} \frac{\mathrm{me}(\widehat{p}_{x,R})}{p_x} = \lim_{p_x\to 0} \sqrt{\frac{1-p_x}{p_x R}} = \infty.$$

This implies that the number of replications needed to achieve certain relative precision grows to infinity as the event becomes rarer. This discussion makes obvious the two following facts. 1) The margin of error is not an appropriate precision measure for rare event simulation; instead, we should look at the relative error as defined above (or equivalently to the *coefficient of variation* defined as the square of the margin of error). 2) We need to turn our attention to alternative collections of Monte Carlo estimators requiring a finite number of replications for achieving certain relative precisions no matter how rare the event is.

(3.1) Efficiency criteria in rare event simulation. First we discuss the efficiency criteria employed in rare event simulation. In a rare event framework, we say that a Monte Carlo estimator $\widetilde{p}_{x,R}$ is *strongly efficient* or has *bounded relative error* if the (single-replicate) estimator has the following property

$$\limsup_{x\to\infty} \frac{\mathbb{V}\mathrm{ar}\widetilde{p}_{x,1}}{p_x^2} < \infty.$$

This efficiency property says that the number of replications required to estimate p_x with certain fixed relative precision remains bounded as $p_x \to 0$. However, it is often difficult to construct such estimators and/or prove that the limsup above remains bounded. For that reason it is common to employ an alternative weaker criterion denominated *logarithmic efficiency*. This is defined as

$$\limsup_{x\to\infty} \frac{\mathbb{V}\mathrm{ar}\widetilde{p}_{x,1}}{p_x^{2-\epsilon}} = 0, \qquad \forall \epsilon > 0.$$

This criterion implies that the number of replications needed for achieving certain relative precision grows at most at rate of order $|\log(p_x)|$. From a practical point of view, there is no substantial difference between these two criteria, but as mentioned before it is often much easier to prove logarithmic efficiency not only because it is a weaker criterion but also due to the equivalent definition given in the following result.

LEMMA (3.1). *An estimator $\widetilde{p}_x$ is logarithmically efficient iff*

$$\liminf_{x\to\infty} \frac{\left|\log \mathbb{V}\mathrm{ar}\,\widetilde{p}_x\right|}{\left|\log p_x\right|} \geq 1.$$

The condition given in the previous Lemma often arises in Large Deviation theory, but are now standard in Rare-Event simulation. The proof of this Lemma is standard but to the best of the author's knowledge, it has seldom appeared in the rare event simulation literature. For sake of completeness, we provide an alternative proof.

Proof of Lemma (3.1). Let us first assume that

$$\liminf_{x\to\infty} \frac{|\log \mathbb{V}\mathrm{ar}\,\widetilde{p}_{x,1}|}{2|\log p_x|} \geq 1$$

Then for all $\epsilon > 0$ there exists x_0 such that the inequality on the left hand side on the following display holds for all $x \geq x_0$

$$\frac{|\log \mathbb{V}\mathrm{ar}\, \widetilde{p}_{x,1}|}{2|\log p_x|} > 1 - \epsilon/2 \qquad \Longleftrightarrow \qquad \frac{\mathbb{V}\mathrm{ar}\, \widetilde{p}_{x,1}}{p_x^{2-\epsilon}} < 1.$$

The inequality on the right hand side above is obtained by simple algebraic manipulations of the inequality in the left hand side. Taking lim sup we obtain

$$\limsup_{x\to\infty} \frac{\mathbb{V}\mathrm{ar}\, \widetilde{p}_{x,1}}{p_x^{2-\epsilon}} < 1.$$

The last inequality holds for all $\epsilon > 0$, hence the the limsup is necessarily smaller or equal to 0. The converse is proved in a similar way. Let us assume

$$\limsup_{x\to\infty} \frac{\mathbb{V}\mathrm{ar}\, \widetilde{p}_{x,1}}{p_x^{2-\epsilon}} = 0, \qquad \forall \epsilon > 0.$$

Hence, for all $0 < \delta < 1$ there exist x_0 such that for all $x \geq x_0$ the inequality on the left hand side of the following display holds

$$\frac{\mathbb{V}\mathrm{ar}\, \widetilde{p}_{x,1}}{p_x^{2-\epsilon}} < c \quad \Rightarrow \quad \left| \frac{\log \mathbb{V}\mathrm{ar}\, \widetilde{p}_{x,1}}{2|\log p_x|} \right| > \left| \frac{\log c}{|\log p_x|} - \frac{2-\epsilon}{2} \right|$$

The inequality on the right hand side follows from standard algebraic manipulations. Taking lim inf on both sides we obtain that

$$\liminf_{x\to\infty} \frac{|\log \mathbb{V}\mathrm{ar}\, \widetilde{p}_{x,1}|}{2|\log p_x|} > 1 - \epsilon/2, \qquad \Longleftrightarrow \qquad \liminf_{x\to\infty} \frac{|\log \mathbb{V}\mathrm{ar}\, \widetilde{p}_{x,1}|}{2|\log p_x|} \geq 1.$$

This completes the proof. □

A stronger efficiency concept is achievable and has been proven for several algorithms in the recent literature. This criterion goes under several different names such as *asymptotically zero relative error* or *vanishing relative error*, and it is defined as follows. We say that an estimator $\widehat{p}_{x,1}$ has asymptotically zero relative error iff

$$\limsup_{x\to\infty} \frac{\mathbb{V}\mathrm{ar}\, \widetilde{p}_{x,1}}{p_x^2} = 0.$$

This criterion is stronger than bounded relative error. In fact, when an estimator has *asymptotically zero relative error*, it can theoretically produce a variance reduction such that the number of replications necessary to attain certain precision is of order $\mathrm{O}(p_x)$ as $x \to \infty$. That means that ultimately the number of replications needed for achieving certain relative precision will continue to decrease as the event becomes rarer until it will be necessary to have a single replication. In addition, there exist, on one hand, efficiency criteria which are weaker than *logarithmic efficiency*. More precisely, for a fixed value $\delta > 0$ we say that an estimator $\widehat{z}(x)$ is δ-*efficient* if

$$\limsup_{u\to\infty} \frac{\mathbb{V}\mathrm{ar}\widehat{z}(u)}{z^{2-\delta}(u)} < \infty.$$

This definition fills the gap between *logarithmic efficiency* and *Crude Monte Carlo efficiency* and it is often used to describe the improvement over Crude Monte Carlo. On the other hand, there are stronger efficiency concepts which take care of the moments of higher order of an estimator. These go under the name *bounded relative error of order k* [19].

Before we move on, we would like to remark that proving that a candidate estimator satisfies any of the efficiency properties listed above is often a very difficult problem. The reason for this is that the variance of an estimator (appearing on the numerator of the efficiency criteria discussed) is often unknown and one has to find an upper bound of the second moment of the estimator which is tight enough so it remains asymptotically bounded by the appropriate power of the first moment.

(3.2) Variance reduction techniques. One of the the most important practical tasks in rare-event simulation is to propose estimators for a given sequence of rare events which may satisfy any of the efficiency properties discussed in the previous section. In a more general framework, the set of techniques employed used to produce estimators which improve the performance of the Crude Monte Carlo estimator go under the name of *variance reduction methods* (c.f. [3]). More precisely, a variance reduction method is an algorithm that modifies an existing estimator (or constructs a new one) in such a way that the resulting estimator remains unbiased and (hopefully) produces a reduction in variance when compared to Crude Monte Carlo. Among the most notorious variance reduction methods one could list *Importance Sampling*, *Control Variates*, *Stratification*, *Conditional Monte Carlo*, *Antithetic sampling*. In addition, variance reduction methods can be divided in *static* and *adaptive*. In a static method, every step of the algorithm is conducted independent of the outcome; in contrast, the evolution of an adaptive algorithm depends on previous outcomes of the algorithm.

While most of these methods are potentially able to produce smaller variances than Crude Monte Carlo, not all of these are well suited for rare event simulation. The main reason is that the demand of variance reduction in the presence of rare events is huge. As discussed in the previous section, one requires a variance which is of much lower order than the one provided by Crude Monte Carlo. Moreover, the quality of these methods is often assessed not only based on the variance reduction itself but also in the amount of computational resources consumed, the theoretical work required and the implementation effort invested.

Among the most powerful methods mentioned above, the most effective ones for rare event simulation are Importance Sampling and Conditional Monte Carlo. We discuss briefly these two methods, but before doing so we point that this review is dedicated to static algorithms; however, it is worth mentioning that adaptive techniques have attracted a considerable amount of attention in recent years due to its effectiveness [10, 14, cf.].

Let us start with Importance Sampling. Assume that all the random variables of interest are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This method relies on the existence of a Radon-Nykodym derivative of the original measure with respect to an alternative probability measure: the *importance sampling distribution*. More precisely, suppose that we are interested in estimating $\mathbb{E}[h(W)]$ where W is a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{E}$ is the expectation operator under the measure $\mathbb{P}$. If $\mathbb{Q}$ is an absolutely continuous measure with respect to $\mathbb{P}$, then it holds that

$$\mathbb{E}[g(W)] = \mathbb{E}^{\mathbb{Q}}[L\, g(W)],$$

where $\mathbb{E}^{\mathbb{Q}}$ is the expectation operator under the measure $\mathbb{Q}$ and $L := \mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}$ is the Radon-Nykodym derivative of $\mathbb{P}$ with respect to $\mathbb{Q}$ (the last also goes under the

name of *likelihood ratio* in the stochastic simulation literature). In particular, if the measures $\mathbb{P}$ and $\mathbb{Q}$ are absolutely continuous, then the Radon-Nykodym derivative/likelihood ratio is simply the ratio of the corresponding density functions.

The main idea of importance sampling is that if W is simulated according to the measure $\mathbb{Q}$, then the random variable $L\,h(W)$ has an expected value which is equal to $\mathbb{E}[h(W)]$, hence it is an unbiased estimator of the quantity of interest. The variance of the estimator is clearly altered as the second moment is given by

$$\mathbb{E}^{\mathbb{Q}}[L^2 h^2(W)] = \mathbb{E}[Lh(W)].$$

Observe that importance sampling does not always produce variance reduction as the expressions above are not necessarily bounded by the second moment of the estimator under the original measure. In fact, one can end up with an increased or even an infinite variance if one chooses the wrong importance sampling distribution. To the best of the author's knowledge there does not exist a general methodology for choosing an appropriate importance sampling (there is however, a large number of strategies that can suggest good importance sampling distributions); most of the time the selection is based on the experience of the simulator or other additional information about the quantity of interest.

However, when the focus is in estimating probabilities, there exists a distribution with *zero variance*, which is simply the original distribution restricted to the event of interest, that is $\mathbb{Q}(\mathrm{d}x) := \mathbb{I}_A(x)/\mathbb{P}(A)\mathbb{P}(dx)$. Clearly the Radon-Nykodym derivative/likelihood ratio is given by $L := \mathbb{P}(A)\mathbb{I}_A(\cdot)$. Thus we have

$$\mathbb{E}^{\mathbb{Q}}[L; A] = L\mathbb{E}^{\mathbb{Q}}[\mathbb{I}(A)] = \mathbb{P}(A).$$

while for the second moment of the estimator we obtain that

$$\mathbb{E}^{\mathbb{Q}}[L^2; A] = \mathbb{P}^2(A)\mathbb{E}^{\mathbb{Q}}[\mathbb{I}_A] = \mathbb{P}^2(A).$$

From the last expression it follows that this estimator has variance 0. At first sight, this observation might appear of no practical use as the implementation of the zero variance estimator is unfeasible since it requires the knowledge of the unknown probability of interest $\mathbb{P}(A)$. However, the zero variance distribution is of great theoretical interest as one can obtain partial information about it an serve as the ideal model when choosing an appropriate distribution; that is a distribution which is as "close" as possible to the zero variance distribution. Intuitively, we would like to choose a distribution in such way that the "important" event A is sampled with higher frequency with respect to the original distribution. However, there is a natural trade-off in the final value of the variance for the new estimator because if we increase the frequency of any subset it would also increase the values of its likelihood ratios. Therefore, the selection of the importance sampling requires a conscious analysis. In fact, a considerable amount of research effort in rare event simulation has been devoted to approximating the zero variance distribution. One of the most prominent cases is that of the Cross-Entropy method, which consists of an iterative method which selects an "optimal" distribution from a parametric family by minimizing the Kullback-Leibler distance with respect to the zero variance distribution. Another prominent case is that of *Exponential Change of Measure* or *Exponential Twisting* where the importance sampling distribution is selected from the so called exponential family generated by the original distribution. The later technique will be discussed in some detail in the following section.

The second variance reduction technique that will be discussed here is *Conditional Monte Carlo*. This is perhaps the most general variance reduction technique and the one requiring more theoretical effort. The intuitive idea behind it is that the variance of a given estimator can be reduced by *extracting* the variability coming from known information. If we add a little bit more of rigor to this idea we simply end up with conditional expectation. Let us consider again a random variable W defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, h an arbitrary function and $\mathcal{G}$ a simulatable sub-σ-algebra of $\mathcal{F}$. Then

$$\mathbb{E}[g(W)] = \mathbb{E}[\mathbb{E}[g(W)|\mathcal{G}]],$$

and in consequence $\mathbb{E}[g(W)|\mathcal{G}]$ is an unbiased estimator of the quantity of interest $\mathbb{E}[g(W)]$. This estimator is unbiased and, in most practical cases one could easily verify the conditions of the Rao-Blackwell Theorem; in such case, the variance is always smaller or equal than the one of the original estimator. The implementation of this algorithm is more involved when compared to other variance reduction methods as it requires two critical steps, 1) simulating from $\mathcal{G}$ and 2) computing explicitly the random variable $\mathbb{E}[g(W)|\mathcal{G}]$. Obviously it is also desired that the resulting estimator provides a substantial variance reduction, and for achieving that, the sub-σ-algebra $\mathcal{G}$ should contain as much information about the occurrence of the event of interest as possible.

Most of the estimators discussed in this paper rely on Conditional Monte Carlo method. A variety of examples will be provided in the following section to help clarifying the use of this method.

4. Main Results

We will be interested in the tail probabilities of a sum of random variables. More precisely,

$$\mathbb{P}(X_1 + \cdots + X_N > u),$$

where $X_1, X_2, \ldots$ is a sequence of random variables and N is possibly random (most algorithms condition on the random number N and then employ a method for simulating the tail probability for a fixed number of random variables). The case of independent and *light-tailed* random variables is well understood via the theory of Large Deviations. Moreover, in terms of Monte Carlo simulation, the standard variance reduction method is Importance Sampling with an *exponential change of measure*. That consists in selecting an importance sampling distribution from a family of probability measures consisting of the normalized measures $F_\theta(\mathrm{d}x) := \mathrm{e}^{\theta x} F(\mathrm{d}x)$ for all possible values of θ in the domain of convergence $\Theta := \{\theta : \mathbb{E}[\mathrm{e}^{\theta X}] < \infty, X \sim F\}$. For estimating $\mathbb{P}(X_1 + \cdots + X_n > u)$ where the X_i's are nonnegative and independent random variables with common distribution F, the importance sampling distribution and its associated parameter θ are chosen in such way that $\mathbb{E}_\theta[X] = u$. From our discussion of light and heavy tails it follows that such parameter always exists in the light-tailed case for all values of u. This selection is asymptotically optimal as Large Deviations results can be used to prove that it converges to the zero variance importance sampling [26, 12, 25, 24]. Moreover, it is known that in the light-tailed case, an exponential change of measure delivers a logarithmically efficient estimator.

However, in the heavy-tailed case the domain of convergence Θ is reduced to the set of non positive values of θ, and therefore we only hope to find a solution of

$\mathbb{E}_\theta[X] = u$ when $u \leq \mathbb{E}[X]$; hence, it is clear that it is not possible to implement an optimal exponential change of measure for large values of u. The (nowadays considered) seminal paper [2] presented a number of examples which further exhibited the inherent difficulty in designing good estimators for probabilities of rare events involving heavy-tailed random variables and the challenges of demonstrating their efficiencies. This paper triggered an intense research activity devoted to rare event simulation of heavy-tailed random variables; during the last fifteen years we have seen a wide variety of new developments including estimators for ad hoc applications, novel simulation methodologies targeting rare events and theoretical advances which allowed to simplify some efficiency proofs. As a result, the literature is quite vast. In the following we list a few notorious early works in the area which exemplify the main ideas that one could find in this expanding area of research. A prevalent idea in most of these works is the exploitation of the principle of the single large jump, either by proposing importance sampling distributions which increase the frequency of single big jumps or conditioning in such a way that the conditional probability of a single big jump can be explicitly calculated.

The first logarithmic efficient algorithm was proposed in [1] for the regularly varying case. This is a Conditional Monte Carlo estimator and based on order statistics. The reference [2] proposes a variant of the latter estimator and proves logarithmic efficiency for regularly varying and the lognormal case. A conditional algorithm, similar to that of [1] was proposed in [4]. That algorithm exploits a symmetry relation of random variables which are i.i.d. and the conditioning involves the lower order statistic. It is proved that this estimator has the stronger bounded relative error efficiency property in the regularly case and it is numerically superior to many similar algorithms. It was later proved in [17] that it also achieves bounded relative error in the lognormal case. An independent proof was provided in [5] and reported in [22]; an extension of this is given in Theorem 4.3 in this review. The performance of this algorithm has been improved over time but it continues to be used as a benchmark of performance for similar algorithms. On the importance sampling front, [18] developed a novel methodology where an importance sampling distribution is selected according to a criteria involving the hazard rate function; accordingly, this method is called *hazard rate twisting*. Early examples of adaptive algorithms include [14] which proposed a state-dependent algorithm for the regularly varying case having bounded relative error. Also [10] proposed a state-dependent algorithm in a queueing context and proved that their estimator has vanishing relative error when applied to a $GI/G/1$ queue for a large class of heavy-tailed distributions.

Here, we will concentrate in the conditional algorithms proposed in [1] and [4]. At the end of this section we include some extended results. Moreover, several algorithms studied in this dissertation build on these early ideas.

(4.1) Conditional Monte Carlo Methods Based on Order Statistics. In this subsection we discuss the algorithms designed by [1] and [4]. As mentioned previously, this algorithms exploit the principle of the single big jump by using order statistics. The idea is neat and simple as one can calculate explicitly the probability that the maximum alone is responsible for the large value of the sum by conditioning on the remaining order statistics. Although the algorithms in [1] and

[4] are both based on this idea, they differ in the way of conditioning. The original idea appeared in [1] but the modified version in [4] provided a more efficient and easier to implement algorithm. We complement these ideas by providing extensions with the corresponding proofs of efficiency. Let us first start with the Asmussen-Binswanger estimator:

We assume that $\{X_1,\dots,X_n\}$ is a collection of i.i.d. heavy-tailed random variables. The idea is to simulate the first $n-1$ order statistics out of n. The procedure is simple as we just simulate $X_1,\dots,X_n$ and discard the largest one. The clever idea here is that we can now calculate explicitly the conditional probability of the rare event $\{S_n \geq u\}$ given the order statistics. This comes out as the following random variable

$$\mathbb{P}\big(S_n > x \,\big|\, X_{(1)},\dots,X_{(n-1)}\big) \;=\; \frac{\overline{F}((x-S_{(n-1)})\vee X_{(n-1)})}{\overline{F}(X_{(n-1)})},$$

where $S_{(n-1)} = X_{(1)}+\cdots+X_{(n-1)}$. This algorithm is logarithmic efficient in the regularly varying [8] and the lognormal case [1]. However, we can easily drop the *identically distributed* assumption. When simulating the order statistics, we just need to keep track of the (random) index of the largest random variable, say K. The conditioning will deliver instead

$$\mathbb{P}\big(S_n > x \,\big|\, X_{(1)},\dots,X_{(n-1)}\big) \;=\; \frac{\overline{F}_K\big(X_{(n-1)}\vee(x-S_{n-1})\big)}{\overline{F}_K\big(X_{(n-1)}\big)}, \tag{4.1}$$

where $F_k(\cdot)$ is the distribution function of the k-th random variable. Clearly, the random variable above is unbiased. Moreover, we prove that it has logarithmic efficiency when all random variables are independent lognormals but not necessarily identically distributed.

THEOREM (4.2). *Let $X_1,\dots,X_n$ be independent lognormal random variables. Then the estimator (4.1) is logarithmic efficient.*

The proof of this theorem is slightly technical and relegated to the appendix. This result can be further extended to the case where the random variable with the heaviest tails is lognormal. The proof of this follows trivially by comparison of the tail asymptotics.

Asmussen-Kroese estimator. A slight tweak in the Asmussen-Binswanger estimator can result in a dramatic variance reduction. The main observation of [4] is that the algorithm above still has a large variability due to the fact that there is a significant large probability of having a big jump among the first $n-1$ order statistics. This probability is dramatically reduced by considering a symmetry argument. The idea is to calculate the probability of the event $\{S_n > x, X_k = M_n\}$ for $k = 1,\dots,n$ and where $M_n = \max\{X_i : i = 1,\dots,n\}$. By symmetry we obtain

$$\mathbb{P}(S_n > x) = n\,\mathbb{P}(S_n > x, X_d = M_n). \tag{4.3}$$

Condition on $\mathcal{F} = \sigma(X_1,\dots,X_{d-1})$ and note that

$$n\,\mathbb{P}(S_d > x, X_d = M_d | X_1,\dots,X_{n-1}) = n\overline{F}(M_{n-1}\vee(x-S_{n-1})).$$

This algorithm has bounded relative error in the regularly varying [4] and the lognormal [5, 17] cases. Moreover, the *identically distributed* assumption can be

dropped by substituting the symmetric argument with

$$\mathbb{P}(S_n > x) = \sum_{k=1}^{n} \mathbb{P}(S_n > x, X_k = M_n).$$

This idea was empirically explored in our technical report [5]. An obvious approach consists in estimating individually each of the terms in the summation above. The resulting estimator has good efficiency properties but requires more computational effort. The following alternative approach delivers much better results. The strategy described can be seen as an hybrid between conditional Monte Carlo and importance sampling where the importance sampling distribution is a mixture. Let $p_k := \mathbb{P}(X_k = M_d)$, the probability that X_k takes the largest value among the X_i's and q_k a discrete probability measure supported over $\{1,2,\ldots,n\}$. Hence

$$\begin{aligned}\mathbb{P}(S_n > x) = \sum_{k=1}^{n} \mathbb{P}(S_n > x | X_k = M_n) p_k \frac{q_k}{q_k} &= \sum_{k=1}^{n} \frac{\mathbb{P}(S_n > x, X_k = M_n)}{q_k} q_k \\ &= \mathbb{E}\Big[\frac{\mathbb{I}(S_n > x, X_K = M_n)}{q_K}\Big].\end{aligned}$$

where K is distributed according to q_k. Further if we condition with respect to the sub-σ-algebra $\mathcal{F} = \sigma(K, X_1, \ldots, X_{K-1}, X_{K+1}, \ldots, X_n)$ we obtain

$$\mathbb{P}(S_n > x) = \mathbb{E}\Big[\frac{\overline{F}_K(M_{-K} \vee (x - S_{-K}))}{q_K} \,\Big|\, \mathcal{F}\Big], \tag{4.4}$$

where F_k is the distribution of the k-th random variable and M_{-k} and S_{-k} are defined as the maximum and sum of the X_i's without considering the k-th random variable. The convenient election of the q_k's should deliver a significant variance reduction. Intuitively, this should be minimized if we choose $q_k^* := P(X_k = M_n | S_n > u)$. That is, the probability that the k-th random variable is largest conditioned to the rare event. However, this probability is not available beyond the independent case. Our suggestion is to use

$$q_k(u) = \frac{\mathbb{P}(X_k > u)}{\sum_{i=1}^{n} \mathbb{P}(X_i > u)}.$$

Empirically, we have verified that this proposal approaches the value of the p_k's as $u \to \infty$. Hence, it is conjectured that asymptotically these are equivalent. Moreover, this estimator

$$\frac{\overline{F}_K(M_{-K} \vee (x - S_{-K}))}{p_K} \tag{4.5}$$

delivers excellent numerical results with little computational effort. Efficiency proofs for nonidentical and independent random variables in the lognormal and regularly varying cases are given in the next two Theorems and their proofs can be found in the appendix.

THEOREM (4.6). *Let $X_1, X_2, \ldots, X_n$ be independent lognormal random variables, K a discrete random variable supported over $\{1,\ldots,n\}$. Then (4.5) is an unbiased estimator of $\mathbb{P}(S_n > x)$ with bounded relative error.*

THEOREM (4.7). *Let $X_1, X_2, \ldots, X_n$ be independent regularly varying random variables with indexes α_i respectively, K a discrete random variable supported over*

$\{1,\dots,n\}$. *Then* (4.5) *is an unbiased estimator of* $\mathbb{P}(S_n > x)$ *with bounded relative error.*

5. Conclusions

Calculating the tail probability of a sum of random variables is a fundamental problem in applied probability. In particular, having sharp approximations of these tail probabilities is of key importance in several disciplines. While in the most common cases this problem is tackled with standard methods, it turns out that it is very challenging to deal with random variables which posses heavy tails. In this paper we provide a review on the standard theory of heavy-tails and subexponentiality; we studied the alternative definitions for heavy-tails and provided a glimpse of their main properties. In particular, we paid attention to the subexponential-type property which is characterized for the principle of the single big jump. One of my main contributions to this area was to demonstrate that this behavior goes beyond the independent case by proving that a collection of correlated lognormals posses the subexponential-type property. This result provides an asymptotic equivalent expression for the tail probability of a sum of correlated lognormals which can be used as an approximation of the real probability.

However, in most applications it is desired to have a better precision. Some of the most precise and reliable methods to obtain approximations is via the Monte Carlo method. In particular, the subarea known as Rare-event simulation is devoted to develop the methodologies for delivering sharp approximations. The second part of this review is dedicated to these aspects. We discussed the main techniques and introduced the efficiency concepts used to assess the theoretical performance of estimators for rare event probabilities. This area of research is quite extensive. However, for this review I decided to focus on Conditional Monte Carlo for independent but not necessarily identically distributed random variables. The original estimators are now considered standard for the i.i.d. case but nevertheless we were able to extend these results by dropping the identically distributed assumption. The efficiency proofs in Theorems 4.1, 4.2 and 4.3 appear in my Phd thesis but otherwise this is unpublished material.

6. Appendix: Proofs

Proof of Theorem (4.2). In order to characterize the dominant tail behavior we define

$$\sigma^2 = \max_{1\le k\le d} \sigma_k^2, \quad \mu = \max_{k:\sigma_k^2=\sigma^2} \mu_k,$$

and let F be the distribution of a lognormal random variable with parameters μ and σ. Note that the index K is a discrete random variable supported over $\{1,\dots,d\}$, so we can simplify our proof using the following inequality

$$\mathbb{E}\big[\widehat{z}_{AB}^2\big] = \mathbb{E}\Big[\frac{\overline{F}_K^2\big(X_{(d-1)} \vee (u - S_{d-1})\big)}{\overline{F}_K^2\big(X_{(d-1)}\big)}\Big] \le \sum_{k=1}^{d} \mathbb{E}\Big[\frac{\overline{F}_k^2\big(X_{(d-1)} \vee (u - S_{d-1})\big)}{\overline{F}_k^2\big(X_{(d-1)}\big)}\Big].$$

The idea is to obtain an asymptotic upper bound for the expectation for a fixed k. Then we break this expectation in two pieces as follows

$$\mathbb{E}\Big[\frac{\overline{F}_k{}^2\big((u-S_{(d-1)})\vee X_{(d-1)}\big)}{\overline{F}_k{}^2(X_{(d-1)})}\Big]=\mathbb{E}\Big[\frac{\overline{F}_k^2\big((u-S_{(d-1)})\vee X_{(d-1)}\big)}{\overline{F}_k^2\big(X_{(d-1)}\big)};X_{(d-1)}<\frac{u}{d}\Big]$$
$$+\mathbb{E}\Big[\frac{\overline{F}_k{}^2((u-S_{(d-1)})\vee X_{(d-1)})}{\overline{F}_k{}^2(X_{(d-1)})};X_{(d-1)}>\frac{u}{d}\Big].$$

The quotient inside the second expectation is always smaller than 1, so we can bound the whole expectation with $\mathbb{P}\big(X_{(d-1)}>u/d\big)$. For the first expectation, it will be useful to note that if $X_{(d-1)}<u/d$ then the following inequalities hold

$$u-S_{(d-1)}\geq u-(d-1)X_{(d-1)}\geq u-\frac{d-1}{d}u=u/d\geq X_{(d-1)}.$$

This implies that in the event $\{X_{(d-1)}<u/d\}$, the following inequality holds true as well

$$\overline{F}_k\big((u-S_{(d-1)})\vee X_{(d-1)}\big)\leq\overline{F}_k(u/d).$$

Inserting these bounds in the expectations we arrive at the following upper bound

$$\mathbb{E}\Big[\frac{\overline{F}_k^2(u/d)}{\overline{F}_k^2(X_{(d-1)})};X_{(d-1)}<\frac{u}{d}\Big]+\mathbb{P}\big(X_{(d-1)}>u/d\big). \tag{6.1}$$

We concentrate on the expectation in the last term. Since $X_{(n-1)}<u/d$ we can apply Lemma (6.3) to get a bound for the quotient in the first expectation to obtain

$$c\,\mathbb{E}\Big[\frac{\overline{F}^2(u/d)}{\overline{F}^2(X_{(d-1)})};X_{(d-1)}<\frac{u}{d}\Big]=c\,\overline{F}^2(u/d)\mathbb{E}\Big[\frac{1}{\overline{F}^2(X_{(d-1)})};X_{(d-1)}<\frac{u}{d}\Big],$$

where c is a constant (recall that F was defined as the distribution with the dominant tail). Letting $F_{(d-1)}$ and $f_{(d-1)}$ be the distribution and density functions of $X_{(d-1)}$ respectively, we rewrite this expectation in integral form and use partial integration to obtain

$$\int_0^{u/d}\frac{f_{(d-1)}(y)}{\overline{F}^2(y)}dy=-\frac{\overline{F}_{(d-1)}(y)}{\overline{F}^2(y)}\Big|_0^{u/d}+2\int_0^{u/d}\frac{\overline{F}_{(d-1)}(y)f(y)}{\overline{F}^3(y)}dy$$
$$=1-\frac{\overline{F}_{(d-1)}(u/2)}{\overline{F}^2(u/2)}+2\int_0^{u/d}\frac{\overline{F}_{(d-1)}(y)}{\overline{F}^2(y)}\frac{f(y)}{\overline{F}(y)}dy.$$

We get a new upper bound by just ignoring the negative term. For dealing with integral it will be useful to note that

$$\frac{\overline{F}_{(d-1)}(t)}{\overline{F}^2(t)}\leq\frac{\sum_{i\neq j}\overline{F}_i(t)\overline{F}_j(t)}{\overline{F}^2(t)}=\mathrm{O}(1),\qquad(0,\infty). \tag{6.2}$$

This is true since the $\overline{F}$ has the heaviest tail so it dominates all $\overline{F}_k$'s, and the quotient remains bounded as $y\to\infty$. Trivially, the same holds true as $y\to 0$.

Then, by a continuity argument this quotient remains bounded all over $(0,\infty)$ by a constant, say $c_1 > 0$. We use this to obtain a new upper bound

$$1 + c_1 \int_0^{u/d} \frac{f(t)}{\overline{F}(t)} dy = 1 - c_1 \log \overline{F}(u/d).$$

Inserting this new bound in (6.1) we have obtained a new bound for $\mathbb{E}\hat{z}^2_{AB}(u)$ which has the following shape

$$c\overline{F}^2(u/d)\big[1 - c_1 \log \overline{F}(u/d)\big] + \overline{F}_{(d-1)}(u/d) \leq c_2 \overline{F}^2(u/d)\big[1 - c_1 \log \overline{F}(u/d)\big],$$

where the last inequality was obtained by using the argument (6.2). So, to prove logarithmic efficiency we need

$$\lim_{u\to\infty} \frac{\mathbb{E}\widehat{z}_{AB}(u)}{\mathbb{P}^{2-\epsilon}(S_d > u)} \leq \lim_{u\to\infty} \frac{c_2 \overline{F}^2(u/d)\big[1 - c_1 \log \overline{F}(u/d)\big]}{\overline{F}^{2-\epsilon}(u)} = 0.$$

Using *Mill's ratio* and some basic calculus it is provable that the last limit is zero for all $\epsilon > 0$. By doing this the proof is complete. □

LEMMA (6.3). *Let F_1 and F_2 lognormal distributions such that F_2 has a heavier tail than F_1. Then, there exists $c \in \mathbb{R}$ such that for all $y \leq x$ it holds that*

$$\frac{\overline{F}_1(x)}{\overline{F}_1(y)} \leq c \frac{\overline{F}_2(x)}{\overline{F}_2(y)}.$$

Proof. Let $\lambda_1(x), \lambda_2(x)$ the corresponding failure rate functions of the lognormal distributions F_1 and F_2. First we will prove that there exist constants $c_1 > 0$ and $y_0 > 0$ such that the following inequality is true

$$-\lambda_1(t) \leq -\lambda_2(t) + c_1 \mathbb{I}_{[0,y_0]}(t).$$

For proving this, we will start from the inequality

$$\begin{aligned}[\lambda_1(t) - \lambda_2(t)]^+ &= \lambda_1(t) - \lambda_2(t) + [\lambda_2(t) - \lambda_1(t)]\, \mathbb{I}_{\{t:\lambda_1(t)<\lambda_2(t)\}}(t) \\ &\leq \lambda_1(t) - \lambda_2(t) + \lambda_2(t) \mathbb{I}_{\{t:\lambda_1(t)<\lambda_2(t)\}}(t),\end{aligned}$$

from where it follows that

$$-\lambda_1(t) \leq -\lambda_2(t) + \lambda(t) \mathbb{I}_{\{t:\lambda_1(t)<\lambda_2(t)\}}(t).$$

Since $\lambda_2(t)$ is real-valued on closed intervals of the type $[0, y_0]$ it remains bounded in there by continuity. So, it is just necessary to prove that $\{t : \lambda_1(t) < \lambda_2(t)\} \subseteq [0, y_0]$ for some $y_0 \in \mathbb{R}^+$. We consider the two possible cases in which $\overline{F}_1$ has heavier tail than $\overline{F}_2$. In the first of them we consider $\sigma_1 < \sigma_2$. So we use the tail asymptotic expression for $\lambda(x)$ to obtain

$$\lim_{x\to\infty} \frac{\lambda_1(x)}{\lambda_2(x)} = \lim_{x\to\infty} \frac{\log x / x\sigma_1^2}{\log x / x\sigma_2^2} = \frac{\sigma_2^2}{\sigma_1^2} > 1,$$

from where the conclusion follows easily. The second case comes when $\sigma_1 = \sigma_2$ and $\mu_1 < \mu_2$. For proving that $\lambda_2(x) \leq \lambda_1(x)$ we will just check that $\lambda(x, \mu)$ is a

decreasing of function of μ. The derivative is given as

$$\frac{d}{d\mu}\lambda(x,\mu) = \frac{\frac{\log x-\mu}{\sigma^2}f(x,\mu)\overline{F}(x,\mu) - f(x,\mu)\int_x^\infty \frac{\log t-\mu}{\sigma^2}f(t,\mu)dt}{\overline{F}^2(t,\mu)}$$
$$= \frac{\log x f(x,\mu)\overline{F}(x,\mu) - f(x,\mu)\int_x^\infty \log t f(t,\mu)dt}{\sigma^2\overline{F}^2(t,\mu)}.$$

The last expression is verified to be negative from the observation

$$\int_x^\infty \log t f(t,\mu)dt > \log x\int_x^\infty f(t,\mu)dt = \log x\overline{F}(x).$$

Then we just use this intermediate result to prove that

$$\frac{\overline{F}_1(x)}{\overline{F}_1(y)} = \exp\Big\{-\int_y^x \lambda_1(t)dt\Big\} \le \exp\Big\{-\int_y^x \lambda_2(t)dt + \int_y^x c_1\mathbb{I}_{[0,y_0]}(t)dt\Big\}$$
$$\le \exp\Big\{-\int_y^x \lambda_2(t)dt + \int_0^{y_0} c_1 dt\Big\}$$
$$= \exp\Big\{\log\frac{\overline{F}_2(x)}{\overline{F}_2(y)} + c_2\Big\} = c\frac{\overline{F}_2(x)}{\overline{F}_2(y)}. \qquad \square$$

Proof of Theorem (4.6). Recall that the condition for *asymptotic bounded relative error* is equivalent to

$$\lim_{u\to\infty}\frac{\mathbb{E}[\hat{z}^2_{AK}(u)]}{\mathbb{P}(S_d>u)} < \infty.$$

By subexponentiality we have that $\mathbb{P}(X_k>u) = \mathrm{O}(\mathbb{P}(S_n>u))$ for all k. Using this relation and the fact that all p_i's are all larger than 0 it will be enough to prove that

$$\limsup_{u\to\infty}\frac{\overline{F}^2_k(M_{-k}\vee(u-S_{-k}))}{\mathbb{P}^2(X_k>u)} < \infty \qquad k=1,\dots,d.$$

The idea will be to provide an upper bound where we get rid of the random variable S_{-k} since its distribution is unknown to us. For doing so, we divide the sample space in two events, namely $A_1=\{M_{-k}\le u/2d\}$ and $A_2=\{M_{-k}>u/2d\}$, and note that in A_1 the following relations hold

$$u-S_{-k}\ge u-nM_{-k}\ge u-u/2 = u/2 > u/2d \ge M_{-k}.$$

Using this we can obtain an upper bound in terms of M_d only

$$\frac{\mathbb{E}[\overline{F}^2_k(M_{-k}\vee(u-S_{-k}))]}{\overline{F}^2_k(u)} \le \mathbb{E}\Big[\frac{\overline{F}^2_k(u-nM_{-k})}{\overline{F}^2_k(u)};M_{-k}<u/2d\Big]$$
$$+\mathbb{E}\Big[\frac{\overline{F}^2_k(M_{-k})}{\overline{F}^2_k(u)};M_{-k}>u/2d\Big].$$

So, with a simple change of variables we can rewrite this expression in integral form as follows

$$\int_0^{u/2} \frac{\overline{F}_k^2(u-y)}{\overline{F}_k^2(u)} f_{M_{-k}}(y/d)dy + \int_{u/2d}^{\infty} \frac{\overline{F}_k^2(y)}{\overline{F}_k^2(u)} f_{M_{-k}}(y)dy.$$

The advantage of this bound is that the density of M_{-k} is known to us. In fact, this density is always smaller than the sum of the individual densities as can be seen from the following expression

$$f_{M_{-k}}(\cdot) = \sum_{i\neq k} f_i(\cdot) \prod_{j\neq i,k} F_j(\cdot) \leq \sum_{i=1}^{d} f_i(\cdot).$$

Inserting this new bound and taking the sum out of the integral we arrive to the conclusion that the estimator will have bounded relative error if

$$\limsup_{u\to\infty} \int_0^{u/2} \frac{\overline{F}_k^2(u-y)}{\overline{F}_k^2(u)} f_i(y/d)dy + \int_{u/2d}^{\infty} \frac{\overline{F}_k^2(y)}{\overline{F}_k^2(u)} f_i(y)dy < \infty, \qquad i,k = 1,\dots,d. \tag{6.4}$$

We prove separately that each of this two integrals remain bounded as $u \to \infty$. The first integral remains bounded due to Lemma (6.5). The second one is the easy since it can be evaluated directly using L'Hopital Theorem,

$$\lim_{u\to\infty} \frac{\int_{u/2d}^{\infty} \overline{F}_k^2(y) f_i(y)dy}{\overline{F}_k^2(u)} = \lim_{u\to\infty} \frac{\overline{F}_k^2(u/2d) f_i(u/2d)}{4d\,\overline{F}_k(u) f_k(u)} \to 0.$$

This limit can be easily verified using *Mill's ratio*. Putting together these results the result follows immediately. □

LEMMA (6.5). *Under the hypothesis of the Theorem (4.6) it holds that*

$$\lim_{u\to\infty} \int_0^{u/2} \frac{\overline{F}_k^2(u-y)}{\overline{F}_k^2(u)} f_i(y/d)dy < \infty.$$

Proof. Consider $\overline{F}_k(u) = \exp\{-\int_0^u \lambda(t)dt\}$, where $\lambda(t)$ is the failure rate of the lognormal distribution and by standard subexponential theory we know that $\lambda(t)$ is asymptotically equivalent to $\frac{\log(u)}{\sigma^2 u}$. By choosing $c > \frac{1}{\sigma}^2$ we obtain that $c\frac{\log t}{t}$ is an asymptotic upper bound for $\lambda(t)$, then

$$\frac{\overline{F}_k(u-y)}{\overline{F}_k(u)} = \exp\Big\{\int_{u-y}^{u} \lambda(t)dt\Big\} < \exp\Big\{c\log u \int_{u-y}^{u} \frac{1}{t}dt\Big\}$$
$$= \exp\Big\{c\log u(\log u - \log(u-y))\Big\}.$$

Using a first order Taylor expansion of $\log(\cdot)$ around $(u-y)$ and the fact that it is a concave function we have that $\log u < \log(u-y) + \frac{y}{u-y}$, so the last expression is bounded by

$$\exp\Big\{c\,\frac{y\log u}{u-y}\Big\}.$$

Take $u > 1$. Our claim is that the set $\{y : \log(2y) > \frac{y \log u}{u-y}\} = (g(u), u/2)$ for some function $g(u) \to 1/2$. This is true since both functions are increasing and equal when $y = u/2$, but $\log(2y)$ is concave and $\frac{y \log u}{y-u}$ is convex proving that there exists a smaller root than $u/2$. Next we verify that for any value $y_0 > 1/2$ there exists a value u_0 such that for all $u > u_0$ the inequality $\log(2y_0) > \frac{y_0 \log u}{u-y_0}$ is fulfilled and therefore $g(u) < y_0$. We use this to get

$$\int_{y_0}^{u/2} \frac{\overline{F}_k^2(u-y)}{\overline{F}_k^2(u)} f_i(y/d)dy < \int_{y_0}^{\infty} c_1 \exp\{c \log y\} f_i(y/d)dy = \int_{y_0}^{\infty} c_2\, y^c f_i(y/d)dy.$$

Since all the moments of a lognormal rancom variable are bounded we can conclude that the last expression is also bounded. For $y \in (0, y_0)$ we simply use the fact that a lognormal random variable belongs to the class $\mathcal{L}$, so we obtain

$$\int_0^{y_0} \frac{\overline{F}_k^2(u-y)}{\overline{F}_k^2(u)} f_i(y/d)dy < \frac{\overline{F}_k^2(u-y_0)}{\overline{F}_k^2(u)} \to 1. \qquad \square$$

Proof of Theorem (4.7). Note that in the proof of Theorem (4.6) we did not make use of the hypothesis about the distribution up to (6.4). Hence, we can retake the proof from there so it remains to prove that the same holds for regularly varying distributions. That is

$$\int_0^{u/2} \frac{\overline{F}_k^2(u-y)}{\overline{F}_k^2(u)} f_i(y/d)dy + \int_{u/2d}^{\infty} \frac{\overline{F}_k^2(y)}{\overline{F}_k^2(u)} f_i(y)dy < \infty,$$

where F_k is a regular varying distribution function with index α_k and f_i are densities of regularly varying random variables with indexes α_i. The first integral can be easy bounded with

$$\frac{\overline{F}_k^2(u/2)}{\overline{F}_k^2(u)} = 2^{-2\alpha_k} + o(1) \qquad u \to \infty.$$

For the second one we can use L'Hopital rule to obtain

$$\frac{\overline{F}_k^2(u/2d)}{\overline{F}_k^2(u)} f_i(u/2d) = (2d)^{2\alpha_k} f_i(u/2d) = o(1).$$

Putting together these two expressions we complete the proof. $\square$

Acknowledgments

The author would like to thank the editors and the referees. The author is supported by the ARC grant DE130100819.

Received June 20, 2013

Final version received October 24, 2013

LEONARDO ROJAS-NANDAYAPA
SCHOOL OF MATHEMATICS AND PHYSICS,
THE UNIVERSITY OF QUEENSLAND,
4072 BRISBANE, AUSTRALIA,
l.rojas@uq.edu.au

REFERENCES

[1] S. ASMUSSEN AND K. BINSWANGER, *Simulation of ruin probabilities for subexponential claims*, ASTIN Bulletin, **27**, (1997), 297–318.

[2] S. ASMUSSEN, K. BINSWANGER, AND B. HØJGAARD, *Rare events simulation for heavy-tailed distributions*, Bernoulli, **6**, (2000), 303–322.

[3] S. ASMUSSEN AND P. W. GLYNN, *Stochastic Simulation: Algorithms and Analysis*, Springer-Verlag, New York, 2007.

[4] S. ASMUSSEN AND D. P. KROESE, *Improved algorithms for rare event simulation with heavy tails*, Advances in Applied Probability, **38**, (2006), 545–558.

[5] S. ASMUSSEN AND L. ROJAS-NANDAYAPA, *Sums of dependent lognormal random variables: Asymptotics and simulation*, in Thiele Research Reports, Aarhus, DK, 2006, Thiele Centre.

[6] ———, *Asymptotics of sums of lognormal random variables with Gaussian copula*, Stat. Probab. Lett., **78**, (2008). To appear.

[7] N. H. BINGHAM, C. M. GOLDIE, AND J. L. TEUGELS, *Regular Variation*, Cambridge University Press, Cambridge, 1987.

[8] K. S. BINSWANGER, *Rare Events and Insurance*, PhD thesis, ETH, Zürich, 1997. 12233.

[9] E. S. BLAKE, C. W. LANDSEA, AND E. J. GIBNEY, *The deadliest, costliest, and most intense united states tropical cyclones from 1851 to 2010 (and other frequently requested hurricane facts)*, tech. report, National Oceanic and Atmospheric Administration, 2011.

[10] J. BLANCHET AND P. W. GLYNN, *Efficient rare-event simulation for the maximum of a heavy-tailed random walk*, Ann. of Appl. Probab., **18**, (2008), 1351–1378.

[11] J. H. BLANCHET AND L. ROJAS-NANDAYAPA, *Efficient simulation of tail probabilities of sums of dependent random variables*, Journal of Applied Probability, **48A**, (2011), 147–164.

[12] J. A. BUCKLEW, J. A. NEY, AND P. SADOWSKY, *Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains*, J. Appl. Prob., **27**, (1990), 44–59.

[13] V. P. CHISTYAKOV, *A theorem of sums of independent positive random variables and its applications to branching processes*, Theory of Probability and its Applications, **9**, (1964), 640–648.

[14] P. DUPUIS, K. LEDER, AND H. WANG, *Importance sampling for sums of random variables with regularly varying tails*, ACM TOMACS, **17**, (2006), 1–21.

[15] P. EMBRECHTS, C. KLÜPPELBERG, AND T. MIKOSCH, *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, New York, 1997.

[16] S. FOSS, D. KORSHUNOV, AND S. ZACHARY, *An introduction to heavy-tailed and subexponential distributions*, Springer, 2011.

[17] J. HARTINGER AND D. KORTSCHAK, *On the efficiency of the Asmussen-Kroese-estimator and its applications to stop-loss transform*, in Proceedings of the 6th International Workshop on Rare Event Simulation, Bamberg, Germany, 2006, RESIM.

[18] S. JUNEJA AND P. SHAHABUDDIN, *Simulating heavy-tailed processes using delayed hazard rate twisting*, ACM TOMACS, **12**, (2002), 94–118.

[19] P. L'ECUYER, J. BLANCHET, B. TUFFIN, AND P. W. GLYNN, *Asymptotic robustness of estimators in rare-event simulation*, ACM TOMACS, , (2008). To appear.

[20] T. MIKOSCH, *Non-Life Insurance Mathematics*, Springer-Verlag, Berlin, 2009.

[21] S. I. RESNICK, *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer-Verlag, New York, 2006.

[22] L. ROJAS-NANDAYAPA, *Risk probabilities: Asymptotics and Simulation*, PhD thesis, Aarhus University, 2008.

[23] L. ROJAS-NANDAYAPA AND S. ASMUSSEN, *Efficient simulation of finite horizon problems in queueing and insurance risk*, Queueing Systems, **57**, (2007), 85–97.

[24] J. S. SADOWSKY, *On Monte Carlo estimation of large deviations probabilities*, Ann. Appl. Prob., **6**, (1996), 399–422.

[25] J. S. SADOWSKY AND J. A. BUCKLEW, *On large deviations theory and asymptotically efficient monte carlo estimation*, IEEE Transactions on Information Theory, **36**, (1990), 579–588.

[26] D. SIEGMUND, *Importance sampling in the Monte Carlo study of sequential tests*, Ann. Stat., **3**, (1976), 673–684.

Bol. Soc. Mat. Mexicana (3) Vol. 19, 2013

UNIFORM LARGE DEVIATIONS FOR HEAVY-TAILED QUEUES UNDER HEAVY TRAFFIC

JOSE BLANCHET AND HENRY LAM

ABSTRACT. We provide a complete large and moderate deviations asymptotic for the steady-state waiting time of a class of subexponential $M/G/1$ queues under heavy traffic. The asymptotic is uniform over the positive axis, and reduces to heavy-traffic asymptotics and heavy-tail asymptotics on two ends, both of which are known to be valid over restricted asymptotic regimes. The link between these two well-known asymptotics is a transition term that is expressible as a convolution-type integral. The class of service times that we consider includes regularly varying and Weibull tails in particular.

It is our pleasure to contribute to this special issue dedicated to the International Year of Statistics. In response to the request of the editors of this special issue we briefly overview the research topics that we have investigated recently. Our research group has pursued several themes in recent years. All of them lie under the scope of applied probability. Some of our projects deal with computational probability. In this context, our goal is to enable efficient computation in stochastic systems using (and often developing) theory of probability to inform the design of algorithm that are optimal and robust in certain sense (see Blanchet and Glynn (2008)). Most of the computations that we study relate to stochastic simulation (also known as Monte Carlo) methods (see Blanchet and Lam (2012)). Other projects that we pursue relate to classical analysis in probability, such as asymptotic approximations, large deviations, and heavy-traffic limits (Blanchet and Glynn (2006) and Lam et al (2011)). All of our research efforts are motivated by models and problems in areas such as: Finance, Insurance, Operations Research, and Statistics.

Here we shall study a class of asymptotic results that lie at the intersection of large deviations and heavy-traffic limit theory. We use a classical model in queueing theory to illustrate these types of results, namely, the classical $M/G/1$ queue. Despite its apparent tractability, most of the *asymptotics for the steady-state waiting time of the* $M/G/1$ *queue* that have been proposed in the literature are only provably valid in restricted regimes. Among them are the well-known heavy-traffic or Kingman asymptotic (see Kingman (1961)) and the heavy-tail or Pakes-Veraberbeke asymptotic (see for example Embrechts and Veraverbeke (1982)). More precisely, in heavy traffic (i.e. when the long-run proportion of time the server is utilized, ρ, is close to 1) one approximates the distribution of the steady-state waiting time in spatial scales of size $1/(1-\rho)$ by the steady-state distribution of reflected Brownian motion (which is exponential). On the other hand, the heavy-tail asymptotic assumes fixed traffic intensity while the tail parameter

2010 *Mathematics Subject Classification*: 60F10.
Keywords and phrases: uniform large deviations, queueing systems, heavy-tails, heavy-traffic.

increases. It states that for service time with so-called stationary excess distribution (in the tradition of renewal theory) $B_0(x)$ lying in the class, $\mathcal{S}$, of subexponential distribution (see, for example, Embrechts et. al. (2003) and Asmussen (2001, 2003)), the probability that the steady-state waiting time is larger than x is asymptotically $(\rho/(1-\rho))\bar{B}_0(x)$.

In this paper we provide a uniform large deviations asymptotic of the steady-state waiting time distribution for heavy-tailed $M/G/1$ under heavy traffic. Our results can handle the case when heavy traffic is present but the tail parameter level is moderate, which is covered by neither heavy-traffic or heavy-tail asymptotics. Our results in this paper extend and unify previous work by Olvera-Cravioto et. al. (2011) and Olvera-Cravioto and Glynn (2011). In these two papers, the authors studied first the regularly varying $M/G/1$ queue and showed that heavy-traffic and heavy-tail asymptotics remain valid on regimes that are respectively smaller and larger than an explicitly identified transition point. Then, a separate argument is given in Olvera-Cravioto and Glynn (2011) in order to deal with Weibull type distributions. Our framework here provides means to develop a unified theory of transitions from heavy-traffic to heavy-tailed asymptotics that covers both regularly varying and Weibullian tails at once. In addition, and in contrast to Olvera-Cravioto et. al. (2011), for regularly varying distributions we provide an explicit asymptote for the behavior of the tail of the steady-state waiting time in the $M/G/1$ queue right at the transition point in complete generality.

We shall use the machinery developed by Rozovskii (1989, 1993) for large and moderate deviations of random walks. Related papers that develop similar methods include Nagaev (1969) and Borokov and Borovkov (2001). A central argument to obtaining these deviations results is finding a suitable truncation level depending on p. In Section 2 we outline this truncation argument and we provide the details of the proofs in Section 3.

1. Statement of Result and Outline of Argument

Let $(X_i : i \geq 1)$ be a sequence of non-negative i.i.d. r.v.'s (independent and identically distributed random variables) and define $S_n = X_1 + \cdots + X_n$, with $S_0 = 0$. Let M be a geometrically distributed random variable with parameter $p > 0$ and independent of the X_i's. In particular, $P(M = k) = pq^k$ for $k = 0, 1, ...$, where $q = 1 - p$. The random variable S_M is said to be a geometric sum. It turns out (see Asmussen (2003)), that the steady-state waiting time of an $M/G/1$ queue can be represented as geometric sum, is $p = 1 - \rho > 0$.

Throughout the rest of the paper we use $F(x)$ to denote the distribution of X_i and we write X to denote a generic copy of X_i. We are interested in $P(S_M > x)$ as $p \searrow 0$ and $x = x(p) \nearrow \infty$.

We use the following notation. Given non-negative functions $f_1(\cdot)$, $f_2(\cdot)$, we write $f_1(x) \ll f_2(x)$ if $f_1(x) = o(f_2(x))$ as $x \to \infty$ (i.e. "f_1 has smaller order than f_2") and $f_1(x) \gg f_2(x)$ if $f_2(x) = o(f_1(x))$ (i.e. "f_1 has larger order than f_2"). Also we use "$\preceq$" and "$\succeq$" to denote "having order smaller than or equal to" and "having order larger than or equal to", respectively. For example, $f_1(x) \preceq f_2(x)$ means that $f_1(x) \leq c f_2(x)$ for some $c > 0$. Lastly, we use "$\sim$" to denote "asymptotically equivalent or the same order" (i.e. $f_1(x)/f_2(x) \to 1$).

We shall consider X_i in class $\mathcal{S}$ of subexponential distribution (i.e. $P(X_1+X_2>x)/\bar{F}(x)\to 2$) together with the assumption

$$\bar{F}(x) := P(X_i > x) = e^{-g(x)},$$

where $g(\cdot)\geq 0$ is clearly non-decreasing, and $g(x)\to\infty$ as $x\to\infty$. We also assume that $g(\cdot)$ is differentiable, and that $g(x)/x^{\delta}\to 0$ and is eventually decreasing for some $0<\delta<1$. We further assume that $EX=\mu<\infty$, $EX^2=1$ and $EX^{2+\epsilon}<\infty$ for some $\epsilon>0$ (we sometimes drop the subscript i in X_i for convenience). We assume that $(2+\epsilon)\log x\preceq g(x)\ll x$. In addition, we assume that $h(x)=g(x+\mu)-2\log x$ is eventually non-decreasing and goes to ∞, which is intuitive given that $EX^{2+\epsilon}<\infty$. We also assume that $h'(x)/h(x)\leq(\delta+\eta)/x$ for some $\eta>0$ eventually. Finally, we also assume that X_i is strongly non-lattice, in the sense that

$$\inf_{|\omega|>v}|1-\chi(\omega)|>0$$

for any $v>0$, where $\chi(\omega)=Ee^{i\omega X}$ is the characteristic function of X.

Set $B_p=1/p$ and define for $x>0$ and $p\in(0,1)$

$$\Gamma(x,p)=\left[e^{-\theta^* x}+\left(\frac{1}{p}\bar{F}(x)+\int_{B_p}^{x}\left(\frac{1}{p}+\frac{x-y}{\mu}\right)e^{-\theta^*(x-y)}dF(y)\right)I(x\geq B_p)\right], \tag{1.1}$$

where θ^* is the solution to the equation $E[e^{\theta X};X\leq B_p]=1/q$. Our main result is the following:

THEOREM (1.2). *Let $B_p=1/p$. We have uniformly over $x>0$ that*

$$\limsup_{p\to 0}\sup_{x>0}\left|\frac{P(S_M>x)}{\Gamma(x,p)}-1\right|=0.$$

Note that we can as well choose B_p to be any quantity having the same order as $1/p$. Blanchet and Glynn (2007) shows that θ^* admits a Taylor series type expansion $\theta^*=p/\mu+c_2p^2+\cdots$. The expansion is valid up to the order of the moment of X_i. Thus if $EX^2<\infty$ we can ensure that θ^* can be expanded up to the second order of p. Note that this gives $e^{-\theta^* x}\sim e^{-px/\mu}$ for $B_p\preceq x\ll 1/p^2$, which coincides with Kingman's asymptotic. On the other hand, the second term in (1.1) is the heavy-tail asymptotic. It can be shown that the first term is dominant for small order of x (with respect to p) while the second term is dominant for large order. The third term can be the dominant component in a neighborhood of the transition between the first and the second. These observations are apparent through the following example.

EXAMPLE (1.3) (Regularly Varying Tail). *Suppose X_i has density $L(x)/x^{1+\alpha}, x>0$ where $\alpha>2$ and $L(\cdot)$ is a slowly varying function, so $\bar{F}(x)\sim L(x)/x^{\alpha}$. We are interested in computing the third term of (1.1). First we have*

$$\frac{1}{p}\int_{B_p}^{x}e^{-\theta^*(x-y)}\frac{L(y)}{y^{1+\alpha}}dy=\frac{1}{p}\,e^{-\theta^* x}\frac{L(x)}{x^{\alpha}}\int_{B_p/x}^{1}e^{\theta^* xu}\frac{1}{u^{1+\alpha}}\frac{L(ux)}{L(x)}du$$
$$\sim\frac{1}{p}\,e^{-\theta^* x}\frac{L(x)}{x^{\alpha}}\int_{B_p/x}^{1}\frac{e^{\theta^* xu}}{u^{1+\alpha}}du.$$

where the first equality follows by a substitution $y = xu$, and the equivalence relation follows from the property of slowly varying function that $L(ux)/L(x) \to 1$ uniformly over compact set as $x \to \infty$. If $\theta^* x = O(1)$, which implies $x \le C_1/\theta^* \sim C_1\mu/p$ for some constant $C_1 > 0$ (see the proof of Lemma (2.4) for the equivalence $\theta^* \sim p/\mu$), then

$$\int_{B_p/x}^1 \frac{e^{\theta^* xu}}{u^{1+\alpha}} du \le e^{\theta^* x} \int_{B_p/x}^1 \frac{1}{u^{1+\alpha}} du \le C_2 e^{\theta^* x}$$

for some $C_2 > 0$ and so

$$\frac{1}{p}\, e^{-\theta^* x} \frac{L(x)}{x^\alpha} \int_{B_p/x}^1 \frac{e^{\theta^* xu}}{u^{1+\alpha}} du \le \frac{C_2}{p} \frac{L(x)}{x^\alpha} \ll e^{-\theta^* x}.$$

On the other hand, if $\theta^* x \nearrow \infty$, then applying Laplace's method yields

$$\frac{1}{p}\, e^{-\theta^* x} \frac{L(x)}{x^\alpha} \int_{B_p/x}^1 \frac{e^{\theta^* xu}}{u^{1+\alpha}} du \le \frac{1}{p} \frac{L(x)}{x^\alpha} \frac{1}{\theta^* x} \ll \frac{1}{p} \bar{F}(x).$$

Now by the same analysis, and noting that $\theta^* \sim p/\mu$, we obtain that

$$\frac{x}{\mu} \int_{B_p}^x e^{-\theta^*(x-y)} \frac{L(y)}{y^{1+\alpha}} dy \le C_3 \frac{L(x)}{x^{\alpha-1}} \ll e^{-\theta^* x}$$

and

$$\frac{1}{\mu} \int_{B_p}^x e^{-\theta^*(x-y)} \frac{L(y)}{y^{\alpha}} dy \le C_4 \frac{L(x)}{x^{\alpha-1}} \ll e^{-\theta^* x}$$

for $\theta^* x = O(1)$, and

$$\frac{x}{\mu} \int_{B_p}^x e^{-\theta^*(x-y)} \frac{L(y)}{y^{1+\alpha}} dy \sim \frac{1}{\mu} \int_{B_p}^x e^{-\theta^*(x-y)} \frac{L(y)}{y^{\alpha}} dy \sim \frac{1}{p} \frac{L(x)}{x^\alpha}$$

for $\theta^* x \nearrow \infty$, which implies that

$$\int_{B_p}^x \frac{x-y}{\mu}\, e^{-\theta^*(x-y)} \frac{L(y)}{y^{1+\alpha}} dy = o\left(\frac{1}{p} \frac{L(x)}{x^\alpha}\right).$$

Hence we have

$$P(S_M > x) = \left[e^{-\theta^* x} + \frac{1}{p}\bar{F}(x)\right](1 + o(1))$$

for any $x > 0$. This recovers a basic result in Olvera-Cravioto et. al. (2011) and identifies the transition point located at $-((\alpha-2)/2)\log(p)/p$.

We now give a brief outline of our argument leading to Theorem (1.2). Detailed proofs will be provided in the next section. Our method is mainly inspired by Rozovskii (1989, 1993) together with the use of uniform renewal theorem in Blanchet and Glynn (2007). We first find an appropriate truncation for X_i, so that the geometric sum of the truncated part can be approximated by uniform renewal theorem while the remaining part follows the big-jump asymptotic. Uniform renewal theory then yields Kingman's asymptotic. On the other hand, the heavy-tail component will boil down to calculating a convolution of negative binomial sum with the increment distribution.

From now on we will adopt the following notations. Recall that $B_p = 1/p$, and δ satisfies $g(x)/x^\delta \to 0$. This allows us to find $\delta' = \delta + \eta < 1$ for some $\eta > 0$. For

convenience of development, when $g(x) \preceq \log x$, we take $\delta = 0$ and δ' be a small number such that $0 < \delta' < 1$. We then let $K_p = (1/p^{2\delta})e^{(1-\delta')g(B_p)}$, and

$$C_{p,M} = \begin{cases} B_p & \text{for } M \le K_p \\ \mu + \sqrt{M} & \text{for } M > K_p \end{cases}$$

Let us state the result on the split into truncated and remaining part:

PROPOSITION (1.4).

$$\begin{aligned} P(S_M > x) \quad &= \quad \left[P\left(S_M > x, \max_{1\le i\le n} X_i \le C_{p,M} \right) \right. \\ &\left. + P\left(S_M > x, \bigcup_{i=1}^{n} \left\{ X_i > C_{p,M}, \max_{j\ne i} X_j \le C_{p,M} \right\} \right) \right] (1 + o(1)) \end{aligned} \tag{1.5}$$

uniformly over $x > 0$.

Note that we have used a truncation level that remains at B_p for small M but grows in order $\sqrt{M}$ for large M. Such level will ensure that the contribution of two or more jumps i.e. $X_i > C_{p,M}$, is negligible for both small and large M. Moreover, as we shall see in Proposition (1.6) below, K_p is chosen such that the truncated part is regular enough to invoke uniform renewal theorem.

Our argument is finished by recognizing the two components in the right hand side of (1.5) as the terms in (1.1), via the following propositions:

PROPOSITION (1.6).

$$P\left(S_M > x, \max_{1\le i\le n} X_i \le C_{p,M} \right) = e^{-\theta^* x} + o\left(e^{-\theta^* x} + \frac{1}{p}\bar{F}(x) I(x \ge B_p) \right)$$

uniformly over $x > z(p)$ *for any* $z(p)$ *such that* $z(p) \to \infty$ *as* $p \to 0$.

PROPOSITION (1.7).

$$\begin{aligned} &P\left(S_M > x, \bigcup_{i=1}^{n} \left\{ X_i > C_{p,M}, \max_{j\ne i} X_j \le C_{p,M} \right\} \right) \\ &= \begin{cases} \left[\frac{1}{p}\bar{F}(x) + \int_{B_p}^{x} \left(\frac{1}{p} + \frac{x-y}{\mu} \right) e^{-\theta^*(x-y)} dF(y) \right] (1 + o(1)) & \text{uniformly over } x \ge B_p \\ \ll e^{-\theta^* x} & \text{uniformly over } x < B_p \end{cases} \end{aligned}$$

2. Proofs

Note that Theorem (1.2) is uniform over $x \ge 0$. The results that follow are obtained uniformly over $x \ge z(p)$ as long as $z(p) \to \infty$ as $p \to 0$. Of course, for $x \le z(p) < B_p$ we have that $\Gamma(x,p) = e^{-\theta^* x}$ and therefore the limit

$$\lim_{p\to 0} \sup_{0\le xp\le z(p)p} \left| \frac{P(pS_M > px)}{\Gamma(x,p)} - 1 \right| = \lim_{p\to 0} \sup_{0\le u\le z(p)p} \left| \frac{P(pS_M > u)}{\exp(-(\theta^*/p)u)} - 1 \right| = 0,$$

is easily established if $z(p)p = o(1)$. Consequently, to obtain Theorem (1.2) it suffices to indeed assume $x \ge z(p)$ as indicated.

Proof of Proposition (1.4). First we write

$$\begin{aligned} P(S_M > x) &= P(S_M > x,\ \max X_i \le B_p,\ M \le K_p) \\ &\quad + P(S_M > x,\ \max X_i \le \mu + \sqrt{M},\ M > K_p) \\ &\quad + P(S_M > x,\ \max X_i > B_p,\ M \le K_p) \\ &\quad + P(S_M > x,\ \max X_i > \mu + \sqrt{M},\ M > K_p) \end{aligned} \tag{2.1}$$

Note that the first two terms constitute the first term in the right hand side of (1.5), and we shall focus on the last two terms. For the third term, we have

$$\begin{aligned} &P(S_M > x,\ \max X_i > B_p,\ M \le K_p) \\ &\quad = P(S_M > x,\ \text{exactly one } X_i > B_p,\ M \le K_p) \\ &\qquad + P(S_M > x,\ \text{more than one } X_i > B_p,\ M \le K_p). \end{aligned} \tag{2.2}$$

We will show that the second term in (2.2) is negligible compared to the first term in (2.1), by following the proof in Lemma 4 of Rozovskii (1993). Using the notation there, we denote

$$Q_{n-k,k}(x) = P(S_n > x,\ X_1, X_2, \ldots, X_k > B_p,\ X_{k+1}, \ldots, X_n \le B_p),$$

$A = \sup_{y \ge 2B_p} I_{B_p}(y)/\bar{F}(y)$ where

$$I_{B_p}(y) = \int_{B_p}^{y-B_p} \bar{F}(y-u)dF(u),$$

and $\xi_{B_p} = \sup_{y \ge B_p} \bar{F}(y)/\bar{F}(y+B_p)$. Now Lemma 4a in Rozovskii (1993), applying to our case, states that for $k \ge 2$, $Q_{n-k,k}(x) \le AQ_{n-k+1,k-1}(x) + \bar{F}(B_p)Q_{n-k+1,k-1}(x - B_p)$ (from equation (64) there), and $Q_{n-k+1,k-1}(x - B_p) \le \xi_{B_p} Q_{n-k+1,k-1}(x)$ (from equation (65) there).

Recognizing that X_i is always non-negative, we have

$$\begin{aligned} Q_{n-k,k}(x) &\le (F(B_p))^{-1} Q_{n-k+1,k}(x) \\ &\le (F(B_p))^{-1}(AQ_{n-k+1,k-1}(x) + \bar{F}(B_p)Q_{n-k+1,k-1}(x - B_p)) \\ &\le (F(B_p))^{-1}(A + \bar{F}(B_p)\xi_{B_p})Q_{n-k+1,k-1}(x) \\ &= H_p Q_{n-k+1,k-1}(x) \end{aligned} \tag{2.3}$$

where $H_p = (F(B_p))^{-1}(A + \bar{F}(B_p)\xi_{B_p})$. Lemma 4c in Rozovskii (1993) yields that (note the slight difference in the definition of $g(x)$ between there and here. We denote $\bar{F}(x) = e^{-g(x)}$ while Rozovskii defines $\bar{F}(x) \sim e^{-g(x)}/x^2$ for the case of finite variance. Thus the $g(x)$'s differs by a term of $2\log x$. We also note that in Rozovskii (1993) $B_n = \sqrt{n}$ in the case of finite variance.)

$$A = O\left(\frac{1}{B_p^2} \max_{y \ge B_p} g(y) y^{2(1-\delta)} e^{-(1-\delta)g(y)}\right), \text{ and } \xi_{B_p} = O\left(\frac{\exp\{\delta g(B_p)\}}{B_p^{2\delta}}\right).$$

Now let $\tilde{g}(x) = g(x) - 2\log x$. Then we have

$$\begin{aligned} A &\le p^2 \max_{y \ge B_p} (\tilde{g}(y) + 2\log y) e^{-(1-\delta)\tilde{g}(y)} \\ &\le p^2 \max_{y \ge B_p} C\tilde{g}(y) e^{-(1-\delta)\tilde{g}(y)} \quad \text{for some constant} C, \\ &\quad \text{by our assumption that } g(x) \ge (2+\epsilon)\log x \\ &= Cp^2 \tilde{g}(B_p) e^{-(1-\delta)\tilde{g}(B_p)} = Cp^{2\delta}(g(B_p) - 2\log B_p) e^{-(1-\delta)g(B_p)} \end{aligned}$$

when p is small enough, and $\xi_{B_p} = O(p^{2\delta} e^{\delta g(B_p)})$. Hence we have

$$K_p \cdot H_p \le C(F(B_p))^{-1}(g(B_p)+1)e^{-\eta g(B_p)} \to 0$$

for some $C > 0$. So

$$\begin{aligned} & \sum_{n=2}^{\lfloor K_p \rfloor} pq^n \sum_{k=2}^{n} \binom{n}{k} Q_{n-k,k}(x) \\ \le & \sum_{n=2}^{\lfloor K_p \rfloor} pq^n \sum_{k=2}^{n} n^k H_p^{k-1} P(S_n > x,\ X_1 > B_p,\ \max_{2 \le i \le n} X_i \le B_p) \ \text{(by iterating (2.3))} \\ \le & \sum_{n=2}^{\lfloor K_p \rfloor} pq^n n \sum_{k=2}^{n} K_p^{k-1} H_p^{k-1} P(S_n > x,\ X_1 > B_p,\ \max_{2 \le i \le n} X_i \le B_p) \\ \le & \sum_{n=2}^{\lfloor K_p \rfloor} pq^n n \frac{K_p H_p}{1 - K_p H_p} P(S_n > x,\ X_1 > B_p,\ \max_{2 \le i \le n} X_i \le B_p) \ \text{(for } p \text{ small enough)} \\ \ll & \sum_{n=2}^{\lfloor K_p \rfloor} pq^n n P(S_n > x,\ X_1 > B_p,\ \max_{2 \le i \le n} X_i \le B_p) \end{aligned}$$

and hence

$$P(S_M > x, \text{ more than one } X_i > B_p,\ M \le K_p) \ll P(S_M > x,\ \max X_i \le B_p,\ M \le K_p)$$

uniformly over $x > 0$.

We are left to prove that the fourth term in (2.1) is equivalent in order to

$$P(S_M > x, \text{ exactly one } X_i > \mu + \sqrt{M}, M > K_p)$$

Using directly the result in Lemma 4 of Rozovskii (1993), and denoting $\tilde{X}_i = X_i - \mu$ and $\tilde{S}_n = \sum_{i=1}^n \tilde{X}_i$ as the centered random variables, we get

$$\begin{aligned} & P(S_M > x,\ \max X_i > \mu + \sqrt{M},\ M > K_p) = \\ = & \sum_{n=\lfloor K_p \rfloor + 1}^{\infty} pq^n P(S_n > x,\ \max X_i > \mu + \sqrt{n}) \\ = & \sum_{n=\lfloor K_p \rfloor + 1}^{\infty} pq^n P(\tilde{S}_n > x - n\mu,\ \max \tilde{X}_i > \sqrt{n}) \\ = & \sum_{n=\lfloor K_p \rfloor + 1}^{\infty} pq^n P(\tilde{S}_n > x - n\mu,\ \text{exactly 1 } \tilde{X}_i > \sqrt{n})(1 + g(x,n)) \end{aligned}$$

where the last equality follows from (61) in Rozovskii (1993) and $g(x,n) \to 0$ uniformly in x as $n \to \infty$.

Since $K_p \to \infty$ as $p \to 0$, for any ϵ, we have

$$\sum_{n=\lfloor K_p \rfloor+1}^{\infty} pq^n P(\tilde{S}_n\mu > x - n\mu, \text{ exactly 1 } \tilde{X}_i > \sqrt{n}) g(x,n)$$
$$\leq \epsilon \sum_{n=\lfloor K_p \rfloor+1}^{\infty} pq^n P(\tilde{S}_n\mu > x - n\mu, \text{ exactly 1 } \tilde{X}_i > \sqrt{n})$$

when p becomes small enough. Therefore

$$P(S_M > x, \max X_i > \mu + \sqrt{M}, M > K_p)$$
$$\leq (1+\epsilon) P(S_M > x, \text{exactly one } X_i > \mu + \sqrt{M}, M > K_p)$$

for p small enough. Since ϵ is arbitrary, we conclude that

$$P(S_M > x, \max X_i > \mu + \sqrt{M}, M > K_p)$$
$$\sim P(S_M > x, \text{exactly one } X_i > \mu + \sqrt{M}, M > K_p)$$

uniformly over $x > 0$.

The proof of Proposition (1.4) is complete. □

Proof of Proposition (1.6). Note that

$$\begin{aligned} & P(S_M > x, \max X_i \leq C_{p,M}) = \\ = \; & P(S_M > x, \max X_i \leq B_p, M \leq K_p) + P(S_M > x, \max X_i \leq \mu + \sqrt{M}, M > K_p) \\ = \; & P(S_M > x, \max X_i \leq B_p) - P(S_M > x, \max X_i \leq B_p, M > K_p) \\ & + P(S_M > x, \max X_i \leq \mu + \sqrt{M}, M > K_p) \end{aligned}$$

The proof of the proposition will be finished by the following two lemmas:

LEMMA (2.4).

$$P(S_M > x, \max X_i \leq B_p) = e^{-\theta^* x}(1 + o(1))$$

uniformly over $x > z(p)$ for any $z(p)$ such that $z(p) \to \infty$ as $p \to 0$.

LEMMA (2.5).

$$P(S_M > x, \max X_i \leq B_p, M > K_p)$$
$$\leq P(S_M > x, \max X_i \leq \mu + \sqrt{M}, M > K_p) \ll e^{-\theta^* x} + \frac{1}{p}\bar{F}(x)$$

uniformly over $x > z(p)$ for any $z(p)$ such that $z(p) \to \infty$ as $p \to 0$. □

Proof of Lemma (2.4). Let $\bar{X}$ be the truncated random variable at level B_p i.e.

$$P(\bar{X} \in B) = P(X \in B | X \leq B_p)$$

for any measurable set B. We denote $P_\theta(\cdot)$ as the exponential change of measure

$$P_\theta(\cdot) = E[e^{\theta\bar{X} - \bar{\psi}(\theta)}; \cdot\,]$$

where $\bar{\psi}(\cdot) = \log\bar{\phi}(\cdot)$ is the logarithmic moment generating function of $\bar{X}$ and $\bar{\phi}(\theta) = Ee^{\theta\bar{X}}$ is the moment generating function of $\bar{X}$. Also denote $E_\theta[\cdot]$ as the expectation under $P_\theta(\cdot)$ and $\bar{\mu}_\theta = E_\theta\bar{X}$.

Consider the change of measure of X_i with θ satisfying

$$\bar{\phi}(\theta) = \frac{1}{qF(B_p)}$$

which gives

$$E[e^{\theta X}; X \le B_p] = \frac{1}{q} \tag{2.6}$$

This equation is similar to (33) in Blanchet and Glynn (2007), but with a different truncation level. Theirs is a truncation level x while here we use B_p regardless of x.

We want to characterize the solution of (2.6), which, as we will see, will give the θ^* in Theorem (1.2). Suppose $0 \le \theta \le Cp$ for some $C > 0$ and p small enough. Write

$$\begin{aligned} E[e^{\theta X}; X \le B_p] &= \int_0^{B_p} e^{\theta y} dF(y) \\ &= \int_0^{B_p} \left(1 + \theta y + \frac{\theta^2 y^2}{2} e^{\nu\theta y}\right) dF(y) \end{aligned}$$

for some $0 \le \nu = \nu(\theta y) \le 1$. The equation is valid by our moment assumptions on X. We then get

$$E[e^{\theta X}; X \le B_p] = F(B_p) + \theta m(B_p) + R(\theta) \tag{2.7}$$

where

$$F(B_p) = 1 - \bar{F}(B_p) = 1 - e^{-g(B_p)}$$

$$m(B_p) = \int_0^{B_p} y dF(y) = \mu(1 + o(1)) \le \mu$$

as $p \to 0$ and

$$R(\theta) = \int_0^{B_p} \frac{\theta^2 y^2}{2} e^{\nu\theta y} dF(y) \le \frac{\theta^2}{2} e^C D(B_p) \le \frac{\theta^2}{2} e^C$$

where $D(B_p) = \int_0^{B_p} y^2 dF(y)$. Equating (2.7) with $1/q = 1 + p + p^2 + \cdots$, and noting that $e^{-g(B_p)} \ll p^2$, we get $\theta^* \sim p/\mu$, which also verifies that there is a unique θ^* that indeed lies in $[0, Cp]$. Henceforth we will identify this as our θ^*.

Next we have, letting T be exponentially distributed with rate θ^* in the fourth equality below,

$$\begin{aligned} P(S_M > x, \max X_i \le B_p) &= \sum_{n=1}^{\infty} pq^n P(S_n > x, \max X_i \le B_p) \\ &= \sum_{n=1}^{\infty} p(qF(B_p))^n P(\bar{S}_n > x) = \sum_{n=1}^{\infty} pE_{\theta^*}[e^{-\theta^* \bar{S}_n}; \bar{S}_n > x] \\ &= p \sum_{n=1}^{\infty} E_{\theta^*}[x < \bar{S}_n < T] = pE_{\theta^*}\left[\sum_{n=1}^{\infty} I(x < \bar{S}_n < T))\right] \\ &= p \int_x^{\infty} \theta^* e^{-\theta^* y} (\bar{U}_{\theta^*}(y) - \bar{U}_{\theta^*}(x)) dy \\ &= pe^{-\theta^* x} \int_0^{\infty} \theta^* e^{-\theta^* y} (\bar{U}_{\theta^*}(y + x) - \bar{U}_{\theta^*}(x)) dy, \end{aligned}$$

where $\bar{U}_{\theta^*}(\cdot) = \sum_{n=1}^{\infty} P_{\theta^*}(\bar{S}_n \le \cdot)$ is the renewal measure of $\bar{X}_i$ under the measure $P_{\theta^*}(\cdot)$.

We shall now find $\bar{U}_{\theta^*}(\cdot)$. More specifically, we will show that uniform renewal theorem, as depicted in Theorem 1, part 2 of Blanchet and Glynn (2007), is valid for $\bar{X}_i$ under the exponential family $P_{\theta^*}(\cdot)$ for all small enough p. Denote $\bar{\chi}_{\theta^*}(\omega) = E_{\theta^*} e^{i\omega\bar{X}} = qF(B_p)Ee^{(i\omega+\theta^*)\bar{X}}$ as the characteristic function of $\bar{X}$ under $P_{\theta^*}(\cdot)$. The theorem requires that such family is uniformly strongly non-lattice i.e.

$$\inf_{0\le p\le\kappa}\inf_{|\omega|>v}|1-\chi_{\theta^*}(\omega)|>0$$

for small enough $\kappa > 0$ and any $v > 0$, and that $\sup_{0\le p\le\kappa} E_{\theta^*}\bar{X}^{2+\epsilon} < \infty$. If these conditions hold then we have (as a weaker conclusion than the stated theorem in Blanchet and Glynn (2007))

$$\sup_{0\le p\le\kappa}\bar{\mu}_{\theta^*}^4\left|\bar{U}_{\theta^*}(t)-\frac{t}{\mu_{\theta^*}}-\frac{E_{\theta^*}\bar{X}^2}{2\bar{\mu}_{\theta^*}^2}\right| = o(1) \tag{2.8}$$

as $t\to\infty$.

We now check the above conditions. First note that

$$\begin{aligned}\bar{\chi}_{\theta^*}(\omega) &= qF(B_p)Ee^{i(\omega+\theta^*)\bar{X}} = qE[e^{(i\omega+\theta^*)X};X\le B_p]\\ &= q(E[e^{i\omega X};X\le B_p]+\theta^* r(\theta^*,\omega)) = q(\chi(\omega)-E[e^{i\omega X};X>B_p]+\theta^* r(\theta^*,\omega))\end{aligned}$$

where $|r(\theta^*,\omega)| \le E[Xe^{v\theta^* X};X\le B_p]$ for some $0\le v = v(\theta^* X)\le 1$ a.s. and the second equality is valid by our moment assumptions on X. Note that

$$|E[e^{i\omega X};X>B_p]|\le\bar{F}(B_p), \text{ and } E[Xe^{(v\theta^*)X};X\le B_p]\le e^C\mu,$$

for some $C > 0$. Since X is non-lattice, given any $v > 0$, we have $|1-\chi(\omega)| > 1-\zeta$ for some $0 < \zeta = \zeta(v) < 1$ for all $|\omega| > v$. So for $|\omega| > v$, we have

$$\begin{aligned}|1-\bar{\chi}_{\theta^*}(\omega)| &= |1-(1-p)(\chi(\omega)-E\left[e^{i\omega X};X>B_p\right]+\theta^* r(\theta^*,\omega)|\\ &\ge |1-\chi(\omega)|-|E[e^{i\omega X};X>B_p]|-\theta^*|r(\theta^*,\omega)|\\ &\quad -p(|\chi(\omega)|+|E[e^{i\omega X};X>B_p]|+\theta^*|r(\theta^*,\omega)|)>1-\zeta'\end{aligned}$$

for some $0 < \zeta' < 1$ and small enough p. This shows that $\bar{X}$ under the exponential family $P_{\theta^*}(\cdot)$ is uniformly strongly non-lattice. Moreover, we have

$$E_{\theta^*}\bar{X}^{2+\epsilon}\le qE[X^{2+\epsilon}e^{\theta^* X};X\le B_p]\le e^C EX^{2+\epsilon}$$

for some $C > 0$. Together with our moment assumption on X, this shows that $\sup_{0\le p\le\kappa} E_{\theta^*}\bar{X}^{2+\epsilon} < \infty$.

Hence we can invoke the uniform renewal theorem in Blanchet and Glynn (2007). Since $P_{\theta^*}(\bar{X} > s) = qE[e^{\theta^* X};s < X\le B_p]\le qe^C\bar{F}(s)$ for $s < B_p$ and is 0 otherwise, we note that $H_1^F(t)$, $H_2^F(t)$ and $H_1^F * H_1^F(t)$ in Theorem 1, part 2 of Blanchet and Glynn (2007) all go to 0 uniformly in our exponential family as $t\to\infty$. Note also that $\bar{\mu}_{\theta^*} = qE[Xe^{\theta^* X};X\le B_p]$. Since $Xe^{\theta^* X}I(X\le B_p)\le Xe^C$ which is integrable, by dominated convergence theorem and that $\theta^*\sim p/\mu$ we have $E[Xe^{\theta^* X};X\le B_p]\to\mu$. Hence $\bar{\mu}_{\theta^*} = \mu + o(1)$. This concludes, from (2.8), that

$$\bar{U}_{\theta^*}(y+x)-\bar{U}_{\theta^*}(x) = \frac{y}{\bar{\mu}_{\theta^*}}+R(y,x,\theta^*)$$

where $\sup_{0\le p\le\kappa,y>0}|R(y,x,\theta^*)|\to 0$ as $x\to\infty$. So

$$pe^{-\theta^* x}\int_0^\infty \theta^* e^{-\theta^* y}(\bar U_{\theta^*}(y+x)-\bar U_{\theta^*}(x))dy$$
$$= pe^{-\theta^* x}\int_0^\infty \theta^* e^{-\theta^* y}\left(\frac{y}{\bar\mu_{\theta^*}}+R(y,x,\theta^*)\right)dy$$
$$= \frac{pe^{-\theta^* x}}{\bar\mu_{\theta^*}}\left(\frac{1}{\theta^*}+o(1)\right)\sim e^{-\theta^* x}$$

uniformly over $x>z(p)$ for any $z(p)$ such that $z(p)\to\infty$ as $p\to 0$. Lemma (2.4) is proved. □

Proof of Lemma (2.5). The first inequality holds obviously when p is small enough. Thus we will focus on the order relation. As in the proof of Proposition (1.5), we let $\tilde X_i = X_i-\mu$ and $\tilde S_n=\sum_{i=1}^n \tilde X_i$ be the centered random variables and their sum. Let $P(\tilde X>x)=e^{-g(x+\mu)}$ and recall that $h(x)=g(x+\mu)-2\log x$ satisfies $h(x)/x^{\delta'}\to 0$ and is eventually decreasing, and

$$\frac{h'(x)}{h(x)}\le\frac{\delta'}{x} \tag{2.9}$$

for large enough x. Note also that $\epsilon\log x\preceq h(x)\ll x$ by construction.

By (70) and (71) in Rozovskii (1993) (note that the $g(x)$ defined there differs from ours by a term of $2\log x$ and hence $h(x)$ here will play the role of $g(x)$ in Rozovskii's paper) we have

$$P(\tilde S_n>x-n\mu,\max\tilde X_i\le\sqrt{n})\le e^{-\beta(x-n\mu)h(\sqrt{n})/\sqrt{n}} \tag{2.10}$$

for $x\ge n\mu+\Lambda_n$, where $\Lambda_n=\alpha h(\sqrt{n})\sqrt{n}$ and $\beta<\alpha/2$, for large enough n and some constant $\alpha>0$. This will be important for our development.

We now define the following functions that will prove useful for our argument. Let $l(n)=n\mu+\Lambda_n$. We extend the domain of the function l to the positive real axis and define $l^{-1}(y)=\inf\{x: l(x)\ge y\}$, so $l^{-1}(x)=x/\mu-(\alpha/\mu)h(\sqrt{l^{-1}(x)})\sqrt{l^{-1}(x)}$. Also let $r(x)=h(x)/x$, so $1/x\ll r(x)\ll 1/x^{1-\delta'}$ as $x\nearrow\infty$. Define $r^{-1}(y)=\inf\{x: r(x)\le y\}$. We then have

$$1/y^{1/(1-\delta')}\ll r^{-1}(y)\ll 1/y \tag{2.11}$$

as $y\searrow 0$.

We shall also prove a monotone property concerning the function (whose use will become clear in the argument that follows)

$$f(n):=-\beta(x-n\mu)\frac{h(\sqrt{n})}{\sqrt{n}}+n\log q$$

We have

$$f'(n)=-\beta(x-n\mu)\left(\frac{h'(\sqrt{n})}{2n}-\frac{1}{2}\frac{h(\sqrt{n})}{n^{3/2}}\right)+\beta\mu\frac{h(\sqrt{n})}{\sqrt{n}}+\log q$$
$$=-\frac{\beta}{2}\left(\frac{x}{n}-\mu\right)\left(h'(\sqrt{n})-\frac{h(\sqrt{n})}{\sqrt{n}}\right)+\beta\mu\frac{h(\sqrt{n})}{\sqrt{n}}+\log q$$

By (2.9), we have $f'(n)\ge\beta\mu h(\sqrt{n})/\sqrt{n}+\log q\ge 0$, or $f(n)$ is increasing, when $n\le(r^{-1}(-(\log q)/(\beta\mu)))^2$.

Let $R_p = r^{-1}(-(\log q)/(\beta\mu))$. We write

$$\begin{aligned}
&P(S_M > x,\ \max X_i \le \mu + \sqrt{M},\ M > K_p)\\
=\ &\sum_{n=\lfloor K_p\rfloor+1}^{\infty} pq^n P(\tilde{S}_n > x - n\mu, \max \tilde{X}_i \le \sqrt{n})\\
=\ &\sum_{n=\lfloor K_p\rfloor+1}^{\lfloor R_p^2\rfloor \wedge \lfloor l^{-1}(x)\rfloor} pq^n P(\tilde{S}_n > x - n\mu, \max \tilde{X}_i \le \sqrt{n}) I(\lfloor R_p^2\rfloor \wedge \lfloor l^{-1}(x)\rfloor > \lfloor K_p\rfloor)\\
&+ \sum_{n=\lfloor R_p^2\rfloor \vee \lfloor K_p\rfloor+1}^{\lfloor l^{-1}(x)\rfloor} pq^n P(\tilde{S}_n > x - n\mu, \max \tilde{X}_i \le \sqrt{n}) I(\lfloor l^{-1}(x)\rfloor > \lfloor R_p^2\rfloor \vee \lfloor K_p\rfloor)\\
&+ \sum_{n=\lfloor K_p\rfloor \vee \lfloor l^{-1}(x)\rfloor+1}^{\infty} pq^n P(\tilde{S}_n > x - n\mu, \max \tilde{X}_i \le \sqrt{n}) \qquad (2.12)
\end{aligned}$$

We will now analyze the terms one by one. Note that $n \le l^{-1}(x)$ implies $x \ge l(n)$. Hence by (2.10) and our monotone property of $f(\cdot)$ we have

$$\begin{aligned}
&\sum_{n=\lfloor K_p\rfloor+1}^{\lfloor R_p^2\rfloor \wedge \lfloor l^{-1}(x)\rfloor} pq^n P(\tilde{S}_n > x - n\mu, \max \tilde{X}_i \le \sqrt{n}) I(\lfloor R_p^2\rfloor \wedge \lfloor l^{-1}(x)\rfloor > \lfloor K_p\rfloor)\\
\le\ &\sum_{n=\lfloor K_p\rfloor+1}^{\lfloor R_p^2\rfloor \wedge \lfloor l^{-1}(x)\rfloor} pq^n e^{-\beta(x-n\mu)r(\sqrt{n})} I(\lfloor R_p^2\rfloor \wedge \lfloor l^{-1}(x)\rfloor > \lfloor K_p\rfloor)\\
\le\ &\begin{cases} \sum_{n=\lfloor K_p\rfloor+1}^{\lfloor l^{-1}(x)\rfloor} pq^{l^{-1}(x)} \exp\left\{-\beta(x - l^{-1}(x)\mu) r(\sqrt{l^{-1}(x)})\right\} & \text{for } \lfloor l^{-1}(x)\rfloor \le \lfloor R_p^2\rfloor \\ \sum_{n=\lfloor K_p\rfloor+1}^{\lfloor R_p^2\rfloor} pq^{R_p^2} \exp\left\{-\beta(x - R_p^2\mu)\left(-\frac{\log q}{\beta\mu}\right)\right\} & \text{for } \lfloor l^{-1}(x)\rfloor > \lfloor R_p^2\rfloor \end{cases}
\end{aligned}$$

$$\le \begin{cases} l^{-1}(x) pq^{l^{-1}(x)} e^{-\beta\alpha h^2(\sqrt{l^{-1}(x)})} & \text{for } \lfloor l^{-1}(x)\rfloor \le \lfloor R_p^2\rfloor \\ R_p^2 p e^{(x/\mu)\log q} & \text{for } \lfloor l^{-1}(x)\rfloor > \lfloor R_p^2\rfloor \end{cases} \qquad (2.13)$$

and for $\lfloor l^{-1}(x)\rfloor > \lfloor K_p\rfloor$. The first part of the last inequality follows by substituting $x = l(l^{-1}(x)) = l^{-1}(x) + \alpha h(\sqrt{l^{-1}(x)})\sqrt{l^{-1}(x)}$. We shall prove that in both cases they are of smaller order than $e^{-\theta^* x} + (1/p)\bar{F}(x) I(x \ge B_p)$.

Consider the first case, and suppose $K_p \le x \le R_p$. Dividing the first part of (2.13) by $e^{-\theta^* x}$ gives

$$\exp\left\{\frac{p\alpha}{\mu} h(\sqrt{l^{-1}(x)})\sqrt{l^{-1}(x)}(1+o(1)) - \beta\alpha h^2(\sqrt{l^{-1}(x)}) + \log(l^{-1}(x)p)\right\} \qquad (2.14)$$

by substituting $l^{-1}(x) = x/\mu - (\alpha/\mu)h(\sqrt{l^{-1}(x)})\sqrt{l^{-1}(x)}$ and using $\log q = -p(1+o(1))$. Note that $x \le R_p$ implies that $r(x) \ge -(\log q)/(\beta\mu) \ge p/(\beta\mu)$. Since $r(\sqrt{l^{-1}(x)}) \ge r(x)$, we have $h(\sqrt{l^{-1}(x)}) \gg (p/(\beta\mu))\sqrt{l^{-1}(x)}$. This gives

$$\beta\alpha h^2(\sqrt{l^{-1}(x)}) \gg (p\alpha/\mu) h(\sqrt{l^{-1}(x)})\sqrt{l^{-1}(x)}(1+o(1)).$$

Since $h(x) \gg \epsilon \log x$, we also have $\beta\alpha h^2(\sqrt{l^{-1}(x)}) \gg \log l^{-1}(x) \ge \log(l^{-1}(x)p)$. Thus (2.14) is equal to $\exp\{-\beta\alpha h^2(\sqrt{l^{-1}(x)})(1+o(1))\} \le \exp\{-\beta\alpha h^2(\sqrt{K_p})(1+o(1))\} = o(1)$.

Now suppose $x \gg R_p$ and $\lfloor l^{-1}(x) \rfloor \le \lfloor R_p^2 \rfloor$. Note that for any $x \gg R_p$, one can always find $a = a(p) \nearrow \infty$ arbitrarily slowly as $p \searrow 0$, such that $x \gg aR_p$. Dividing the first part of (2.13) by $(1/p)\bar{F}(x)$ gives

$$\exp\{l^{-1}(x)\log q - \beta a h^2(\sqrt{l^{-1}(x)}) + \log(l^{-1}(x)p) + h(x-\mu) + 2\log x + \log p\}. \tag{2.15}$$

Note that $x \gg aR_p$ implies $r(x) \ll r(x/a) \le p/(\beta\mu)$ and hence $pl^{-1}(x) \gg h(x-\mu)$, by using $l^{-1}(x) = x/\mu - (a/\mu)h(\sqrt{l^{-1}(x)})\sqrt{l^{-1}(x)}$. By substituting $y = R_p$ in (2.11) we have $paR_p \ge 1/p^{\delta'/(1-\delta')} \gg -\log p$ which implies $px/\mu \gg \log x$ for $x \gg aR_p$. Hence (2.15) is equal to $\exp\{-pl^{-1}(x)(1+o(1))\} \le \exp\{-pl^{-1}(aR_p)(1+o(1))\} = o(1)$.

We now proceed to the second part of (2.13). But $l^{-1}(x) \ge R_p^2$ implies $x \gg R_p$, since $h^2(\sqrt{l^{-1}(x)})/l^{-1}(x) \to 0$. Hence by the same argument $px/\mu \gg h(x)$ and $px/\mu \gg \log x \ge \log R_p$. Hence dividing the expression by $(1/p)\bar{F}(x)$ gives

$$\exp\left\{\frac{x}{\mu}\log q + 2\log R_p + \log p + h(x-\mu) + 2\log x + \log p\right\}$$
$$= \exp\left\{-\frac{px}{\mu}(1+o(1))\right\} \le \exp\left\{-\frac{pR_p^2}{\mu}(1+o(1))\right\} = o(1)$$

We now analyze the second term of (2.12). We have, for $\lfloor l^{-1}(x) \rfloor > \lfloor R_p^2 \rfloor \vee \lfloor K_p \rfloor$,

$$\begin{aligned}
&\sum_{n=\lfloor R_p^2\rfloor \vee \lfloor K_p\rfloor + 1}^{\lfloor l^{-1}(x)\rfloor} pq^n P(\tilde{S}_n > x - n\mu, \max \tilde{X}_i \le \sqrt{n}) \\
\le\ &\sum_{n=\lfloor R_p^2\rfloor + 1}^{\lfloor l^{-1}(x)\rfloor} pq^n e^{-\beta(x-n\mu)r(\sqrt{n})} \le \sum_{n=\lfloor R_p^2\rfloor + 1}^{\lfloor l^{-1}(x)\rfloor} p(qe^{\beta\mu r(\sqrt{n})})^n e^{-\beta x r(\sqrt{n})} \\
\le\ &\sum_{n=\lfloor R_p^2\rfloor + 1}^{\lfloor l^{-1}(x)\rfloor} p(qe^{\beta\mu \frac{-\log q}{\beta\mu}})^n e^{-\beta x r(\sqrt{l^{-1}(x)})} \le \sum_{n=\lfloor R_p^2\rfloor + 1}^{\lfloor l^{-1}(x)\rfloor} pe^{-\beta x r(\sqrt{l^{-1}(x)})} \\
&\le l^{-1}(x) p e^{-\beta x r(\sqrt{l^{-1}(x)})} \qquad (2.16)
\end{aligned}$$

where the third inequality holds because $R_p = r^{-1}(-\log q/(\beta\mu))$ and $r(x)$ is eventually decreasing. Note that $xr(\sqrt{l^{-1}(x)}) \sim \mu\sqrt{l^{-1}(x)}h(\sqrt{l^{-1}(x)}) \gg h(\sqrt{l^{-1}(x)})$. Also, since $\epsilon \log x \ll h(x) \ll x$, we have $xr\sqrt{l^{-1}(x)} \ge \sqrt{l^{-1}(x)}h(\sqrt{l^{-1}(x)}) \gg \log x$. Hence dividing (2.16) by $(1/p)\bar{F}(x)$ gives

$$\begin{aligned}
&\exp\{-\beta x r(\sqrt{l^{-1}(x)}) + \log(l^{-1}(x)p) + h(x-\mu) + 2\log x + \log p\} \\
=\ &\exp\{-\beta x r(\sqrt{l^{-1}(x)})(1+o(1))\} \le \exp\{-\beta R_p^2 r(\sqrt{l^{-1}(R_p^2)})(1+o(1))\} = o(1)
\end{aligned}$$

We now analyze the final term of (2.12). We have

$$\begin{aligned}
&\sum_{n=\lfloor K_p\rfloor \vee \lfloor l^{-1}(x)\rfloor + 1}^{\infty} pq^n P(\tilde{S}_n > x - n\mu, \max \tilde{X}_i \le \sqrt{n}) \le q^{l^{-1}(x)\vee K_p + 1} \\
\le\ &\begin{cases} q^{K_p/a'+1} & \text{for } l^{-1}(x) \ll K_p/a' \\ q^{l^{-1}(x)+1} & \text{for } l^{-1}(x) \ge K_p/a' \end{cases} \le \begin{cases} e^{-pK_p/a'} & \text{for } l^{-1}(x) \ll K_p/a' \\ e^{-pl^{-1}(x)} & \text{for } l^{-1}(x) \ge K_p/a' \end{cases} \qquad (2.17)
\end{aligned}$$

where $a' = a'(p) \nearrow 0$ as $p \searrow 0$ at a rate that will be chosen later on.

For the first part, dividing by $e^{-\theta^* x}$ yields

$$\exp\left\{-p\frac{K_p}{a'}+\theta^* x\right\}=\exp\left\{-p\frac{K_p}{a'}(1+o(1))\right\}=o(1)$$

For the second part observe that

$$r\left(\frac{K_p}{a'}\right)=\frac{h((1/a'^{2\delta})e^{(1-\delta')g(B_p)})}{(1/a'^{2\delta})e^{(1-\delta')g(B_p)}}\leq\frac{Ca'^{1-\delta'}p^{2\delta(1-\delta')}}{e^{(1-\delta'^2 g(B_p)}}\ll p$$

for some constant $C>0$, for a suitably chosen a', since $h(x)\leq x^\delta$ eventually. We then get $K_p/a'^{-1}(p)$ which implies $r(x)\leq r(K_p/a')\ll p$ for $l^{-1}(x)\geq K_p/a'$. This gives $pl^{-1}(x)\gg h(x)$. Note that $pK_p/a'=(1/a'^{1-2\delta}e^{(1-\delta')g(B_p)}\gg -2\delta\log p+(1-\delta')g(B_p)=\log K_p$, so $pl^{-1}(x)\gg\log x$ for $x\geq K_p/a'$. Hence dividing the second part of (2.17) by $(1/p)\bar{F}(x)$ gives

$$\begin{aligned}\exp\{-pl^{-1}(x)+h(x-\mu)+2\log x+\log p\}&=\exp\{-pl^{-1}(x)(1+o(1))\}\\&\leq\exp\left\{-pl^{-1}\left(\frac{K_p}{a'}\right)(1+o(1))\right\}=o(1)\end{aligned}$$

This concludes our proof of Lemma (2.5). □

Proof of Proposition (1.7). The case for $x<B_p$ is obvious, since we have $P(S_n>x,\text{ exactly one }X_i>C_{p,n})\leq n\bar{F}(B_p)$ and hence

$$P(S_M>x,\text{ exactly one }X_i>C_{p,M})\leq\sum_{n=1}^{\infty}pq^n n\bar{F}(B_p)=\frac{q}{p}\bar{F}(B_p)=o(1)\ll e^{-\theta^* x},$$

uniformly over $x<B_p$. Thus we shall focus on $x\geq B_p$. Note that

$$\begin{aligned}&P(S_M>x,\text{ exactly one }X_i>C_{p,M})\\
\leq\;&\sum_{n=1}^{\lfloor K_p\rfloor}pq^n n\left(\bar{F}(x)(F(B_p))^{n-1}+\int_{B_p}^{x}P(S_{n-1}>x-y,\max X_i\leq B_p)dF(y)\right)\\
&+\sum_{n=\lfloor K_p\rfloor+1}^{\infty}pq^n n\left(\bar{F}(x)(F(\mu+\sqrt{n}))^{n-1}\right.\\
&\left.+\int_{\mu+\sqrt{n}}^{x}P(S_{n-1}>x-y,\max X_i\leq\mu+\sqrt{n})dF(y)I(x\geq\mu+\sqrt{n})\right)\\
\leq\;&\frac{q}{p}\bar{F}(x)+\sum_{n=2}^{\lfloor K_p\rfloor}pq^n n\int_{B_p}^{x}P(S_{n-1}>x-y,\max X_i\leq B_p)dF(y)\\
&+\sum_{n=\lfloor K_p\rfloor+1}^{\infty}pq^n n\int_{\mu+\sqrt{n}}^{x}P(S_{n-1}>x-y,\max X_i\leq\mu+\sqrt{n})dF(y)I(x\geq\mu+\sqrt{n})\\
=\;&\frac{q}{p}\bar{F}(x)+\sum_{n=2}^{\infty}pq^n n\int_{B_p}^{x}P(S_{n-1}>x-y,\max X_i\leq B_p)dF(y)\\
&-\sum_{n=\lfloor K_p\rfloor+1}^{\infty}pq^n n\int_{B_p}^{x}P(S_{n-1}>x-y,\max X_i\leq B_p)dF(y)\\
&+\sum_{n=\lfloor K_p\rfloor+1}^{\infty}pq^n n\int_{\mu+\sqrt{n}}^{x}P(S_{n-1}>x-y,\max X_i\leq\mu+\sqrt{n})dF(y)I(x\geq\mu+\sqrt{n}).\end{aligned}$$

We will finish the proof by invoking the following lemmas:

LEMMA (2.18).

$$\sum_{n=2}^{\infty} pq^n n \int_{B_p}^{x} P(S_{n-1} > x-y, \max X_i \le B_p) dF(y)$$
$$= \int_{B_p}^{x} \left(\frac{1}{p} + \frac{x-y}{\mu} \right) e^{-\theta^*(x-y)} dF(y)(1+o(1))$$

uniformly over $x \ge B_p$.

LEMMA (2.19). *We have* $\mu + \sqrt{K_p} \ge B_p$ *for* p *small enough, and*

$$\sum_{n=\lfloor K_p \rfloor + 1}^{\infty} pq^n n \int_{B_p}^{x} P(S_{n-1} > x-y, \max X_i \le B_p) dF(y)$$
$$\le \sum_{n=\lfloor K_p \rfloor + 1}^{\infty} pq^n n \int_{B_p}^{x} P(S_{n-1} > x-y, \max X_i \le \mu + \sqrt{n}) dF(y)$$
$$\ll \int_{B_p}^{x} \left(\frac{1}{p} + \frac{x-y}{\mu} \right) e^{-\theta^*(x-y)} dF(y) + \frac{1}{p} \bar{F}(x)$$

uniformly over $x \ge B_p$. □

Proof of Lemma (2.18). We write

$$\sum_{n=2}^{\infty} pq^n n P(S_{n-1} > x, \max X_i \le B_p) = \frac{q}{p} \sum_{n=1}^{\infty} p^2 q^n (n+1) P(S_n > x, \max X_i \le B_p)$$
$$= \frac{q}{p} P(S_N > x, \max X_i \le B_p)$$

where N is a negative binomial variable with parameter 2 and p. Let $\{X_i'\}_{i=1,2,\ldots}$, M' and S_M' be independent and identical copies of $\{X_i\}_{i=1,2,\ldots}$, M and S_M, and let $F_M(x) = P(S_M \le x, \max X_i \le B_p)$ and $\bar{F}_M(x)$ be its complement defined by $P(S_M > x, \max X_i \le B_p)$. Note that by Lemma (2.4) we have $P(S_M > x, \max X_i \le B_p) = e^{-\theta^* x}(1+u(x,p))$ where $\sup_{x > B_p} |u(x,p)| \to 0$. We have

$$\begin{aligned}
& P(S_N > x, \max X_i \le B_p) \\
= \; & P(S_M + S_M' > x, \max_{1 \le i \le M} X_i \le B_p, \max_{1 \le j \le M'} X_j' \le B_p) \\
= \; & \int_0^x \bar{F}_M(x-y) dF_M(y) + \bar{F}_M(x)\bar{F}_M(0) \\
& = \int_0^x e^{-\theta^*(x-y)}(1+u(x-y,p)) dF_M(y) + \bar{F}_M(x)\bar{F}_M(0) \\
\sim \; & \int_0^x e^{-\theta^*(x-y)} dF_M(y) + \bar{F}_M(x)\bar{F}_M(0) \\
& = e^{-\theta^* x}\bar{F}_M(0) - \bar{F}_M(x) + \int_0^x \bar{F}_M(y)\theta^* e^{-\theta^*(x-y)} dy + \bar{F}_M(x)\bar{F}_M(0) \\
= \; & e^{-\theta^* x}\bar{F}_M(0) + \int_0^x \theta^* e^{-\theta^* x}(1+u(y,p)) dy - \bar{F}_M(x)(1-\bar{F}_M(0)) \sim e^{-\theta^* x} + \theta^* x e^{-\theta^* x}
\end{aligned}$$

uniformly over $x > B_p$. The fourth equality is obtained using integration by parts, and the last equality follows from the observation that $\bar{F}_M(0) = P(\max_{1 \le i \le M} X_i' \le B_p) = \sum_{n=0}^{\infty} pq^n F(B_p)^n \to 1$ as $p \to 0$. Noting that $\theta^* \sim p/\mu$, the conclusion of the lemma is then an easy consequence. □

Proof of Lemma (2.19). The inequality is obvious. We will thus focus on the order relation. We first prove that

$$\sum_{n=\lfloor K_p\rfloor+1}^{\infty} pq^n nP(S_{n-1}>x, \max X_i \le \mu+\sqrt{n}) \ll e^{-\theta^* x} + \frac{1}{p}\bar{F}(x)$$

uniformly over $x \ge B_p$. The proof is very similar to that of Lemma (2.5). Adopting the notation there, we can write

$$\begin{aligned}
&\sum_{n=\lfloor K_p\rfloor+1}^{\infty} pq^n nP(S_{n-1}>x, \max X_i \le \mu+\sqrt{n}) \\
=& \sum_{n=\lfloor K_p\rfloor+1}^{\lfloor R_p^2\rfloor\wedge\lfloor l^{-1}(x)\rfloor} pq^n nP(\tilde{S}_{n-1}>x-(n-1)\mu, \max \tilde{X}_i \le \sqrt{n}) I(\lfloor R_p^2\rfloor\wedge\lfloor l^{-1}(x)\rfloor > \lfloor K_p\rfloor) \\
&+ \sum_{n=\lfloor R_p^2\rfloor+1}^{\lfloor l^{-1}(x)\rfloor} pq^n nP(\tilde{S}_{n-1}>x-(n-1)\mu, \max \tilde{X}_i \le \sqrt{n}) I(\lfloor l^{-1}(x)\rfloor > \lfloor R_p^2\rfloor > \lfloor K_p\rfloor) \\
&+ \sum_{n=\lfloor K_p\rfloor\vee\lfloor l^{-1}(x)\rfloor+1}^{\infty} pq^n nP(\tilde{S}_{n-1}>x-(n-1)\mu, \max \tilde{X}_i \le \sqrt{n})
\end{aligned}$$

Using the same analysis, the first term will be less than or equal to

$$\begin{cases} \frac{l^{-1}(x)(l^{-1}(x)+1)}{2} pq^{l^{-1}(x)} e^{-\beta\alpha h^2(\sqrt{l^{-1}(x)})} & \text{for } \lfloor l^{-1}(x)\rfloor \le \lfloor R_p^2\rfloor \\ \frac{R_p^2(R_p^2+1)}{2} pe^{(x/\mu)\log q} & \text{for } \lfloor l^{-1}(x)\rfloor > \lfloor R_p^2\rfloor \end{cases}$$

the second term will be less than or equal to

$$\frac{l^{-1}(x)(l^{-1}(x)+1)}{2} pe^{-\beta x r(\sqrt{l^{-1}(x)})}$$

and the third term will be less than or equal to

$$\begin{cases} \left(K_p - \frac{1}{p}+1\right) e^{-pK_p/a'} & \text{for } l^{-1}(x) \ll K_p/a' \\ \left(l^{-1}(x) - \frac{1}{p}+1\right) e^{-pl^{-1}(x)} & \text{for } l^{-1}(x) \ge K_p/a' \end{cases}$$

For the first two terms the same analysis carries over while for the last one we only have to observe again that $px \gg \log x$ for $x \ge K_p/a'$, which will show our claim.

Hence we have

$$\begin{aligned}
&\sum_{n=\lfloor K_p\rfloor+1}^{\infty} pq^n n \int_{B_p}^{x} P(S_{n-1}>x-y, \max X_i \le \mu+\sqrt{n}) dF(y) \\
&\ll \int_{B_p}^{x} \left(e^{-\theta^*(x-y)} + \frac{1}{p}\bar{F}(x-y)\right) dF(y) \le \int_{B_p}^{x} \left(\frac{1}{p} + \frac{x-y}{\mu}\right) e^{-\theta^*(x-y)} dF(y) + \frac{1}{p}\bar{F}(x)
\end{aligned}$$

where the last order relation is obtained by using property of class $\mathcal{S}$ that $\int_0^x \bar{F}(x-y)/\bar{F}(x) dF(y) \to 2$ as $x \to \infty$. We conclude our proof of Lemma (2.19). □

Acknowledgements

We thank the Referee for the careful review and the useful recommendations. Blanchet acknowledges support from the NSF through the grants CMMI-0846816 and 1069064.

Received June 10, 2013

Final version received October 17, 2013

JOSE BLANCHET
ASSOCIATE PROFESSOR OF INDUSTRIAL ENGINEERING
AND OPERATIONS RESEARCH,
ASSOCIATE PROFESSOR OF STATITICS.
COLUMBIA UNIVERSITY
jose.blanchet@columbia.edu

HENRY K. LAM
ASSISTANT PROFESSOR OF MATHEMATICS AND STATISTICS
BOSTON UNIVERSITY
khlam@bu.edu

REFERENCES

[1] S. ASMUSSEN, Ruin Probabilities, World Scientific, Singapore, 2001.
[2] S. ASMUSSEN, Applied Probability and Queues, Springer, New York, 2003.
[3] J. BLANCHET AND P. GLYNN, *Complete Corrected Diffusion Approximations for the Maximum of a Random Walk*, Annals of Applied Probability **16** (2), (2006), 951–983.
[4] J. BLANCHET AND P. GLYNN, *Uniform renewal theory with applications to expansions of geometric sums*, Adv. Appl. Prob., **39**(4), (2007), 1070–1097.
[5] J. BLANCHET AND P. GLYNN, *Efficient rare-event simulation for the maximum of heavy-tailed random walks*, Annals of applied Probability **18** (4), (2008), 1351–1378.
[6] J. BLANCHET AND H. LAM, *State-dependent importance sampling for rare-event simulation: An overview and recent advances*, Surveys in Operations Research and Management Science**17** (1), (2012), 38–59.
[7] A. A. BOROVKOV AND K. A. BOROVKOV, *On probabilities of large deviations for random walks. I. Regularly varying distribution tails*, *Theory Prob. Appl.* **49** (2), (2001), 189–205.
[8] P. EMBRECHTS, C. KLUPPELBERG, AND T. MIKOSCH, Modelling Extremal Events for Insurance and Finance, Springer, 2003.
[9] P. EMBRECHTS AND N. VERAVERBEKE, *Estimates for the probability of ruin with special emphasis on the possibility of large claims*, Insurance: Mathematics and Economics **1** (1), (1982), 55–72.
[10] J. KINGMAN, *The single server queue in heavy traffic*, Proc. Camb. Phil. Soc. **57**, (1961), 902–904.
[11] H. LAM, J. BLANCHET, M. BAZANT, AND D. BURCH, *Corrections to the Central Limit Theorem for heavy-tailed probability densities*, Journal of Theoretical Probability **24** (4), (2011), 895–927.
[12] A. V. NAGAEV, *Integral limit theorems taking large deviations into account when Cramer's condition does not hold II*, Theory Prob. Appl. **14** (2), (1969), 193–208.
[13] M. OLVERA-CRAVIOTO, J. BLANCHET, AND P. GLYNN, *On the transition from heavy traffic to heavy tails for the $M/G/1$ queue: The regularly varying case*, Annals of Applied Probability **21** (2), (2011), 645–668.
[14] M. OLVERA-CRAVIOTO AND P. GLYNN, *Uniform Approximations for the $M/G/1$ Queue with Subexponential Processing Times*, Queueing Systems **68** (1), (2011), 1–50.
[15] L. V. ROZOVSKII, *Probabilities of large deviations of sums of independent random variables with common distribution function in the domain of attraction of the normal law*, Theory Prob. Appl. **34** (4), (1989), 625–644.
[16] L. V. ROZOVSKII, *Probabilities of large deviations on the whole axis*, Theory Prob. Appl. **38** (1), (1993), 53–79.

Bol. Soc. Mat. Mexicana (3) Vol. 19, 2013

SELF-SIMILAR MARKOV PROCESSES

JUAN CARLOS PARDO AND VÍCTOR RIVERO

ABSTRACT. This note surveys some recent results on self-similar Markov processes. Since the research around the topic has been very rich during the last fifteen years we do not pretend to cover all the recent developments in the field, and hence we focus mainly in giving a panorama of the areas where the authors have made contributions.

1. Introduction

As the title shows the main object of study in this paper is the class of real valued self-similar Markov processes, and in fact much of the results that will be summarised here concern the class of positive self-similar Markov processes. Before going into the detail we lift the following definition from Lamperti's pioneering work [43].

Definition (1.1). A stochastic process $X = \{X_t, t \geq 0\}$ defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, (\mathbb{P}_x)_{x\in\mathbb{R}^d})$ and $\mathbb{R}^d$-valued is said *semi-stable*, now a days called *self-similar*, if there exists an $\alpha \in \mathbb{R}$, such that for any $c > 0$,

$$\mathrm{Law}(\{cX_{c^{-\alpha}t}, t \geq 0\}, \mathbb{P}_x) = \mathrm{Law}(\{X_t, t \geq 0\}, \mathbb{P}_{cx})$$

that is, both processes have the same finite dimensional laws, viz. for any $0 < t_1 < t_2 < \cdots < t_n < \infty$

$$\begin{aligned}(1.2)\quad &\mathbb{P}_x\left(cX_{c^{-\alpha}t_1} \in \mathrm{d}x_1, cX_{c^{-\alpha}t_2} \in \mathrm{d}x_2, \ldots, cX_{c^{-\alpha}t_n} \in \mathrm{d}x_n\right)\\ &\qquad = \mathbb{P}_{cx}\left(X_{t_1} \in \mathrm{d}x_1, X_{t_2} \in \mathrm{d}x_2, \ldots, X_{t_n} \in \mathrm{d}x_n\right),\end{aligned}$$

where by $\mathbb{P}_x$ we understand the law of the process stating at x. Whenever $\alpha \neq 0$ we will say that the process is $1/\alpha$-self-similar.

Lamperti's main contribution in [43] has been to fully answer the question: *which are all the stochastic processes that can be obtained as the weak limit of some process on which we have applied an infinite sequence of contractions of the scale of time and space?* This question has been motivated by some rather transcendental results about weak convergence of normalised processes as for instance the famous result by Donsker [23] about convergence of a random walk towards a Brownian motion. Lamperti's [43] main results are summarised in the following theorem.

2010 *Mathematics Subject Classification*: 60G18, 60J99, 60G51.

Keywords and phrases: self-similar Markov process, Lévy processes, Markov additive processes, Lamperti representation, exponential functionals, entrance laws, quasi-stationary distributions, fluctuation theory, law of the iterated logarithm.

THEOREM (1.3). *Let $Y = \{Y_t, t > 0\}$ be a stochastic process $\mathbb{R}^d$-valued and $f : \mathbb{R} \to \mathbb{R}$ a function such that the process Y^ζ defined by*

$$Y_t^\zeta = \frac{Y_{\zeta t}}{f(\zeta)}, \qquad t > 0,$$

converges in the sense of finite dimensional laws, towards a non degenerated process, X, that is for any $0 < t_1 < t_2 < \cdots < t_n < \infty$, the convergence in law between random vector holds

$$\left(Y_{t_1}^\zeta, Y_{t_2}^\zeta, \ldots, Y_{t_n}^\zeta\right) \xrightarrow[\zeta\to\infty]{W} \left(X_{t_1}, X_{t_2}, \ldots, X_{t_n}\right). \tag{1.4}$$

Then, the process X is self-similar with an index α, for some $\alpha \in \mathbb{R}$. The function f is regularly varying with index α, that is $f(\zeta) = \zeta^\alpha L(\zeta)$, with L a function such that

$$\lim_{\zeta\to\infty} \frac{L(c\zeta)}{L(\zeta)} = 1, \qquad \text{for all } c > 0.$$

In this case it is nowadays said that X is the scaling limit of Y. Furthermore, any self-similar Markov process can be obtained this way.

Among the class of self-similar processes there are several important subfamilies that permit a better understanding of these. Some of them are the self-similar Gaussian processes; the class of additive self-similar processes, that is those with independent increments, those with homogeneous increments, and, those which are of particular interest to us, which have the strong Markov property. For properties and references about the former classes of processes see for instance the thorough survey by Embrechts and Maejima [26]. In the sequel we will restrict ourselves to the class of real valued self-similar Markov processes.

A real valued self-similar Markov process $X^{(x)}$, starting from x is a càdlàg strong Markov process which fulfills the above described *scaling property*. Real valued self-similar processes often arise in various parts of probability theory as limit of re-scaled processes. These processes are involved for instance in branching processes, Lévy processes, coalescent processes and fragmentation theory. Some particularly well known examples are Brownian motion, Bessel processes, stable subordinators, stable processes, stable Lévy processes conditioned to stay positive, etc.

Our main purpose in this paper is to give a panorama of properties of rssMp that have been obtained since the early sixties under the impulse of Lamperti's work [44], where the study of the case of positive self-similar Markov processes is initiated, and we will put particular emphasis in topics where the authors of this note have contributed.

2. Positive self-similar Markov processes

Throughout this paper we will assume that the self-similarity index, say $1/\alpha$, is strictly positive so $\alpha > 0$. When we restrict self-similar Markov processes to take values on the positive half-line we have an interesting relationship between this class and $\mathbb{R} \cup \{-\infty\}$-valued Lévy processes, such relation was obtained by Lamperti [44] and it is now known as the Lamperti representation of positive valued self-similar Markov processes, **pssMp** for short.

To state a precise result we recall that an $\mathbb{R}\cup\{-\infty\}$-valued stochastic process $\xi=(\xi_t,t\geq 0)$ is a Lévy process if its paths are càdlàg, the state $\{-\infty\}$ is an absorbing point, and it has stationary and independent increments. The state $\{-\infty\}$ is understood as an isolated point and hence the process hits this state and dies at an independent exponential time ζ, with some parameter $q\geq 0$, the case $q=0$ is included to allow this time to be infinite a.s. The law of ξ is characterized completely by its Lévy-Khintchine exponent Ψ, which takes the following form

$$\log\mathbf{E}\left[e^{z\xi_1},1<\zeta\right]=\Psi(z)=-q+bz+\frac{\sigma^2}{2}z^2+\int_{-\infty}^{\infty}\left(e^{zy}-1-zy\mathbb{I}_{\{|y|<1\}}\right)\Pi(dy), \tag{2.1}$$

for any $z\in i\mathbb{R}$, where $\sigma,b\in\mathbb{R}$ and Π is a Lévy measure satisfying the condition $\int_{\mathbb{R}}(y^2\wedge 1)\Pi(dy)<\infty$. For background about Lévy processes see [4], [40], [57].

Hereafter, for $x>0$ the measure $\mathbb{P}_x$ denotes the law of a pssMp issued from x, and to refer to a pssMp issued from $x>0$ we will use indistinctly $(X,\mathbb{P}_x)$, $(X^{(x)},\mathbb{P}_x)$ and $X^{(x)}$.

Lamperti's representation of self-similar $\mathbb{R}_+$–valued Markov processes killed at their first hitting time of 0, enables us to construct the paths of any such process starting from a strictly positive point from those of a Lévy process, and viceversa. More precisely, Lamperti [44] found the representation

$$X_t^{(x)}=\begin{cases}x\exp\xi_{\tau(tx^{-\alpha})}, & 0\leq t\leq x^{\alpha}I(\alpha\xi),\\ 0, & t\geq x^{\alpha}I(\alpha\xi),\end{cases} \tag{2.2}$$

under $\mathbb{P}_x$, for $x>0$, where

$$\tau(t)=\inf\{s>0:I_s(\alpha\xi)\geq t\},\quad I_s(\alpha\xi)=\int_0^s\exp\alpha\xi_u\,du\,,\quad I(\alpha\xi)=\lim_{t\to\zeta}I_t(\alpha\xi),$$

and where ξ is a $\mathbb{R}\cup\{-\infty\}$–Lévy process with law $\mathbf{P}$. Note that for $t<I(\alpha\xi)$, we have the equality $\tau_t=\int_0^t\left(X_s\right)^{-\alpha}ds$, so that (2.2) is invertible. Indeed, any $\mathbb{R}\cup\{-\infty\}$-valued Lévy process ξ can be represented as

$$\xi_t=\begin{cases}\log\left(\frac{X_{\gamma_t}}{X_0}\right), & 0\leq t<\int_0^{T_0-}X_s^{-\alpha}ds,\\ -\infty, & \int_0^{T_0-}X_s^{-\alpha}ds\leq t,\end{cases}$$

where X is some $1/\alpha$-pssMp, $\{\gamma_t,t\geq 0\}$ is the inverse of the additive functional

$$\int_0^t X_s^{-\alpha}\mathrm{d}s,\quad 0\leq t<T_0=\inf\{u>0:X_u=0\}.$$

Observe that the process ξ does not depend on the starting point of X. Hence, we will denote the law of ξ by $\mathbf{P}$, and it is obtained as the image measure of $\mathbb{P}_x$ under the latter transformation, independently of the starting point $x>0$. Reciprocally given a Lévy process $(\xi,\mathbf{P})$ using the former transformation we construct a family of Markovian measures $(\mathbb{P}_x)_{x>0}$, sharing the same semigroup. Hence Lamperti's transformation yields a one to one relation between the class of pssMp killed at their first hitting time of 0 and the one of Lévy processes. Unless otherwise stated, in the sequel we will denote a $1/\alpha$-pssMp by $(X,(\mathbb{P}_x)_{x>0})$, and by $(\xi,\mathbf{P})$ the Lévy process associated to it.

A first implication of Lamperti's transformation is that the first hitting time of 0, for a $1/\alpha$–pssMp and the exponential functional of a Lévy process, $I(\alpha\xi)$, are

equal in law, more precisely

$$(T_0, \mathbb{P}_x) \overset{\text{Law}}{=} (x^\alpha I(\alpha\xi), \mathbf{P}).$$

Another useful consequence is the following classification of pssMp's:

(LC1) $\zeta < \infty$ $\mathbf{P}$–a.s. if and only if

$$\mathbb{P}_x(T_0 < \infty,\ X_{T_0-} > 0,\ X_{T_0+s} = 0,\ \forall s \geq 0) = 1, \tag{2.3}$$

for all $x > 0$.

(LC2) $\zeta = \infty$ and $\lim_{t\to\infty} \alpha\xi_t = -\infty$ $\mathbf{P}$–a.s. if and only if

$$\mathbb{P}_x\left(T_0 < \infty,\ X_{T_0-} = 0,\ X_{T_0+s} = 0,\ \forall s \geq 0\right) = 1, \tag{2.4}$$

for all $x > 0$.

(LC3) $\zeta = \infty$ and $\limsup_{t\to\infty} \alpha\xi_t = \infty$ $\mathbf{P}$–a.s. if and only if

$$\mathbb{P}_x(T_0 = \infty) = 1, \qquad \text{for all } x > 0. \tag{2.5}$$

Furthermore, a useful way to characterise the pssMp associated to a Lévy process ξ is via its infinitesimal generator. Indeed, Volkonskii's Theorem allow us to ensure that the infinitesimal generator of X, say $\mathcal{L}$, evaluated in a function $f : \mathbb{R}^+ \to \mathbb{R}$, such that $\tilde{f}(\cdot) = f(e^\cdot)$ is in the domain of the infinitesimal generator of ξ, that we denote $\mathcal{A}$, takes the form

$$\begin{aligned}\mathcal{L}f(x) &= x^{-\alpha}\mathcal{A}\tilde{f}(\log x)\\ &= -qx^{-\alpha}f(x) + x^{1-\alpha}(-b + \frac{1}{2}\sigma^2)f'(x) + x^{2-\alpha}\frac{1}{2}\sigma^2 f''(x)\\ &\quad + x^{-\alpha}\int_{\mathbb{R}}\left(f(xe^y) - f(x) - yxf'(x)1_{\{|y|<1\}}\right)\Pi(dy),\end{aligned}$$

where (b, σ, Π) is the characteristic triple of ξ, and q is the rate at which it is killed. Using this characterisation we can easily obtain the following examples.

EXAMPLE (2.6) (Continuous pssMp and Bessel processes). *Given that the totality of Lévy processes with continuous paths is necessarily of the form $\xi_t = \epsilon B_t + \mu t$, $t \geq 0$, with $(B_t, t \geq 0)$ a standard Brownian motion and some $\epsilon, \mu \in \mathbb{R}$, we get that the totality of pssMp with continuous paths killed at its first hitting time of zero is obtained as a Lamperti transformation of a process of the latter form, with ϵ, μ and the self-similarity index appropriately chosen. For instance, when X is a standard Brownian motion killed at its first hitting time of 0, the self-similarity index is $1/2$, and the Lévy process is $\xi_t = B_t - \frac{t}{2}$, $t \geq 0$. Furthermore, taking $\epsilon = 1$, and $\mu = \frac{d}{2} - 1$, with $d > 0$, and $\alpha = 2$ we obtain a d-dimensional Bessel process. It is also an interesting exercise to prove this assertion using stochastic calculus.*

EXAMPLE (2.7) (Stable subordinators). *Let X be an α-stable Lévy process with non-decreasing paths, $0 < \alpha < 1$. X is a $1/\alpha$-pssMp. Its infinitesimal generator is*

$$\tilde{\mathcal{A}}f(x) = \int_0^\infty \Big(f(x+y) - f(x)\Big)\alpha c\frac{\mathrm{d}x}{x^{1+\alpha}}, \qquad c > 0.$$

By a change of variables

$$\tilde{\mathcal{A}}f(x) = x^{-\alpha}\int_0^\infty \left(f(xe^z) - f(x)\right)\frac{c\alpha e^z}{(e^x - 1)^{1+\alpha}}\mathrm{d}x,$$

and by Volkonskii's formula we obtain that the underlying Lévy process has jump measure

$$\Pi(\mathrm{d}x) = \frac{c\alpha e^z}{(e^x-1)^{1+\alpha}}\mathrm{d}x.$$

A family of processes associated to stable processes will be described in Section 2.2.

REMARK (2.8). *Although this will not be used in what follows, it is worth pointing out that the assumption made at the beginning of this section asking that the self-similarity index* $1/\alpha$ *is strictly positive is not essential. Indeed, if in Lamperti's transformation we take* $\alpha < 0$, *in order to make things consistent, we should just change the absorbing state of the pssMp to* $\{\infty\}$, *which will be reached in a finite time when the Lévy process* $\alpha\xi$ *will either jump in a finite time or drift towards* $-\infty$. *So, for a general self-similarity index* $1/\alpha$ *we should consider* $(0,\infty)\cup\{\Delta\}$*–valued pssMp where* Δ *is a cemetery state, that it is interpreted as* 0 *if* $\alpha > 0$, *and as* ∞ *if* $\alpha < 0$.

An useful property of pssMp is that the self-similarity property remains valid when we take powers or we make time changes with power functions. More precisely, given a $1/\alpha$-pssMp and a $\beta \in \mathbb{R}\setminus\{0\}$, the process Y defined by

$$Y_t := (X_t)^\beta, \qquad t \geq 0,$$

with $1/0$ taken as ∞, is a α/β-pssMp and it is the Lamperti transform of the Lévy process $\beta\xi$. This makes that in most of the cases there is no loss of generality in assuming that the self-similarity index equals 1. Now, for a $\gamma \neq -\alpha$, define a time change

$$D_t^\gamma = \inf\{u > 0 : \int_0^u (X_s)^\gamma \,\mathrm{d}s > t\}, \qquad t \geq 0.$$

The process W defined by

$$W_t = X_{D_t^\gamma}, t \geq 0,$$

is a $1/(\alpha+\gamma)$-pMasp and the underlying Lévy process remains ξ. The proof of this result is an easy consequence of the fact that time changes with respect to additive functionals preserve the strong Markov property and that, in this case, the time change also preserves the scaling property of the process X. This can also be easily seen by using Lamperti's transform and understanding the composition of time changes. These properties, together with some duality properties, where studied in [60], see also [41]. Other duality properties where obtained in [8].

(2.1) Defining a positive self-similar Markov process starting at 0. In his seminal paper, Lamperti [44] studied the forms in which a self-similar diffusion could be started from the state 0. Lamperti findings lead to the following question

> *Given a positive* $1/\alpha$*-self-similar Markov process,* $(X,(\mathbb{P}_x)_{x>0})$, *constructed via the Lamperti transformation of some Lévy process* ξ, *when does there exists a pssMp that behaves like* X *when it is in* $(0,\infty)$ *and that it is not trivially started from* 0*?*

In the case where the pssMp $(X,(\mathbb{P}_x)_{x>0})$ never hits zero this question has been answered in full generality by Bertoin and Caballero [5], Bertoin and Yor [8], Caballero and Chaumont [11], and Chaumont, Kyprianou, Pardo and Rivero [16], by providing necessary and sufficient conditions for the existence of a probability

measure $\mathbb{P}_{0+}$, that can be obtained as the weak limit of $\mathbb{P}_x$ as $x \downarrow 0+$, and under which the canonical process has the same transition semigroup as the one associated to $(X, (\mathbb{P}_x)_{x>0})$. Equivalently, we may ask under which conditions there exists a non-degenerate process $X^{(0)}$ that is the weak limit of $X^{(x)}$ as $x \to 0$.

Besides, when $(X, (\mathbb{P}_x)_{x>0})$ hits zero in a finite time to answer the above posed question one should look for all the recurrent extensions of it, that is the totality of positive self-similar Markov process that behave like the latter process before hitting zero for the first time but for which 0 is a recurrent and regular state. This problem has been studied by Lamperti [44] and Vuolle-Apiala [59], and solved in whole generality by Rivero [54, 55] and Fitzsimmons [27].

The main contributions of the papers quoted above will be summarized below.

2.1.1. *Entrance laws.* Bertoin and Caballero [5] and Bertoin and Yor [8] proved that, whenever the process drifts towards ∞, $\lim_{t\to\infty} X_t^{(x)} = \infty$, $\mathbb{P}_x$-a.s., the family of processes $X^{(x)}$ converges, as $x \downarrow 0$, in the sense of finite dimensional distributions towards $X^{(0)}$ if and only if the underlying Lévy process ξ in the Lamperti representation is such that

$$(H) \qquad \xi \text{ is non lattice and } \ 0 < m =: \mathbf{E}(\xi_1) \le \mathbf{E}(|\xi_1|) < +\infty.$$

In fact, the condition of ξ being non-lattice is not essential, whenever the process is lattice the limit exists when taken along adequately chosen subsequences. As proved by Caballero and Chaumont in [11], the latter condition is also a NASC for the weak convergence of the family $(X^{(x)})$, $x > 0$ on the Skohorod space of càdlàg trajectories. In the same article, the authors also provided a path construction of the process $X^{(0)}$. The entrance law of $X^{(0)}$ has been described in [5] and [8] as follows: for every $t > 0$ and for every measurable function $f : \mathbb{R}_+ \to \mathbb{R}_+$,

$$\mathbb{E}\left(f\left(X_t^{(0)}\right)\right) = \frac{1}{m}\mathbb{E}\left(I(-\alpha\xi)^{-1} f\left(\left(tI(-\alpha\xi)^{-1}\right)^{1/\alpha}\right)\right)$$

Caballero and Chaumont [11] actually studied the more general case where X is only required to be such that

$$\limsup_{t\to\infty} X_t^{(x)} = \infty, \qquad \mathbb{P}_x - \text{a.s. } \forall x > 0, \tag{2.9}$$

and they were able to prove that a necessary and sufficient condition for the weak convergence to hold is that the mean of the upward ladder height process, say $h = (h_t, t \ge 0)$, associated to ξ, is finite and a further technical condition. For further details see [11]. Latter Chaumont, Kyprianou Pardo, and Rivero [16] improved the result of Chaumont and Caballero [11] by showing that the technical condition is irrelevant. Moreover, these authors obtained an expression for the entrance law which extend that obtained in [5] and [8], namely

$$\mathbb{E}_{0+}(f(X_t)) = \int_0^\infty f\left(\frac{t^{1/\alpha}}{x^{1/\alpha}}\right)\frac{1}{x}\eta(\mathrm{d}x),$$

where η is a measure defined by

$$\eta(f) = \frac{1}{\alpha\mu_+}\int_{\mathbb{R}_+^3} \mathbf{P}(\widetilde{I} \in \mathrm{d}t)\widehat{\mathcal{V}}(\mathrm{d}x)\mathbf{P}_x^\dagger\left(\int_0^{\varsigma_0} e^{-\alpha\xi_u}\,\mathrm{d}u \in \mathrm{d}s\right) f\left(e^{\alpha x}(t+s)\right),$$

and $\int_0^\infty x^{-1}\eta(\mathrm{d}x) = 1$, with $\widetilde{I}$ the exponential functional of the negative of the Lévy process ξ conditioned to stay positive, $\widehat{\mathcal{V}}$ the renewal measure of the downward

ladder height process associated to ξ, and $\mathbf{P}^{\dagger}$ being the law of ξ conditioned to hit zero continuously.

Knowing the results of Bertoin and Caballero [5] and Bertoin and Yor [8], it may appear surprising that the searched necessary and sufficient condition for the weak convergence to hold is the finiteness of the mean of the upward ladder height process. An heuristic for the necessity of this condition is as follows. Assume there is a process $X^{(0)}$ that is obtained as a weak limit of $X^{(x)}$ as $x \downarrow 0+$. This process has the scaling property and hence for any $c > 0$

$$(cX^{(0)}_{tc^{-\alpha}}, t \geq 0) \overset{\text{Law}}{=} (X^{(0)}_t, t \geq 0).$$

For $a > 0$, let $T_a = \inf\{t > 0 : X^{(0)}_t > a\}$, this r.v. is a finite stopping time because the original process satisfies (2.9). By the scaling property we have the equality in law

$$X^{(0)}_{T_a} \overset{\text{Law}}{=} cX^{(0)}_{T_{a/c}}, \qquad \text{for all } c > 0.$$

Making $c \to \infty$ and some elementary manipulations we get

$$X^{(0)}_{T_a} \overset{\text{Law}}{=} a \exp\left\{\lim_{c\to\infty}\left(\log(X^{(0)}_{T_{a/c}}) - \log(a/c)\right)\right\}. \tag{2.10}$$

Finally, the process $\log X^{(0)}$ should have the same hitting probabilities as ξ, because $\log X^{(\cdot)}$ is obtained by time changing ξ, and hence we should have the equality in law

$$\lim_{c\to\infty}\left(\log\left(X^{(0)}_{T_{a/c}}\right) - \log(a/c)\right\} \overset{\text{Law}}{=} \lim_{z\to\infty} \xi_{T^+_{\log(z)}} - \log(z),$$

with $T^+_{\log(z)} = \inf\{t > 0 : \xi_t > \log(z)\}$. We conclude from (2.10) that the latter limit exists and it is not degenerate. It is well known that this is equivalent to the weak convergence of the overshoots of the underlying upward ladder height subordinator. The latter condition is in turn equivalent to the finiteness of the mean of the upward ladder height subodinator, see for instance [22]. The details about the whole argument can be found in [42] where the authors obtain precise descriptions about the distribution of random variables associated to the events of first passage above a level and last passage below a level, among other results.

Besides, observe that if the mean of the upward ladder height is finite then by the results in [5] and [11] the pssMp, $H = (H_t, t \geq 0)$, associated to the upward ladder height subordinator has a non-degenerate weak limit, $H^{(0)}$, and so one may wonder whether it is possible to understand the limit process $X^{(0)}$ using the process $H^{(0)}$. We remark that we may understand the process H as the process of the past supremum of X in an adequate time scale. The main motivation of the paper [16] was to construct the process H from the process X, via a time change, establish the convergence of H, and finally construct the limit process $X^{(0)}$ from $H^{(0)}$, by *hanging* into the paths of $H^{(0)}$ the excursions from the supremum. One of the main results in [16] is the following description.

THEOREM (2.11). *Let X be the Lamperti transform of a L.p. ξ that does not jump or drift towards $-\infty$, and define the maximum process $(M_s, s \geq 0)$, by $M_t := \sup_{s\leq t} X_s$, $t \geq 0$. We have the following facts.*

(i) There exists a function $j(\varepsilon)$ for $\varepsilon > 0$, such that

$$\lim_{\varepsilon\to 0} \frac{1}{j(\log(1+\varepsilon))} \int_0^t 1_{\left\{\frac{M_s}{X_s} \in [1, 1+\varepsilon[\right\}} \mathrm{d}s = L^{\Theta}_t,$$

uniformly over bounded intervals in probability. The process L^{Θ} is a local time at the past supremum for X, i.e. and additive functional whose support is given by the closure of the random set

$$\Theta := \{t \geq 0 : X_t = \sup_{0 \leq s \leq t} M_s\}.$$

(ii) Let $\{R_t, t \geq 0\}$ be the right continuous inverse of L^{Θ}, that is

$$R_t = \inf\{s > 0 : L^{\Theta}_s > t\}, \qquad t \geq 0,$$

and put $H_t := X_{R_t}$, $t \geq 0$. The process H is the Lamperti transform of the upward ladder height subordinator of ξ.

The term "hanging into the paths of H^{0+} the excursions from the supremum" is made precise in [16] by constructing an exit system associated to the random set Θ defined above, and then using this exit system to prove the convergence of the resolvent of X, as the starting point tends to 0, which gives the finite dimensional convergence. Finally, the weak convergence is obtained by proving tightness. We do not provide further details. Instead we mention that the papers [16], [20] and [42] contributed to the foundation of a fluctuation theory for pssMp analogous to the well developed one of real valued Lévy processes. In the paper [20] the path of a pssMp is decomposed into the path of the pre- and post-overall infimum. A precise description of these segments of paths is provided, and the limit of the path when the overall minimum tends to zero is obtained. In particular, it is shown that the post-minimum process converges to the path of $X^{(0)}$, and hence that the pre-infimum path squeezes to the path equal to zero with 0 length when the starting point tends to 0.

2.1.2. *Recurrent extensions.* We now deal with the results around the question *What are the positive α–self-similar Markov processes $\widetilde{X}$ which behave like $(X, \mathbb{P})$ up to the first hitting time of 0 for $\widetilde{X}$ and such that 0 is a regular and recurrent state?* A process that has this characteristics is usually called a recurrent extension of the process $(X, \mathbb{P})$. Lamperti [44] solved this question in the special case of a Brownian motion killed at 0, using results specific to Brownian motion. After Lamperti, Vuolle-Apiala [59] used excursion theory to give a more general answer under some regularity assumptions for the resolvent of the process $(X, \mathbb{P})$. Then the question was solved in full generality by Rivero [54], [55], and Fitzsimmons [27]. In order to describe the results in those papers we introduce some further notions, but before we mention that a different approach using stochastic differential equations has been used in [24] (see Section 2.1.3).

Definition (2.12) (Self-similar excursion measures). A measure $\mathbf{n}$ on $(\mathbb{D}^+, \mathcal{G}_\infty)$ having infinite mass is a self-similar excursion measure compatible with $(X, \mathbb{P})$ if

(i) $\mathbf{n}$ is carried by

$$\{\omega \in \mathbb{D} + |0 < T_0 < \infty \text{ and } X_t(\omega) = 0, \ \forall t \geq T_0\};$$

(ii) For every bounded $\mathcal{G}_\infty$-measurable functional H and each $t > 0$ and $\Lambda \in \mathcal{G}_t$,

$$\mathbf{n}(H \circ \theta_t, \Lambda \cap \{t < T_0\}) = \mathbf{n}\left(\mathbb{E}_{X_t}(H), \Lambda \cap \{t < T_0\}\right),$$

with θ_t the shift operator;

(iii) $\mathbf{n}\left(1 - e^{-T_0}\right) < \infty$;

(iv) there exists a $\gamma \in]0,\alpha[$ such that for every $c>0$ the image of $\mathbf{n}$ under the mapping $H_c : \mathbb{D}^+ \to \mathbb{D}^+$, defined by $H_c(\omega)(t) = c\omega(tc^{-1/\alpha})$, for $t>0$, is

$$\mathbf{n} \circ H_c = c^{\gamma/\alpha}\,\mathbf{n}.$$

It is well known in the theory of Markov processes that a way to construct recurrent extensions of self-similar Markov processes is by means of the so called Itô's program or pathwise approach, which consists on *pasting together excursions*. Precise results about this topic can be found in [10] and [54]. The main results from the latter references allow us to ensure that there is a bijection between the existence of self-similar recurrent extensions and self-similar excursion measures compatible to $(X,\mathbb{P})$. Actually, the latter is the Itô excursion measure for the excursions from zero of the self-similar recurrent extension of $(X,\mathbb{P})$. We will now describe necessary and sufficient conditions for the existence of such a measure.

Vuolle-Apiala [59] proved that there are two types of excursion measures, namely those for which the recurrent extension *exits* 0 *by jumps*, which in terms of the excursion measure means $\mathbf{n}(X_{0+}=0)=0$, and those for which the recurrent extension leaves zero continuously, $\mathbf{n}(X_{0+}>0)=0$. And furthermore, a self-similar excursion measure is either of one type or the other, but not both. It has been shown in [54] that the reason for this is that they have different self-similarity index. Vuolle-Apiala proved that a consequence of the scaling property is that all the self-similar excursion measures such that $\mathbf{n}(X_{0+}=0)=0$, can be written as

$$\mathbf{n}(\cdot) = c_{\alpha,\beta}\int_{x>0}\frac{\mathrm{d}x}{x^{1+\beta}}\mathbb{P}_x(\cdot), \tag{2.13}$$

for some β such that $\beta/\alpha \in]0,1[$ and $c_{\alpha,\beta}\in]0,\infty[$, is a normalizing constant. Thus, to determine the existence of a recurrent extension that leaves 0 by a jump all we are ought to do is to verify when a measure of this form bears all the condition to be a self-similar excursion measure. That is the purpose of the following theorem.

THEOREM (2.14). *Let $(X,\mathbb{P})$ be an α-self-similar Markov process that hits the cemetery point 0 in a finite time a.s. and $(\xi,\mathbf{P})$ the Lévy process associated to it via Lamperti's transformation. For $0<\beta<\alpha$, the following are equivalent*

(i) $\mathbf{E}(e^{\beta\xi_1}, 1<\zeta)<1$,

(ii) $\mathbf{E}\left(\left(\int_0^\infty \exp\{\alpha\xi_s\}\mathrm{d}s\right)^{\beta/\alpha}\right)<\infty$,

(iii) There exists a recurrent extension of $(X,\mathbb{P})$, say $X^{(\beta)}$, that leaves 0 by a jump and its associated excursion measure $\mathbf{n}^\beta$ is such that

$$\mathbf{n}^\beta(X_{0+}\in \mathrm{d}x) = c_{\alpha,\beta}\beta x^{-1-\beta}\mathrm{d}x, \qquad x>0,$$

where $c_{\alpha,\beta}$ is a constant.

In this case, the process $X^{(\beta)}$ is the unique recurrent extension of $(X,\mathbb{P})$ that leaves 0 by a jump distributed as above.

The proof of this theorem resides in the fact that a measure of the form in (2.13) satisfies the conditions (i), (ii) and (iv) from the Definition (2.12), as it can be easily verified, but the condition (iii) is only satisfied when the condition (ii) of Theorem (2.14) is satisfied, and hence what is left to prove is the equivalence between (i) and (ii), because the equivalence between (ii) and (iii) is obtained from the previous discussion.

It is actually more difficult to establish the existence of entrance laws that are carried by the paths that leave 0 continuously. The definitive answer is given in the following result obtained by Rivero [54, 55] and Fitzsimmons [27].

THEOREM (2.15). *Let $(X,\mathbb{P})$ be an α-self-similar Markov process that hits its cemetery state 0 in a finite time $\mathbb{P}$-a.s. and $(\xi,\mathbf{P})$ be the Lévy process associated to $(X,\mathbb{P})$ via Lamperti's transformation. The following are equivalent:*

(i) $\exists\, \theta \in]0,\alpha[$, *s. t.* $\mathbf{E}(e^{\theta\xi_1}, 1<\zeta) = 1$, ***Cramér's condition.***

(ii) There exists a recurrent extension of $(X,\mathbb{P})$ that leaves 0 continuously and such that its associated excursion measure from 0, say $\mathbf{N}$, satisfies $\mathbf{N}(1-e^{-T_0}) = 1$.

In this case, the recurrent extension in (ii) is unique and the entrance law associated to the excursion measure $\mathbf{N}$ is given by, for any f positive and measurable

$$\mathbf{N}(f(X_t), t<T_0) = \frac{1}{t^{\theta/\alpha}\Gamma(1-(\theta/\alpha))\mathbf{E}^\natural(J^{(\theta/\alpha)-1})}\,\mathbf{E}^\natural\left(f\left(\frac{t^{1/\alpha}}{J^{1/\alpha}}\right)J^{(\theta/\alpha)-1}\right),$$

for $t>0$, where $\mathbf{P}^\natural = e^{\theta\xi_t}\mathbf{P}$ on $\sigma(\xi_s, s\le t)$ and $J = \int_0^\infty \exp\{-\alpha\xi_s\}\mathrm{d}s$.

The above description of the measure $\mathbf{N}$ is reminiscent of Imhof's construction of the excursion measure of the Brownian process out from 0 which relates the law of a brownian motion conditioned to stay positive and started from zero, that is a Bessel process issued from 0, and the excursion measure. Further results in this direction and a description of the excursion measure conditionally on the length as well as its image under time reversal are provided in [54]. A description of the excursion measure in terms of the height of the excursion is provided in [3].

2.1.3. *A stochastic differential equation approach.* Motivated by the problem of existing zero and the description of the recurrent extension of positive self-similar Markov process, Barczy and Döring [24] proposed a stochastic differential equation (SDE for short) approach. More precisely, recall that the Lévy-Itô representation of a Lévy process $\xi = (\xi_t, t\ge 0)$ is given as follows

$$\xi_t = bt + \sigma B_t + \int_0^t\int_{\{|u|\le 1\}} u\widetilde{\mathcal{N}}(\mathrm{d}s,\mathrm{d}u) + \int_0^t\int_{\{|u|\ge 1\}} u\mathcal{N}(\mathrm{d}s,\mathrm{d}u),$$

where $b\in\mathbb{R}$, $\sigma\ge 0$, B is a Brownian motion and $\mathcal{N}$ is an independent Poisson random measure on $(0,\infty)\times\mathbb{R}$ with intensity $\mathrm{d}s\otimes\Pi(\mathrm{d}u)$ and $\widetilde{\mathcal{N}}$ represents its compensated version. Assuming that $\mathbf{E}[e^{\xi_1}]<\infty$, hence the proposed SDE can be written as follows

$$\begin{aligned} X_t = x + \left(\log\mathbf{E}[e^{\xi_1};\zeta>1]\right)t + \sigma\int_0^t\sqrt{X_t}\mathrm{d}B_s - \int_0^t\int_0^\infty \mathbf{1}_{\{rX_{s-}\le 1\}}X_{s-}\widetilde{\mathcal{M}}(\mathrm{d}s,\mathrm{d}r) \\ + \int_0^t\int_0^\infty\int_{-\infty}^\infty \mathbf{1}_{\{rX_{s-}\le 1\}}X_{s-}(e^u-1)\widetilde{\mathcal{N}}_1(\mathrm{d}s,\mathrm{d}r,\mathrm{d}u), \end{aligned}$$

for $t\le T_0$. Here B is a Brownian motion, $\mathcal{N}_1$ is an independent Poisson random measure on $(0,\infty)\times(0,\infty)\times\mathbb{R}$ with intensity $\mathrm{d}s\otimes\mathrm{d}r\otimes\Pi(\mathrm{d}u)$ and $\mathcal{M}$ is an independent Poisson random measure on $(0,\infty)\times(0,\infty)$ with intensity $q\mathrm{d}s\otimes\mathrm{d}r$. The random measures $\widetilde{\mathcal{M}}$ and $\widetilde{\mathcal{N}}$ represents the compensated version of $\mathcal{M}$ and $\mathcal{N}$, respectively. The intuition of the above SDE follows from applying Itô's formula to e^{ξ_t} and afterwards including a correction which is given by the random time change.

It is important to note that whenever $0 < \log \mathbf{E}[e^{\xi_1}; \zeta > 1] < \infty$, then this new representation is not restricted to $t \le T_0$. We also note that the SDE defined above posses weak solutions up to a first hitting time. If ξ posses only negative jumps and satisfies that $0 < \log \mathbf{E}[e^{\xi_1}; \zeta > 1]$, then for any initial condition $x > 0$ there is a pathwise unique non-negative strong solution $(X_t, t \ge 0)$ which is self-similar with index $\alpha = 1$ and such that $(X_t, 0 \le t \le T_0)$ its underlying Lévy process in the Lamperti transform has the same law as ξ killed at rate q. If the Lévy process ξ satisfies the Cramér condition then $(X_t, t \ge 0)$ is the unique recurrent self-similar extension of $(X_t, 0 \le t \le T_0)$. Moreover if ξ does not drift to $-\infty$ or ξ drifts to $-\infty$ and it satisfies the Cramér condition, then the process $(X_t, t \ge 0)$ is the unique strong solution of the above SDE with initial condition $X_0 = 0$.

(2.2) Exponential functionals. As we have seen in previous sections a recurrent object in the theory of pssMp is the so-called *exponential functional* of a Lévy process ξ with lifetime ζ, i.e.

$$I_\zeta(\xi) := \int_0^\zeta \exp\left\{\xi_s\right\} \mathrm{d}s.$$

For instance, a consequence of Lamperti's transformation is that the first hitting time of a positive valued ssMp has the same law as the exponential functional of a Lévy process; we have also seen that it describes the entrance law of a pssMp that never hits zero, or the entrance law under the excursion measure of the recurrent extension associated to a pssMp that hits zero in a finite time. Latter in this note, Section 2.3, we will see that the density of an exponential functional plays a crucial role in establishing integral test for describing the upper and lower envelopes of a pssMp; also we will see in Section 2.4 that the Yaglom limit for a pssMp is determined by the asymptotic behaviour of the tail distribution of the first hitting time of zero. Hence a good understanding of the law of an exponential functional is necessary in order to obtain precise information about pssMp's. Furthermore, one could say that exponential functionals of Lévy processes and pssMp leave in symbiosis because facts about pssMp have been used to obtain properties of exponential functionals.

Moreover, this is not the only fact that has motivated many research works on the topic over the last two decades. The law of $I_\zeta(\xi)$ plays an important role in many other areas of probability theory, for instance in fragmentation, coalescence and branching processes, financial and insurance mathematics, Brownian motion in hyperbolic spaces, random processes in random environment, etc. For more details about these topics and other aspects not covered in this section we refer to the survey paper [9]. In this section we intend to provide a collection of results that partially complements the latter paper.

Because in our setting ζ is taken as the life time of the Lévy process ξ, we will hence focus in the case where $\zeta = \mathbf{e}_q$, an exponential random variable with parameter $q \ge 0$ which is independent of the process ξ. Many authors have been interested in the existence (and determine explicitly, as well) of the density associated to $I_{\mathbf{e}_q}(\xi)$. When $q = 0$, then $\mathbf{e}_q$ is understood as ∞. In this case, we assume that the process ξ drifts towards $-\infty$ since it is a necessary and sufficient condition for the almost sure finiteness of $I(\xi) := I_\infty(\xi)$, see for instance Theorem 1 in Bertoin and Yor [9]. Carmona, Petit and Yor [15] were the first in studying the existence of the density of $I_\infty(\xi)$. More precisely, they proved the existence of

such density in the case when the jump structure of the Lévy process is of finite variation and also provided an integral equation that the density must hold, we will recall the integral equation below. Recently Bertoin et al. [6] (see Theorem 3.9) proved the existence of the density in the general case.

The first result that we present in this section is about the existence of the density of $I_{\mathbf{e}_q}(\xi)$ in the case $q>0$, in terms of its associated 1-positive self-similar process $(X,\mathbb{P}_1)$, it has been obtained in [50].

THEOREM (2.16). *Let $q>0$, then the function*

$$h(t):=q\mathbb{E}_1\left[\frac{1}{X_t}\mathbf{1}_{\{t<T_0\}}\right],\qquad t\geq 0,$$

is a density for the law of $I_{\mathbf{e}_q}(\xi)$.

A consequence of this result in the case where ξ is a subordinator gives the following important property of h.

COROLLARY (2.17). *Assume $q>0$ and that ξ is a subordinator. Then the law of the r.v. $I_{\mathbf{e}_q}(\xi)$ is a mixture of exponentials, that is its law has a density h on $(0,\infty)$ which is completely monotone. Furthermore, $\lim_{t\downarrow 0}h(t)=q$.*

Carmona, et al. [15] integral equation give some information of the density and allow us to compute it explicitly in some particular cases. A generalizaton and extension of this integral equation in the case of the negative of a subordinator has been obtained in [50]

THEOREM (2.18). *Assume that $\xi=-\sigma$, with σ a subordinator with drift $c\geq 0$, killing term q and Lévy measure Π. Let $q\geq 0$. The random variable $I_{\mathbf{e}_q}$ has a density that we denote by k, and it solves the equations*

$$\int_y^\infty k(x)\mathrm{d}x=\int_0^\infty k(ye^x)U_q(\mathrm{d}x),\qquad \textit{almost everywhere,} \tag{2.19}$$

and

$$(1-cx)k(x)=\int_x^\infty \overline{\Pi}(\log(y/x))k(y)\mathrm{d}y+q\int_x^\infty k(y)\mathrm{d}y,\qquad x\in(0,1/c). \tag{2.20}$$

with $\mathbf{E}\left[\int_0^{\mathbf{e}_q}1_{\{\sigma_t\in dx\}}\mathrm{d}t\right]=U_q(\mathrm{d}x)$, on $x\geq 0$, and $\overline{\Pi}(y):=\Pi(y,\infty)$, for $y>0$. Conversely, if a density on $(0,1/c)$ satisfies any of the equations (2.19) or (2.20) then it is the density of $I_{\mathbf{e}_q}$.

There are two main approaches which have been developed and used to extract more information about the law of the exponential functional. The first one uses the fact that the Mellin transform of $I_{\mathbf{e}_q}(\xi)$ is solution of the functional equation,

$$\mathbf{E}\left[I_{\mathbf{e}_q}(\xi)^{s-1}\right]=\frac{\psi_q(s)}{s}\mathbf{E}\left[I_{\mathbf{e}_q}(\xi)^s\right], \tag{2.21}$$

where $\psi_q(\lambda)=-\ln\mathbf{E}[e^{\lambda\xi_1},1<\mathbf{e}_q]$. The above equation, when $q=0$, appears for the first time in Carmona et al. [15] and was extended by Maulik and Zwart [45]. When $q=0$, the equation (2.21) can be solved explicitly in the case when ξ is the negative of a subordinator or a spectrally positive Lévy process, which, in both cases, determine the law of $I_{\mathbf{e}_q}(\xi)$. More precisely, let $-\xi$ be a subordinator and $\Phi(\lambda)=-\ln\mathbf{E}[e^{\lambda\xi_1}]$. Carmona et al. [15] noted that the law of the exponential

functional of a subordinator is determined by its entire moments which are given by the identity

$$\mathbf{E}\left[I(\xi)^k\right] = \frac{k}{\Phi(k)}\mathbf{E}\left[I(\xi)^{k-1}\right] = \frac{k!}{\Phi(1)\cdots\Phi(k)}, \qquad \text{for} \quad k = 1,2,\ldots$$

We note that this equation can be solved explicitly in many situations, see for instance Bertoin and Yor [9]. Similarly, if ξ is a spectrally positive Lévy process, Bertoin and Yor [9] proved that the law of $I(\xi)$ is determined by its negative entire moments and can be expressed in the form

$$\mathbf{E}\left[I(\xi)^{-(k+1)}\right] = \frac{\Psi(k)}{k}\mathbf{E}\left[I(\xi)^{-k}\right] = m\frac{\Psi(1)\cdots\Psi(k-1)}{(k-1)!}, \qquad \text{for} \quad k = 1,2,\ldots$$

where $\Psi(\lambda) = \ln \mathbf{E}[e^{-\lambda\xi_1}]$ and with the convention that the right-hand side equals m for $k = 1$.

One can prove that if Cramér's condition is satisfied then the Mellin transform of $I_{\mathbf{e}_q}(\xi)$ satisfies the functional identity (2.21), however it is clear that there are infinitely many functions which satisfy the same functional identity. The next result obtained by Kuznetsov and Pardo [39] tells us that if we have found a function $f(s)$ which satisfies (2.21), and if we can verify two conditions about the zeros of this function and its asymptotic behaviour, then we can in fact uniquely identify the Mellin transform of $I_{\mathbf{e}_q}(\xi)$.

PROPOSITION (2.22). *Assume that there exists $z_0 > 0$ such that $\psi_q(z)$ is finite for all $z \in (0, z_0)$ and $\psi_q(\theta) = 0$ for some $\theta \in (0, z_0)$. If $f(s)$ satisfies the following three properties*

(i) $f(s)$ *is analytic and zero-free in the strip* $\mathrm{Re}(s) \in (0, 1+\theta)$,
(ii) $f(1) = 1$ *and* $f(s+1) = sf(s)/\psi_q(s)$ *for all* $s \in (0,\theta)$,
(iii) $|f(s)|^{-1} = o(\exp(2\pi|\mathrm{Im}(s)|))$ *as* $\mathrm{Im}(s) \to \infty$, *uniformly in* $\mathrm{Re}(s) \in (0, 1+\theta)$,

then $\mathbf{E}[I_{\mathbf{e}_q}(\xi)^{s-1}] \equiv f(s)$ *for* $\mathrm{Re}(s) \in (0, 1+\theta)$.

In particular, this proposition can be used to provide a very simple and short proof of the well-known result on exponential functional of Brownian motion with drift and of the recent results on exponential functionals of processes with double-sided hyper-exponential jumps (see [14]).

Recently in [39], the authors found a particular class of Lévy processes, called hypergeometric Lévy processes, for which the solution of the functional equation can directly be guessed from (2.21) and verified using Proposition 1, and derived the law of $I_{\mathbf{e}_q}(\alpha\xi)$ for an specific value of q.

Hypergeometric Lévy processes were first introduced in [42] and constructed using Vigon's theory of philanthropy (see [58]). The class of processes that we present next should be considered as a subclass of the hypergeometric processes studied in [42] and as a generalization of Lamperti-stable processes, which were introduced by Caballero and Chaumont [11].

We start by defining its Laplace exponent $\psi_q(z)$ as

$$\psi_q(z) = \frac{\Gamma(1-\beta+\gamma-z)}{\Gamma(1-\beta-z)}\frac{\Gamma(\hat{\beta}+\hat{\gamma}+z)}{\Gamma(\hat{\beta}+z)}, \tag{2.23}$$

where $(\beta, \gamma, \hat{\beta}, \hat{\gamma})$ belong to the admissible set of parameters

$$\mathcal{A} = \{\beta \le 1,\ \gamma \in (0,1),\ \hat{\beta} \ge 0,\ \hat{\gamma} \in (0,1)\}.$$

Let

$$\eta = 1-\beta+\gamma+\hat{\beta}+\hat{\gamma}.$$

The Levy density of hypergeometric Lévy processes can be computed explicitly, see Proposition 1 in [39]. Moreover, if $\beta<1$ and $\hat{\beta}>0$ the process ξ is killed at rate

$$q=\psi_q(0)=\frac{\Gamma(1-\beta+\gamma)}{\Gamma(1-\beta)}\frac{\Gamma(\hat{\beta}+\hat{\gamma})}{\Gamma(\hat{\beta})}.$$

The process ξ drifts to $+\infty$, $-\infty$ or oscillates whenever $\beta=1$ and $\hat{\beta}>0$, $\beta<1$ and $\hat{\beta}=0$ or $\beta=1$ and $\hat{\beta}=0$. The process ξ has no Gaussian component. When $\gamma+\hat{\gamma}<1$ ($1\le\gamma+\hat{\gamma}<2$) the process has paths of bounded variation and no linear drift (paths of unbounded variation).

Three Lamperti-stable processes $\xi^*,\xi^{\uparrow}$ and $\xi^{\downarrow}$ were introduced by Caballero and Chaumont [11] by applying the Lamperti transformation to the positive self-similar Markov processes construced from a stable process. In particular, the process ξ^* is obtained from a stable process started at $x>0$ and killed upon exit from the positive half-line, while the process $\xi^{\uparrow}$ $\{\xi^{\downarrow}\}$ is obtained from a stable process conditioned to stay positive {conditioned to hit zero continuously}. We refer to [11, 12, 17] for all the details on these processes.

The Lamperti-stable processes ξ^*, $\xi^{\uparrow}$, $\xi^{\downarrow}$ can be identified as hypergeometric processes with the following sets of parameters. Let $\delta=1/\alpha$. From the definition

	β	γ	$\hat{\beta}$	$\hat{\gamma}$
ξ^*	$1-\alpha(1-\rho)$	$\alpha\rho$	$1-\alpha(1-\rho)$	$\alpha(1-\rho)$
$\xi^{\uparrow}$	1	$\alpha\rho$	1	$\alpha(1-\rho)$
$\xi^{\downarrow}$	0	$\alpha\rho$	0	$\alpha(1-\rho)$

of the Laplace exponent (2.23) we find that ξ satisfies Cramér's condition, that is to say $\mathbf{E}[\exp(\hat{\beta}\xi_1)]=1$, therefore applying Lemma 2 from Rivero [55] we conclude that the Mellin transform of $I_{\mathbf{e}_q}(\alpha\xi)$ exists for $s\in(0,1+\hat{\beta}\delta)$. In order to describe the results about the law of $I_{\mathbf{e}_q}(\alpha\xi)$, we need to define the double gamma function, $G(z;\tau)$. The double gamma function is defined by an infinite product in Weierstrass's form

$$G(z;\tau)=\frac{z}{\tau}e^{a\frac{z}{\tau}+b\frac{z^2}{2\tau}}\prod_{m\ge0}{\prod_{n\ge0}}'\left(1+\frac{z}{m\tau+n}\right)e^{-\frac{z}{m\tau+n}+\frac{z^2}{2(m\tau+n)^2}},\quad|\arg(\tau)|<\pi,\quad z\in\mathbb{C}.$$

Here the prime in the second product means that the term corresponding to $m=n=0$ is omitted. Note that by definition $G(z;\tau)$ is an entire function in z and if $\tau\notin\mathbb{Q}$ it has simple zeros on the lattice $m\tau+n$, $m\le0$, $n\le0$. We refer to Kuznetsov [36] or Kuznetsov and Pardo [39] for more properties of this function. The following result, lifted from [39], characterize the Mellin transform of the exponential functional of hypergeometric Lévy processes.

THEOREM (2.24). *Assume that $\alpha>0$, $(\beta,\gamma,\hat{\beta},\hat{\gamma})\in\mathcal{A}$ and $\hat{\beta}>0$. Then for $s\in\mathbb{C}$ we have*

$$\mathbf{E}[I_{\mathbf{e}_q}(\alpha\xi)^{s-1}]\equiv C\Gamma(s)\frac{G((1-\beta)\delta+s;\delta)}{G((1-\beta+\gamma)\delta+s;\delta)}\frac{G((\hat{\beta}+\hat{\gamma})\delta+1-s;\delta)}{G(\hat{\beta}\delta+1-s;\delta)},\tag{2.25}$$

where the constant C is such that the above identity equals 1 when $s = 1$.

We now want to study the density of the exponential functional, which is defined by

$$p(x) = \frac{\mathrm{d}}{\mathrm{d}x}\mathbf{P}(I_{\mathbf{e}_q}(\alpha\xi) \le x), \quad x \ge 0.$$

In order to do so, we have to compute the inverse Mellin transform of (2.25) which is not a simple inversion exercise. From the paper by Kuznetsov & Pardo [39] in the next result it is provided an asymptotic expansion of the density $p(x)$ in the case when $\alpha \notin \mathbb{Q}$.

THEOREM (2.26). *Assume that $\alpha \notin \mathbb{Q}$. Then*

$$p(x) \sim \sum_{n\ge 0} a_n x^n + \sum_{m\ge 0}\sum_{n\ge 0} b_{m,n} x^{(m+1-\beta+\gamma)\delta+n}, \quad x \to 0^+,$$

$$p(x) \sim \sum_{m\ge 0}\sum_{n\ge 0} c_{m,n} x^{-(m+\hat{\beta})\delta-n-1}, \qquad x \to +\infty.$$

The series $(a_n)_{n\ge 0}, (b_{m,n})_{m,n\ge 0}$ and $(c_{m,n})_{m,n\ge 0}$ can be computed explicitly. We refer to Kuznetsov and Pardo [39] for more details about this series.

It turns out that for almost all parameters α the asymptotic series from above converge to $p(x)$ for all $x > 0$. In order to state this result, we need to define the following set of real numbers.

Definition (2.27). Let $\mathcal{L}$ be the set of real irrational numbers x, for which there exists a constant $b > 1$ such that the inequality

$$\left|x - \frac{p}{q}\right| < \frac{1}{b^q}$$

is satisfied for infinitely many integers p and q.

For more details about this set of irrational numbers see Kuznetsov [36] and Kuznetsov and Hubalek [32]. The following result was obtained in [39].

THEOREM (2.28). *Assume that $\alpha \notin \mathcal{L} \cup \mathbb{Q}$. Then for all $x > 0$*

$$p(x) = \begin{cases} \sum_{n\ge 0} a_n x^n + \sum_{m\ge 0}\sum_{n\ge 0} b_{m,n} x^{(m+1-\beta+\gamma)\delta+n}, & \text{if } \gamma+\hat{\gamma} < 1, \\ \sum_{m\ge 0}\sum_{n\ge 0} c_{m,n} x^{-(m+\hat{\beta})\delta-n-1}, & \text{if } \gamma+\hat{\gamma} > 1. \end{cases}$$

It is worth recalling that, in general, it is not an easy exercise to invert the Mellin (or moments) transform of $I_{\mathbf{e}_q}(\xi)$ since a fine analysis of its asymptotic behavior is required.

The second methodology to obtain information about the law of $I_{\mathbf{e}_q}(\xi)$ is based on the well-known relation between this and the distribution of the absorption time of positive self-similar Markov processes. Indeed, in Carmona et al. [15] it is shown that the law of $I_{\mathbf{e}_q}(\xi)$ can be expressed as an invariant function of a transient Ornstein-Uhlenbeck process associated to self-similar Markov process.

In Pardo et al. [49], starting from a large class of Lévy processes and assuming that $q = 0$, it is shown that the law of $I(\xi)$ can be factorized into the product of independent exponential functionals associated with two companion Lévy processes, namely the descending ladder height process of ξ and a spectrally positive Lévy process constructed from its ascending ladder height process. It is known

that these two subordinators appear in the Wiener-Hopf factorization of Lévy processes. The laws of these exponential functionals are uniquely determined either by their positive or negative integer moments. Moreover, whenever the law of any of these can be expanded in series we can in general develop the law of $I(\xi)$ in series. Thus, for example, the requirements put on the Lévy measure of ξ in Kuznetsov and Pardo [39] can be relaxed to conditions only on the positive jumps (the Lévy measure on the positive half-line) of ξ thus enlarging considerably the class of Lévy processes ξ, for which we can obtain a series expansion of the law of $I(\xi)$.

Before stating the next results let us introduce some notation. First, since in our setting ξ drifts to $-\infty$, it is well-known that the ascending (resp. descending) ladder height process $H^+ = (H^+(t))_{t\geq 0}$ (resp. $H^- = (-H^-(t))_{t\geq 0}$) is a killed (resp. proper) subordinator. Then, we write, for any $z \in i\mathbb{R}$,

$$\phi_+(z) = \log \mathbf{E}\left[\exp(zH^+(1))\right] = \delta_+ z + \int_{(0,\infty)} (e^{zy} - 1)\mu_+(\mathrm{y}) - k_+, \tag{2.29}$$

where $\delta_+ \geq 0$ is the drift and $k_+ > 0$ is the killing rate. Similarly, with $\delta_- \geq 0$, we have

$$\phi_-(z) = \log \mathbf{E}[\exp(zH^-(1))] = -\delta_- z - \int_{(0,\infty)} (1 - e^{-zy})\mu_-(\mathrm{y}). \tag{2.30}$$

We recall that the integrability condition $\int_0^\infty (1 \wedge y)\mu_\pm(dy) < \infty$ holds. The Wiener-Hopf factorization then reads off as follows

$$\Psi(z) = -\phi_+(z)\phi_-(z), \text{ for any } z \in i\mathbb{R}. \tag{2.31}$$

Definition (2.32). We denote by $\mathcal{P}$ the set of positive measures on $\mathbb{R}_+$ which admit a non-increasing density.

Before we formulate the next result we introduce the two main hypothesis:

($\mathcal{H}_1$) Assume further that $-\infty < \mathbf{E}[\xi_1]$ and that one of the following conditions holds:
- **i):** $\mu_+ \in \mathcal{P}$ and there exists $z_+ > 0$ such that for all z with, $\Re(z) \in (0, z_+)$, we have $|\Psi(z)| < \infty$.
- **ii):** $\Pi_+ \in \mathcal{P}$.

($\mathcal{H}_2$) Assume that
- **i):** $\mu_+ \in \mathcal{P}$, $k_+ > 0$ and $\mu_- \in \mathcal{P}$.

Then the following result has been proved by Pardo, Patie & Savov [49].

THEOREM (2.33). *Assume that ξ is a Lévy process that drifts to $-\infty$ with characteristics of the ladder height processes as in (2.29) and (2.30). Let either ($\mathcal{H}_1$) or ($\mathcal{H}_2$) holds. Then, in both cases, there exists a spectrally positive Lévy process Y with a negative mean whose Laplace exponent ψ_+ takes the form*

$$\psi_+(-s) = -s\phi_+(-s) = \delta_+ s^2 + k_+ s + s^2 \int_0^\infty e^{-sy}\mu_+(y,\infty)dy,\ s \geq 0, \tag{2.34}$$

and the following factorization holds

$$\mathrm{I}(\xi) \stackrel{d}{=} \mathrm{I}(-H^-) \times \mathrm{I}(Y) \tag{2.35}$$

where $\stackrel{d}{=}$ stands for the identity in law and $\times$ for the product of independent random variables.

The above result has been recently improved by Patie and Savov [52]. The obtained identity can be looked from another perspective. Let us have two subordinators with Lévy measures $\mu_\pm$ such that $\mu_+ \in \mathcal{P}$, $k_+ > 0$ and $\mu_- \in \mathcal{P}$. Then according to Vigon's theory of philanthropy, see [58], we can construct a process ξ such that its ladder height processes have exponents as in (2.29) and (2.30) and hence ξ satisfies the conditions of the previous Theorem. Therefore this method can be used to synthesize examples starting from the building blocks, i.e. the ladder height processes. This was observed in [49].

COROLLARY (2.36). *Let $\mu_\pm$ be the Lévy measures of two subordinators and $\mu_+ \in \mathcal{P}$, $k_+ > 0$ and $\mu_- \in \mathcal{P}$. Then there exists a Lévy process which drifts to $-\infty$ whose ascending and descending ladder height processes have the Laplace exponents respectively given by (2.29) and (2.30). Then all the claims of the Theorem (2.33) hold and in particular we have the factorization (2.35).*

Another interesting problem is determining the behaviour of the density of the exponential functional $I(\xi)$ at 0 and at ∞. This problem has been recently studied by Kuznetzov [37] for Lévy processes with rational Laplace exponent (at 0 and at ∞) and by Patie [51] for spectrally negative Lévy processes (at ∞). In most of the applications, it is enough to have estimates of the tail behaviour $\mathbf{P}(I(\xi) \le t)$ when t goes to 0 and/or $\mathbf{P}(I(\xi) \ge t)$ when t goes to ∞. The tail behaviour $\mathbf{P}(I(\xi) \le t)$ was studied by Pardo [48] in the case where $-\xi$ is spectrally positive Lévy process and its Laplace exponent is regularly varying at infinity with index $\gamma \in (1,2)$. That is the content of the following result.

PROPOSITION (2.37). *Let $I(\xi)$ be the exponential functional associated to a spectrally negative Lévy process ξ. Suppose that ψ, the Laplace exponent of $-\xi$, varies regularly at $+\infty$ with index $\beta \in (1,2)$. Then*

$$-\log \mathbf{P}\left(I\left(\xi\right) < 1/x\right) \sim (\beta-1)\overleftarrow{H}(x) \quad \text{as} \quad x \to +\infty, \tag{2.38}$$

where

$$\overleftarrow{H}(x) = \inf\left\{s > 0\,,\ \psi(s)/s > x\right\}.$$

In the case where the Lévy process is the negative of a subordinator several results are available to describe the left and right tail distribution of the exponential functional. Haas and Rivero [30] studied the case when ξ is the negative of a subordinator under several different frameworks, obtaining very precise estimates of the right tail behavior of the law of I, and described the maximum domain of attraction of I. One of the main results in [30] is the following description of the hazard rate function of an exponential functional of the negative of a subordinator.

THEOREM (2.39). *Let $-\xi$ be a subordinator, $q \ge 0$, and $\boldsymbol{\varphi}_{\Pi,\boldsymbol{q}}$ the function defined in (2.64) below. $I_{\mathbf{e}_q}(\xi)$ has a density k such that*

1. *if $a = 0$ and $\liminf_{x\to 0+} \dfrac{x\overline{\Pi}(x)}{\displaystyle\int_0^x \overline{\Pi}(u)du} > 0$,*

$$\frac{k(t)}{\mathbf{P}(I_{\mathbf{e}_q}(\xi) > t)} \sim_{t\to\infty} \frac{\boldsymbol{\varphi}_{\Pi,\boldsymbol{q}}(t)}{t}, \quad -\log \mathbf{P}(I_{\mathbf{e}_q}(\xi) > t) \sim_{t\to\infty} \int_c^t \frac{\boldsymbol{\varphi}_{\Pi,\boldsymbol{q}}(u)}{u} du.$$

2. *if* $a>0$ *and* $0<\liminf_{x\to 0+}\dfrac{x\overline{\Pi}(x)}{\int_0^x\overline{\Pi}(u)du}\le\limsup_{x\to 0+}\dfrac{x\overline{\Pi}(x)}{\int_0^x\overline{\Pi}(u)du}<1$,

$$\frac{k(t)}{\mathbf{P}(I_{\mathbf{e}_q}(\xi)>t)}\sim_{t\to 1/a} a\boldsymbol{\varphi}_{\Pi,\boldsymbol{q}}\left(\frac{t}{1-at}\right).$$

3. *if* $a>0$ *and* $\Pi(0,\infty)<\infty$, $\left(\dfrac{1}{a}-t\right)\dfrac{k(t)}{\mathbf{P}(I_{\mathbf{e}_q}(\xi)>t)}\sim_{t\to 1/a}\dfrac{\Pi(0,\infty)+q}{a}$.

In the first two cases the Von-mises condition is satisfied

$$k(t)\int_t^\infty \mathbf{P}(I_{\mathbf{e}_q}(\xi)>s)ds/(\mathbf{P}(I_{\mathbf{e}_q}(\xi)>t))^2\xrightarrow[t\to t_F]{}1.$$

To describe the behaviour of the distribution at 0 of the negative of a subordinator we introduce the following assumption.

(A) *The Lévy measure* Π *belongs to the class* $\mathcal{L}_\alpha$ *for some* $\alpha\ge 0$, *that is to say that the tail Lévy measure* $\overline{\Pi}$ *satisfies*

$$\lim_{x\to\infty}\frac{\overline{\Pi}(x+y)}{\overline{\Pi}(x)}=e^{-\alpha y},\quad \textit{for all } y\in\mathbb{R}. \tag{2.40}$$

Observe that regularly varying and subexponential tail Lévy measures satisfy this assumption with $\alpha=0$ and that convolution equivalent Lévy measures are examples of Lévy measures satisfying (2.40) for some index $\alpha>0$.

THEOREM (2.41). *Let* $q\ge 0$ *and* $\xi=-\sigma$, *where* σ *is a subordinator such that when* $q=0$ *the Lévy measure* Π *satisfies assumption* **(A)**. *The following asymptotic behaviour holds for the density function* k *of the exponential functional* $I_{\mathbf{e}_q}$.

i) If $q>0$, *then*

$$k(x)\to q\qquad as\ x\downarrow 0.$$

ii) If $q=0$, *then* $\mathbf{E}[I^{-\alpha}]<\infty$ *and*

$$k(x)\sim\mathbf{E}\left[I^{-\alpha}\right]\overline{\Pi}(\log 1/x)\quad as\ x\downarrow 0.$$

From this result it is possible to derive estimates for the exponential functional of the negative of a type of spectrally negative Lévy process, see [50].

Furthermore, the tail behaviour $\mathbb{P}(I\ge t)$ has been studied in a general setting, see for instance [19, 45, 54, 56], which is far from being an exhaustive list of references. We quote the following result from [54, 56].

THEOREM (2.42). (i) *Assume that* ξ_1 *is no-lattice and it satisfies the conditions*

$$\mathbf{E}\left[\exp(\gamma\xi_1)\right]=1\quad and\quad \mathbf{E}\left[\xi_1^+\exp(\gamma\xi_1)\right]<\infty.$$

In this case we have that $t^\gamma\mathbf{P}(I>t)\xrightarrow[t\to\infty]{}C_\gamma\in(0,\infty)$.

(ii) *Assume* $q=0$, *that there exists a* $\gamma>0$ *s.t.* ξ *is convolution equivalent with index* γ,

$$\lim_{t\to\infty}\frac{\mathbf{P}(\xi_1>t+s)}{\mathbf{P}(\xi_1>t)}=e^{-\gamma s},\ s\in\mathbb{R},\quad \lim_{t\to\infty}\frac{\mathbf{P}(\xi_2>t)}{\mathbf{P}(\xi_1>t)}=2\mathbf{E}[e^{\gamma\xi_1}],$$

and $\mathbf{E}[e^{\gamma\xi_1}]<1$. *If* $0<\gamma\le 1$, *we assume furthermore that* $\mathbf{E}[\xi_1]\in(-\infty,0)$. *Under these hypotheses*

$$\mathbf{P}(I>t)\sim c_\gamma\mathbf{E}(I^{|1-\gamma|})\Pi_\xi(\log(t),\infty)=t^{-\gamma}\ell(t),\qquad t\to\infty,$$

with ℓ an slowly varying function.

More results about exponential functionals and their relations with other areas of probability theory can be found in the thorough review by Bertoin and Yor [9].

(2.3) Asymptotic behaviour. The asymptotic behaviour of positive self-similar Markov processes $X^{(x)}$ with initial state $x > 0$ was studied by Lamperti (Theorem 7.1 in [44]). This property is inherited by the asymptotic behaviour of its associated Lévy process ξ and the fact that,

$$\lim_{t\to 0} \frac{\tau(t)}{t} = 1 \qquad \mathbb{P}_x - a.s.$$

Particularly, we have the following result due to Lamperti [44].

THEOREM (2.43). *Let ξ a Lévy process that admits a law of the iterated logarithm, i.e. for some function $g : [0,+\infty) \to [0,+\infty)$ and some constant $c \in \mathbb{R}$*

$$\liminf_{t\to 0} \frac{\xi_t}{g(t)} = c \quad or \quad \limsup_{t\to 0} \frac{\xi_t}{g(t)} = c, \quad almost\ surely.$$

Then for $x > 0$, $X^{(x)}$, its associated self-similar Markov process, satisfies

$$\liminf_{t\to 0} \frac{X_t^{(x)} - x}{g(t)} = C(x,c) \quad or \quad \limsup_{t\to 0} \frac{X_t^{(x)} - x}{g(t)} = C(x,c), \quad almost\ surely,$$

where $C(x,c)$ is a constant that only depends on x and c.

Of course, we would like to know if we can use the Lamperti representation to the study of the asymptotic behaviour of $X^{(x)}$ at $+\infty$. Also, we would like to know if we can study the lower and upper functions of positive self-similar Markov processes starting from 0 at small times.

Several partial results on the lower envelope of $X^{(0)}$ have already been established before, the oldest of which are due to Dvoretsky and Erdös [25] and Motoo [46] who studied the special case of Bessel processes. More precisely, when $X^{(0)}$ is a Bessel process with dimension $\delta > 2$, we have the following integral test at 0: if f is an increasing function then

$$\mathbb{P}(X_t^{(0)} < f(t), \text{i.o., as } t \to 0) = \begin{cases} 0 \\ 1 \end{cases} \quad \text{according as} \quad \int_{0+} \left(\frac{f(t)}{t}\right)^{\frac{\delta-2}{4}} \frac{dt}{t} \begin{cases} <\infty \\ =\infty \end{cases}.$$

The time inversion property of Bessel processes, induces the same integral test for the behaviour at $+\infty$ of $X^{(x)}$, $x \geq 0$.

Rivero [53] studied the lower functions of increasing self-similar Markov processes via the Lamperti representation. Following the method of Motoo [46] applied to $(e^{-t}X^{(x)}_{e^t-1}, t \geq 0)$, the Ornstein-Uhlenbeck process associated to $X^{(x)}$ (see Carmona et al. [15] for a proper definition), and under the assumption that the density ρ is decreasing in a neighborhood of $+\infty$ and bounded, Rivero [53] gave the following integral test for the lower envelope at $+\infty$.

THEOREM (2.44). *Let $x > 0$ and $X^{(x)}$ an increasing self-similar Markov processes starting from x. If h is a decreasing function then*

$$\mathbb{P}\left(X_s^{(x)} < s^{1/\alpha} h(s),\ i.o.,\ as\ s \to +\infty\right) = 0\ or\ 1$$

according as

$$\int^{\infty} \rho\big(1/h(s)\big)\frac{\mathrm{d}s}{s} \qquad \textit{is finite or infinite.}$$

A similar integral test for the lower envelope at 0 is obtained by Rivero via some reversal properties of $X^{(x)}$. From estimates of ρ and from the above result, Rivero [53] deduced the following laws of the iterated logarithm.

THEOREM (2.45). *Let $\alpha > 0$, and ξ be a subordinator whose Laplace exponent ϕ is regularly varying at $+\infty$ with index $\beta \in (0,1)$. Suppose that the density ρ, of the Lévy exponential functional $I(-\alpha\xi)$ of ξ satisfies that is decreasing in a neighborhood of $+\infty$, and bounded. For $x \geq 0$, let $X^{(x)}$ be the increasing positive self-similar Markov process associated to ξ with scaling index $1/\alpha$. Define*

$$f(t) = \frac{\phi(\log|\log t|)}{\log|\log t|}, \qquad t \neq e, \quad t > 0,$$

then for any $x \geq 0$

$$\liminf_{t\to+\infty} \frac{X^{(x)}}{(tf(t))^{1/\alpha}} = \alpha^{\beta/\alpha}(1-\beta)^{(1-\beta)/\alpha} \qquad \textit{almost surely},$$

and

$$\liminf_{t\to 0} \frac{X^{(0)}}{(tf(t))^{1/\alpha}} = \alpha^{\beta/\alpha}(1-\beta)^{(1-\beta)/\alpha} \qquad \textit{almost surely.}$$

This result extends the laws of the iterated logarithm of Friested [28] and Xiao [61].

We now present some general results on the lower envelope of $X^{(0)}$ at 0 and at ∞. The next result obtained by Chaumont and Pardo [19] means in particular that the asymptotic behaviour of $X^{(0)}$ only depends on the tail behaviour of the law of $I(-\xi)$, and on this of the law of

$$I_q(-\xi) \overset{(\mathrm{def})}{=} \int_0^{\hat{T}_{-q}} \exp\Big\{-\xi_s\Big\}\mathrm{d}s,$$

with $\hat{T}_x = \inf\{t : -\xi_t \leq x\}$, for $x \leq 0$. So also we set

$$F(t) \overset{(\mathrm{def})}{=} \mathbf{P}(I(-\xi) > t) \qquad \text{and} \qquad F_q(t) \overset{(\mathrm{def})}{=} \mathbf{P}(I_q(-\xi) > t).$$

THEOREM (2.46). *The lower envelope of $X^{(0)}$ at 0 is described as follows. Let f be an increasing function.*

(*i*) *If*

$$\int_{0+} F\left(\frac{t}{f(t)}\right)\frac{dt}{t} < \infty,$$

then for all $\varepsilon > 0$,

$$\mathbb{P}(X_t^{(0)} < (1-\varepsilon)f(t), i.o.,\ as\ t \to 0) = 0.$$

(*ii*) *If for all $q > 0$,*

$$\int_{0+} F_q\left(\frac{t}{f(t)}\right)\frac{dt}{t} = \infty,$$

then for all $\varepsilon > 0$,

$$\mathbb{P}(X_t^{(0)} < (1+\varepsilon)f(t), i.o.,\ as\ t \to 0) = 1.$$

(*iii*) *Suppose that* $t \mapsto f(t)/t$ *is increasing. If there exists* $\gamma > 1$ *such that,*

$$\limsup_{t\to+\infty} \mathbf{P}(I > \gamma t)/\mathbf{P}(I > t) < 1 \text{ and if } \int_{0+} F\left(\frac{t}{f(t)}\right)\frac{dt}{t} = \infty,$$

then for all $\varepsilon > 0$,

$$\mathbb{P}(X_t^{(0)} < (1+\varepsilon)f(t), i.o., \text{ as } t \to 0) = 1.$$

As can be expected, there is a version of the last result for large times but Chaumont and Pardo proved that it could be extended also for $X^{(x)}$, for $x > 0$.

We now consider two types of behaviour of $F(t)$. The first type of tail behaviour that we consider is the case where F is regularly varying at infinity, i.e.

$$F(t) \sim \lambda t^{-\gamma} L(t), \quad t \to +\infty, \tag{2.47}$$

where $\gamma > 0$ and L is a slowly varying function at $+\infty$. It is not difficult to see that, under this assumption, for any $q > 0$ the functions F_q and F are equivalent, i.e. $F_q \asymp F$. More precisely, if (2.47) holds then for all $q > 0$,

$$(1 - e^{-\gamma q})F(t) \le F_q(t) \le F(t), \tag{2.48}$$

for all t large enough. This last inequality is consequence from (2.47) and the independence of the processes $(\xi_s, 0 \le s \le \hat{T}_{-q})$ and $(\xi_{s+\hat{T}_{-q}} - \xi_{\hat{T}_{-q}}, s \ge 0)$. It is important to note that the Dvoretzky and Erdös integral test is consequence of this result when $X^{(0)}$ is a transient Bessel process. Another example of such behaviour is when the process ξ satisfies the hypotheses of Theorem (2.42).

The second type of behaviour that we shall consider is when $\log F$ is regularly varying at $+\infty$, i.e.

$$-\log F(t) \sim \lambda t^{\beta} L(t), \quad \text{as } t \to \infty, \tag{2.49}$$

where $\lambda > 0$, $\beta > 0$ and L is a function which varies slowly at $+\infty$. Under this assumption, the conditions of part (iii) of the general integral tests due to Chaumont and Pardo are satisfied.

Define the function ψ by

$$\psi(t) \overset{(\text{def})}{=} \frac{t}{\inf\{s : 1/F(s) > |\log t|\}}, \quad t > 0, t \neq 1. \tag{2.50}$$

Chaumont and Pardo found that the lower envelope of $X^{(0)}$ satisfies the following law of the iterated logarithm:

(*i*)

$$\liminf_{t\to 0} \frac{X_t^{(0)}}{\psi(t)} = 1, \text{ almost surely.} \tag{2.51}$$

(*ii*) For all $x \ge 0$,

$$\liminf_{t\to+\infty} \frac{X_t^{(x)}}{\psi(t)} = 1, \text{ almost surely.} \tag{2.52}$$

Note that this result extends the laws of the iterated logarithm found by Xiao [61] and Rivero [53] in the increasing case, but not their integral tests. Also, note that the assumption that the density of $I(-\xi)$ is decreasing in a neigbourhood of $+\infty$ in the law of the iterated logarithm due to Rivero [53] is not necessary but it is really important for his integral tests. In what follows we will describe some integral tests obtained in [47].

Let us define

$$\bar{G}(t) \stackrel{\text{(def)}}{=} \mathbf{P}\Big(S_1 < t\Big) \quad \text{and} \quad \bar{F}(t) \stackrel{\text{(def)}}{=} \mathbf{P}\Big(I(-\xi) < t\Big),$$

where S_1 denotes the first passage time of $X^{(0)}$ above the level 1. We also denote by $\mathcal{H}_0$ the totality of positive increasing functions $h(t)$ on $(0,\infty)$ that satisfy

i) $h(0) = 0$, and

ii) there exists $\beta \in (0,1)$ such that $\sup_{t<\beta} \dfrac{t}{h(t)} < \infty$.

The following result is extracted from [47].

THEOREM (2.53). *Let $h \in \mathcal{H}_0$.*

i) If

$$\int_{0^+} \bar{G}\left(\frac{t}{h(t)}\right) \frac{\mathrm{d}t}{t} < \infty,$$

then for all $\epsilon > 0$

$$\mathbb{P}_0\Big(X_t^{(0)} > (1+\epsilon)h(t), \text{ i.o., as } t \to 0\Big) = 0.$$

ii) If

$$\int_{0^+} \bar{F}\left(\frac{t}{h(t)}\right) \frac{\mathrm{d}t}{t} = \infty,$$

then for all $\epsilon > 0$

$$\mathbb{P}_0\Big(X_t > (1-\epsilon)h(t), \text{ i.o., as } t \to 0\Big) = 1.$$

A similar integral test holds for large times. As in the case for the lower envelope of $X^{(0)}$, these results can be applied to two type of estimates of the tail behaviour of I and vI. First, we discuss the case when $\bar{F}$ and $\bar{G}$ satisfy

$$ct^{\alpha}L(t) \le \bar{F}(t) \le \bar{G}(t) \le Ct^{\alpha}L(t) \qquad \text{as} \quad t \to 0, \tag{2.54}$$

where $\alpha > 0$, c and C are two positive constants such that $c \le C$ and L is a slowly varying function at 0. An important example included in this case is when $\bar{F}$ and $\bar{G}$ are regularly varying functions at 0. The "regularity" of the behaviour of $\bar{F}$ and $\bar{G}$ gives the following integral tests obtained in [47].

THEOREM (2.55) (Regular case). *Under condition (2.54), the upper envelope of $X^{(0)}$ at 0 and at $+\infty$ is as follows. Let $h \in \mathcal{H}_0$, such that either $\lim_{t\to 0} t/h(t) = 0$ or $\liminf_{t\to 0} t/h(t) > 0$, then*

$$\mathbb{P}\Big(X_t^{(0)} > h(t), \text{ i.o., as } t \to 0\Big) = 0 \text{ or } 1,$$

according as

$$\int_{0^+} \bar{F}\left(\frac{t}{h(t)}\right) \frac{\mathrm{d}t}{t} < \infty \qquad \textit{is finite or infinite.}$$

Note that under condition (2.54), we may drop the factor $(1+\epsilon)$ and that the previous integral test depends only of $\bar{F}$. This result extends the integral test of Khintchine [34] for a stable subordinator. This is consequence of the following estimate of $\bar{F}$,

$$\bar{F}(t) \sim kt^{\beta+1} \qquad \text{as} \qquad t \to 0,$$

and since in the increasing case this integral test determines the upper envelope of increasing pssMp.

The second type of behaviour that we shall consider is when $\log \bar{F}$ and $\log \bar{G}$ are regularly varying at 0, i.e

$$-\log \bar{G}(1/t) \sim -\log \bar{F}(1/t) \sim \lambda t^{\beta} L(t), \quad \text{as } t \to +\infty, \tag{2.56}$$

where $\lambda > 0$, $\beta > 0$ and L is a slowly varying function at $+\infty$. Under this assumption, the upper envelope of $X^{(0)}$ may be described as follows. Define the function

$$\phi(t) \overset{(\text{def})}{=} t \inf\{s : 1/\bar{F}(1/s) > |\log t|\}, \quad t > 0, \quad t \neq 1.$$

The following result has been obtained in [47].

THEOREM (2.57) (Log-regular case). *Under condition (2.56), the future infimum process satisfies the following law of the iterated logarithm:*

i)

$$\limsup_{t \to 0} \frac{X_t^{(0)}}{\phi(t)} = 1, \qquad \textit{almost surely.}$$

ii) For all $x \geq 0$,

$$\limsup_{t \to +\infty} \frac{X_t^{(x)}}{\phi(t)} = 1, \qquad \textit{almost surely.}$$

It can be seen using the results in Theorem (2.37) that there is a large family that satisfies the condition (2.56). From this estimate Pardo obtained the following law of iterated logarithm for the future infimum process in terms of the following function.
Let us define the function

$$f(t) = \frac{\psi(\log|\log t|)}{\log|\log t|} \qquad \text{for} \qquad t > 1, \quad t \neq e,$$

with ψ the Laplace exponent of $-\xi$, with ξ a spectrally negative Lévy process. By integration by parts, we can see that the function $\psi(\lambda)/\lambda$ is increasing, hence it is straightforward that the function $tf(t)$ is also increasing in a neighbourhood of ∞. Using this in [47] it has been proved that if ψ is regularly varying at $+\infty$ with index $\beta \in (1,2)$, then

$$\limsup_{t \to 0} \frac{X_t^{(0)}}{\big(tf(t)\big)^{1/\alpha}} = \alpha^{\beta/\alpha}(\beta-1)^{-(\beta-1)/\alpha} \qquad \text{almost surely,} \tag{2.58}$$

and,

$$\limsup_{t \to +\infty} \frac{X_t^{(x)}}{\big(tf(t)\big)^{1/\alpha}} = \alpha^{\beta/\alpha}(\beta-1)^{-(\beta-1)/\alpha} \qquad \text{almost surely.} \tag{2.59}$$

We finish this section with the following interpretation of the result of existence of a limit measure $\mathbb{P}_0$. If $(\xi, \mathbf{P})$ is a subordinator, associated to a pssMp $(X, \mathbb{P})$ via Lamperti's transformation, and has finite mean $m := \mathbf{E}(\xi_1) < \infty$, we know that there exists a measure $\mathbb{P}_{0+}$ such that

$$\mathbb{P}_x(X_1 \in dy) \xrightarrow[x \to 0+]{\text{weakly}} \mathbb{P}_{0+}(X_1 \in dx) = m^{-1} y^{\alpha} \mathbf{P}\left(I^{-1/\alpha} \in dy\right),$$

where $I = \int_0^{\infty} e^{-\alpha \xi_s} ds$. It is furthermore known that if $\mathbf{E}(\xi_1) = \infty$, then

$$\mathbb{P}_x(X_1 \in dy) \xrightarrow[x \to 0+]{\text{weakly}} \delta_{\infty}(dy).$$

Due to the self-similarity

$$\mathbb{P}_x(X_1 \in dy) = \mathbb{P}_1(xX_{1/x^\alpha} \in dy),$$

hence the latter is equivalent to

$$\frac{X_t}{t^{1/\alpha}} \xrightarrow[t\to\infty]{\text{Law}} \begin{cases} Z, & \text{if } \mathbf{E}(\xi_1) < \infty, \ Z \text{ has the same law as } X_1 \text{ under } \mathbb{P}_{0+}; \\ \infty, & \text{if } \mathbf{E}(\xi_1) = \infty. \end{cases}$$

A further problem that has been addressed in [13] describes the rate at which $X_t/t^{1/\alpha}$ tends towards ∞, when the mean of the underlying subordinator ξ is not finite. The main result by Caballero and Rivero [13] is the following.

THEOREM (2.60). *Let $\{X(t), t \geq 0\}$ be a positive $1/\alpha$-self-similar Markov process with increasing paths. The following assertions are equivalent:*

(i) *$\phi : \mathbb{R}^+ \to \mathbb{R}^+$, is regularly varying at 0 with an index $\beta \in [0,1]$.*
(ii) *Under $\mathbb{P}_1$ the random variables $\{\log(X(t)/t^{1/\alpha})/\log(t), t > 1\}$ converge weakly as $t \to \infty$ towards a r.v. V.*
(iii) *For any $x > 0$, under $\mathbb{P}_x$ the random variables $\{\log(X(t)/t^{1/\alpha})/\log(t), t > 1\}$ converge weakly as $t \to \infty$ towards a r.v. V.*

In this case, the law of V is given by: $V = 0$ a.s. if $\beta = 1$; $V = \infty$, a.s. if $\beta = 0$, and if $\beta \in]0,1[$, its law has a density given by

$$\frac{\alpha^{1-\beta} 2^\beta \sin(\beta\pi)}{\pi} v^{-\beta}(2 + \alpha v)^{-1} \mathrm{d}v, \qquad v > 0.$$

(2.4) Quasi-stationary distributions. Another topic that has been studied related with the asymptotic behaviour of pssMp is the existence of Yaglom limits and quasi-stationary distributions. The main difference with the asymptotic behaviour described before relies on the idea that pssMp that hit zero at a finite time may be at an equilibrium state before being absorbed at 0. The main questions that were addressed in the paper [30] are the following. Assuming that the self-similar Markov process hits zero in a finite time with probability one, determine

QS-I whether there exists a probability measure μ on $\mathbb{R}^+$ such that for any $t > 0$,

$$\int_{\mathbb{R}^+} \mu(dx)\mathbb{P}_x(X_t \in dy | t < T_0) = \mu(dy), \qquad y \geq 0,$$

i.e. μ is a *quasi-stationary measure* for the pssMp $(X, \mathbb{P})$;

QS-II whether there exists a function $g : \mathbb{R}^+ \to \mathbb{R}^+ \setminus \{0\}$ and a non-degenerate probability measure ν on $\mathbb{R}^+$ such that

$$\mathbb{P}_1\left(\frac{X_t}{g(t)} \in dy \middle| t < T_0\right) \xrightarrow[t\to t_F]{\text{weakly}} \nu(dy),$$

we will say that *ν is the limit in the Yaglom sense of X normalized by g*; $t_F = \sup\{t > 0 : \mathbb{P}_1(T_0 > t) > 0\}$.

QS-III what is the relation between μ and ν.

These questions were studied by Haas in [29] under the assumption that X has non-increasing paths and that the Lévy measure of the associated Lévy process is

regularly varying at 0. In [30] the general case has been studied. We will assume wlog that the self-similarity index is equal to 1. In other case, the process

$$Y_t = X_t^{\alpha}, t \geq 0,$$

is a 1-pssMp. A QS-law for X exists iff a QS-law for Y exists. Analogously for Yaglom limits.

The problem of existence of quasi-stationary distributions was tangentially studied by Bertoin and Yor [7] in the case where the process has non-increasing paths. A slight modification of their main results read as follows.

THEOREM (2.61). *If X has non-increasing paths, that is $-\xi$ is a subordinator, then*

- *there exists a QS-law for X,*

$$\int_{\mathbb{R}^+} \mu(dx)\mathbb{P}_x(X_t \in dy | t < T_0) = \mu(dy), \qquad y \geq 0.$$

- *μ is characterized by its entire moments; there is a $\beta > 0$*

$$\int_{\mathbb{R}^+} x^n \mu(dx) = \beta^{-n} \prod_{i=1}^{n} \phi(i), \qquad n \geq 1,$$

  with $\phi(\lambda) = -\log \mathbf{E}(e^{\lambda \xi_1})$, $\lambda > 0$,
- *let R follow the law μ and assume it is independent of ξ, then*

$$R \times \int_0^{\infty} e^{\xi_s} ds \sim Exponential(\beta).$$

So, in order to answer the question (QS-I) one should deal with the case where the paths are allowed to increase. In Theorem 1.1 in [30] it is proved that a necessary and sufficient condition for the existence of a QS-law for a pssMp is that the process has non-increasing paths. The main argument in proving the necessity of this condition is the following. Assume μ is a QS-law for X, the simple Markov property implies that there exists an index $\theta > 0$ such that

$$\int_{(0,\infty)} \mu(\mathrm{d}x)\mathbb{P}_x(t < T_0) = e^{-\theta t}, \qquad t \geq 0.$$

The self-similarity of X implies

$$e^{-\theta t} = \int_{(0,\infty)} \mu(\mathrm{d}x)\mathbb{P}_x(t < T_0) = \int_{[0,\infty)} \mu(\mathrm{d}x)\mathbb{P}_1(t < xT_0).$$

Recall then that $(T_0, \mathbb{P}_1) \overset{\text{Law}}{=} (I, \mathbf{P})$, where $I = \int_0^{\zeta} e^{\xi_s} \mathrm{d}s$. Then if R $\sim \mu$ and R is independent of I, we have that

$$\mathrm{R}I \overset{\text{Law}}{=} \mathbf{e}/\theta.$$

This identity and the independence imply that I has moments of all positive orders. An easy argument allows to prove that the totality of Lévy processes for which the exponential functional I has moments of all positive orders are exactly non-increasing Lévy processes. This concludes the argument because Lamperti's transform preserves the path of the process.

A by product of the above discussion is that the existence of QS-laws is closely related to factorizations of he exponential r.v. as the product of two independent r.v., one of which is an exponential functional. As we will see below this extends

to Yaglom limits and give also rise to factorizations of Pareto and Beta r.v. Before giving a precise result we state a key observation that allows to transform the problem of existence of Yaglom limits into a problem of maximum domain of attraction for exponential functionals.

LEMMA (2.62). *Let X be a pssMp that hits 0 in a finite time and ξ the Lévy process associated to it via Lamperti's transformation, which necessarily drift towards $-\infty$ or has a finite lifetime. We denote $I := \int_0^\infty e^{\xi_s} ds$. For $t > 0$, we have the equality of measures*

$$\mathbf{P}(I - t \in dy | t < I) = \mathbb{P}_1(X_t \widetilde{I} \in dy | t < T_0),$$

where $\widetilde{I}$ has the same law as I and is independent of $(X_s, s \le t)$.

This result is an straithforward consequence of Lamperti's transformation. With this result at hand we have the following answer to question (QS-II) about Yaglom limits.

THEOREM (2.63). *Let X be a pssMp that hits 0 in a finite time. The following assertions are equivalent.*

1. *The process $(X_t, t \ge 0)$ admits a Yaglom limit.*
2. *The process $(X_t \widetilde{I}, t \ge 0)$ admits a Yaglom limit, with $\widetilde{I} \overset{Law}{=} I$ and independent of X.*
3. *There exists a function $g : \mathbb{R}^+ \to \mathbb{R}^+ \setminus \{0\}$ and a non-degenerate probability measure $\widetilde{\Lambda}$ on $\mathbb{R}^+$ s.t.*
$$\mathbf{P}\left(\frac{I - t}{g(t)} \in dy \middle| t < I \right) \xrightarrow[t \to \infty]{weakly} \widetilde{\Lambda}(dy).$$
4. *I is in the maximum domain of attraction of a Gumbel, Weibul or Fréchet distribution.*

In this case,

- *if $I \in MDA(Gumbel)$, $\widetilde{\Lambda}(dy) = e^{-y} dy$, $y \ge 0$;*
- *if $I \in MDA(Weibul)$, $\exists \gamma > 0$ s.t. $\widetilde{\Lambda}(dy) = \gamma (1-y)^{\gamma - 1} dy$, $y \in (0,1)$;*
- *if $I \in MDA(Frechet)$, $\exists \gamma > 0$ s.t. $\widetilde{\Lambda}(dy) = \gamma (1+y)^{-\gamma - 1} dy$, $y \ge 0$.*

As we mentioned before the problem of existence of Yaglom limits is also related to factorisations of r.v. The Lemma (2.62) and the latter theorem have the following consequence: if X admits a Yaglom limit, then there exists a non trivial independent random variable

$$R \overset{\text{Law}}{=} \text{weak-}\lim_{t \to \infty} \frac{X_t}{g(t)} | t < T_0,$$

s.t.

$$R \times I \sim \begin{cases} \text{Exponential}(\beta), & \text{for some } \beta > 0, \\ \text{Beta}(1, \gamma), & \text{for some } \gamma > 0, \\ \text{Pareto}(\gamma), & \text{for some } \gamma > 0. \end{cases}$$

A systematic study of this type of factorisations and some consequences is carried out in [31].

We will next quote two of the main results from [30] providing some necessary and sufficient conditions for I to be in a maximum domain of attraction of a

Gumbel or Frechet distribution. Further cases are treated in [30], together with precise results about descriptions of the Weibull case.

Assume X has non-increasing paths and let ξ the underlying Lévy process. It is known that

$$-\log \mathbf{E}(e^{\lambda \xi_1}) = q + a\lambda + \int_0^\infty 1 - e^{-\lambda x}\Pi(dx), \qquad \lambda > 0,$$

with Π a measure on $(0,\infty)$ such that $\int_{(0,\infty)} 1 \wedge x\Pi(dx) < \infty$, $q, a \geq 0$. We denote by $\boldsymbol{\varphi}_{\Pi,\boldsymbol{q}}$ the inverse function of the mapping

$$t \mapsto \frac{t}{\int_0^\infty (1 - e^{-tx})\Pi(\mathrm{d}x) + q}; \tag{2.64}$$

the inverse is well defined on $[0,\infty)$ if $q > 0$ and on $\left[\left(\int_0^\infty x\Pi(\mathrm{d}x)\right)^{-1}, \infty\right)$ in other case. We denote $\overline{\Pi}(x) := \Pi(x,\infty)$.

THEOREM (2.65). *Let X be a pssMp with non-increasing paths, and ξ the underlying Lévy process. Assume that $-\xi$ is a subordinator with killing term $q \geq 0$, drift $a = 0$ and Lévy measure Π such that*

$$\liminf_{x \to 0} \frac{x\overline{\Pi}(x)}{\int_0^x \overline{\Pi}(u)\mathrm{d}u} > 0. \tag{2.66}$$

In this case $I \in \mathrm{MDA}_{\mathrm{Gumbel}}$, $t_F = \infty$ and

$$\mathbb{P}_1\left(\frac{\boldsymbol{\varphi}_{\Pi,\boldsymbol{q}}(t)X_t}{t} \in \cdot \;\middle|\; t < T_0\right) \xrightarrow[t\to\infty]{} \mu_I^{(\mathbf{e})}.$$

$\mu_I^{(\mathbf{e})}$ is the unique probability measure such that if $R \sim \mu_I^{(\mathbf{e})}$, and $R \perp I$ then $RI \sim Exp(1)$.

Reciprocally, if $I \in \mathrm{MDA}_{\mathrm{Gumbel}}$ and $t_F = \infty$, then $-\xi$ is a subordinator with drift zero and $g(t) \sim \dfrac{\int_t^\infty \mathbf{P}(I > s)ds}{\mathbf{P}(I > t)} = \mathbb{E}_1(X_t | t < T_0)\mathbf{E}(I)$.

The proof of this result rely on precise estimates for the tail distribution of an exponential functional of a non-increasing Lévy processes, some of which are described in Theorem (2.39) here, and this allow to verify that the so-called Von-Mises condition is satisfied which is well known to be a necessary and sufficient condition for a r.v. to be in the maximum domain of attraction of a Gumbel distribution.

To deal with the non-monotone case, which happens to be the one corresponding to the domain of attraction of a Frechet distribution, the following result has been obtained in [30].

THEOREM (2.67). *The following are equivalent:*

- *X is a pssMp with non-monotone paths and that admits a Yaglom limit.*
- *$I \in \mathrm{MDA}_{\mathrm{Fréchet}}$.*
- *$t \mapsto \mathbb{P}(I > t)$ is regularly varying at infinity with some index $-\gamma$, $\gamma > 0$.*

In this case

$$\mathbb{P}_1\left(\frac{X_t}{t} \in \cdot \;\middle|\; t < T_0\right) \xrightarrow[t\to\infty]{} \mu_I^{(\mathbf{P}_\gamma)}. \tag{2.68}$$

The probability measure $\mu_I^{(\mathbf{P}_\gamma)}$ *is the unique p.m. such that if* $R \sim \mu_I^{(\mathbf{P}_\gamma)}$, *and* $R \perp I$ *then*

$$\mathbf{P}(RI \in dy) = \gamma(1+y)^{-\gamma}, y > 0.$$

A necessary condition for $I \in \mathrm{MDA}_{\text{Fréchet}}$ *is*

$$\mathbb{E}[e^{\theta\xi_1}] \le 1,\ \forall 0 \le \theta \le \gamma, \quad \textit{and } \mathbb{E}[e^{\delta\xi_1}] > 1, \forall \delta > \gamma,$$

for some $\gamma > 0$.

Sufficient conditions for the above theorem to hold were given in Theorem (2.42).

3. Real valued self-similar Markov processes

In previous sections, we studied positive self-similar Markov processes and their relationship with Lévy processes via the Lamperti representation. In this section we survey some recent results on real valued self-similar Markov processes which turn out to be associated to Markov additive processes via a Lamperti-type representation.

The structure of real valued self-similar Markov processes has been investigated by Chybiryakov [21] in the symmetric case, and by Kiu [35] and Chaumont et al. [18] in general. Inspired from [38], here we give an interpretation of those authors' results in terms of a two-state Markov additive process. We begin with some relevant definitions.

We focus on real-valued self-similar Markov processes (rssMp) $X = (X_t, t \ge 0)$ with self- similarity index $\alpha > 0$ and starting from $x \in \mathbb{R}\setminus\{0\}$. Let $(\mathbb{P}_x)_{x\in\mathbb{R}\setminus\{0\}}$ denote its probability laws starting from x.

In [18], the authors confine their attention to processes in "class **C.4**". A real valued self-similar Markov process X is in this class if,

$$\mathbb{P}_x(\exists t > 0 : X_t X_{t-} < 0) = 1, \qquad \forall x \neq 0.$$

By the strong Markov property this implies that, with probability one, the process X changes sign infinitely often. This assumption will be in force in the sequel.

The other cases introduced in [18], namely **C.1-3**, are those where either the process changes once, and only once, of sign, and those where the process never changes of sign. In the former cases the construction of real valued ssMp can be easily deduced using the ideas to deal with the cases where the process changes of sign infinitely many times. In the latter case, the real-ssMp is constructed by applying Lamperti's transformation to the process when started at a positive (negative) position using two Lévy processes, one to describe the positive values of the path and the other for the negative values. We do not provide further details about these cases.

(3.1) Markov additive processes. Let E be a finite state space and $(\mathcal{G}_t)_{t\ge0}$ a standard filtration. A càdlàg process (ξ, J) in $\mathbb{R}\times E$ with law $\mathbf{P}$ is called a Markov additive process (MAP) with respect to $(\mathcal{G}_t)_{t\ge0}$ if $(J(t))_{t\ge0}$ is a continuous-time Markov chain in E, and the following property is satisfied, for any $i \in E$, $s, t \ge 0$:

(3.1) given $\{J(t) = i\}$, the pair $(\xi(t+s) - \xi(t), J(t+s))$ is independent of $\mathcal{G}_t$, and has the same distribution as $(\xi(s) - \xi(0), J(s))$ given $\{J(0) = i\}$.

Aspects of the theory of Markov additive processes are covered in a number of texts, among them [1] and [2].

Let us introduce some notation. We write $\mathbf{P}_i = \mathbf{P}(\cdot|\xi(0)=0, J(0)=i)$; and if μ is a probability distribution on E, we write $\mathbf{P}_\mu = \mathbb{P}(\cdot|\xi(0)=0, J(0)\sim\mu) = \sum_{i\in E}\mu(i)\mathbf{P}_i$. We adopt a similar convention for expectations.

It is well-known that a Markov additive process (ξ, J) also satisfies (3.1) with t replaced by a finite stopping time. Furthermore, it has the structure given by the following proposition; see [2, §XI.2a] and [33, Proposition 2.5].

PROPOSITION (3.2). *The pair (ξ, J) is a Markov additive process if and only if, for each $i, j \in E$, there exist a sequence of iid Lévy processes $(\xi_i^n)_{n\geq 0}$ and a sequence of iid random variables $(U_{ij}^n)_{n\geq 0}$, independent of the chain J, such that if $T_0 = 0$ and $(T_n)_{n\geq 1}$ are the jump times of J, the process ξ has the representation*

$$\xi(t) = \mathbf{1}_{\{n>0\}}(\xi(T_n-) + U^n_{J(T_n-),J(T_n)}) + \xi^n_{J(T_n)}(t-T_n), \text{for } t\in[T_n, T_{n+1}),\ n\geq 0.$$

For each $i \in E$, it will be convenient to define, on the same probability space, ξ_i as a Lévy process whose distribution is the common law of the ξ_i^n processes in the above representation; and similarly, for each $i, j \in E$, define U_{ij} to be a random variable having the common law of the U_{ij}^n variables.

Let us now fix the following setup. Firstly, we confine ourselves to irreducible Markov chains J. Let the state space E be the finite set $\{1,\dots,N\}$, for some $N \in \mathbb{N}$. Denote the transition rate matrix of the chain J by $Q = (q_{ij})_{i,j\in E}$. For each $i \in E$, the Laplace exponent of the Lévy process ξ_i will be written ψ_i, in the sense that $e^{\psi_i(z)} = \mathbf{E}[e^{z\xi_i(1)}]$, for all $z \in \mathbb{C}$ for which the right-hand side exists. For each pair of $i, j \in E$, define the Laplace transform $G_{ij}(z) = \mathbf{E}[e^{zU_{ij}}]$ of the jump distribution U_{ij}, where this exists; write $G(z)$ for the $N\times N$ matrix whose (i,j)–th element is $G_{ij}(z)$. We will adopt the convention that $U_{ij} = 0$ if $q_{ij} = 0$, $i \neq j$, and also set $U_{ii} = 0$ for each $i \in E$.

A multidimensional analogue of the Laplace exponent of a Lévy process is provided by the matrix-valued function

$$F(z) = \operatorname{diag}(\psi_1(z),\dots,\psi_N(z)) + Q\circ G(z), \tag{3.3}$$

for all $z \in \mathbb{C}$ where the elements on the right are defined, where $\circ$ indicates elementwise multiplication, also called Hadamard multiplication. It is then known that

$$\mathbf{E}_i[e^{z\xi(t)}; J(t) = j] = \left(e^{F(z)t}\right)_{ij}, \text{ for } i, j \in E,$$

for all $z \in \mathbb{C}$ where one side of the equality is defined. For this reason, F is called the *matrix exponent* of the MAP ξ.

(3.2) Lamperti type representation of real valued self similar Markov processes. Let X be a real valued self-similar Markov process. In [18] it has been proved that X may be identified up to the first hitting time of 0,

$$T_0 = \inf\{t \geq 0 : X_{t-} = 0 \text{ or } X_t = 0\},$$

as the time-changed exponential of a certain complex-valued process $\mathcal{E}$, which from the terminology used in [18] it will be called the *Lamperti-Kiu representation* of X. The main result in [18] is summarised in the following theorem.

THEOREM (3.4). *Let X be a rssMp in class **C.4**, and let $x \neq 0$. It is possible to define independent sequences $(\xi^{\pm,k})_{k\geq 0}, (\zeta^{\pm,k})_{k\geq 0}, (U^{\pm,k})_{k\geq 0}$ of iid random objects with the following proprties:*

1. *The elements of these sequences are distributed such that: the $\xi^{\pm}$ are real-valued Lévy processes; $\zeta^{\pm}$ are exponential random variables with parameters $q^{\pm}$; and $U^{\pm}$ are real-valued random variables.*
2. *For each $x \neq 0$, define the following objects:*

$$(\xi^{(x,k)}, \zeta^{(x,k)}, U^{(x,k)}) = \begin{cases} (\xi^{+,k}, \zeta^{+,k}, U^{+,k}), & \text{if } \operatorname{sgn}(x)(-1)^k = 1 \\ (\xi^{-,k}, \zeta^{-,k}, U^{-,k}), & \text{if } \operatorname{sgn}(x)(-1)^k = -1, \end{cases}$$

$$\mathcal{T}_0^{(x)} = 0, \quad \mathcal{T}_n^{(x)} = \sum_{k=0}^{n-1} \zeta^{(x,k)},$$

$$N_t^{(x)} = \max\{n \geq 0 : \mathcal{T}_n^{(x)} \leq t\},$$

$$\sigma_t^{(x)} = t - \mathcal{T}_{N_t^{(x)}}^{(x)},$$

$$\mathcal{E}_t^{(x)} = \xi^{(N_t^{(x)})}_{\sigma_t^{(x)}} + \sum_{k=0}^{N_t-1} (\xi^{(x,k)}_{\zeta^{(x,k)}} + U^{(x,k)}) + i\pi N_t^{(x)}, \qquad t \geq 0,$$

$$\tau(t) = \inf\Big\{s > 0 : \int_0^s |\exp(\alpha \mathcal{E}_u^{(x)})|\, \mathrm{d}u > t|x|^{-\alpha}\Big\}, \quad t < T_0.$$

Then, the process X under the measure $\mathbb{P}_x$ has the representation

$$X_t = x \exp(\mathcal{E}^{(x)}_{\tau(t)}), \qquad 0 \leq t < T_0.$$

3. *Reciprocally, any process constructed in this form is a real-valued ssMp.*

The case where X is a stable process killed at its first hitting time of 0 or conditioned to avoid zero is studied in detail in [18].

The abundance of notation necessary to be precise in this context may obscure the fundamental idea, which is as follows. At any given time, the process $\mathcal{E}$ evolves as a Lévy process $\xi^{\pm}$, moving along a line $\Im(z) = \pi N$, up until an exponential 'clock' $\zeta^{\pm}$ (corresponding to the process X changing sign) rings. At this point the imaginary part of $\mathcal{E}$ is incremented by π, the real part jumps by $U^{\pm}$, and the process begins to evolve as the other Lévy process, $\xi^{\mp}$.

Particularly in light of the discussion in the previous section, the latter result can be formulated using Markov additive processes. This is the purpose of the following result proved in [38].

PROPOSITION (3.5). *Let X be an rssMp, with Lamperti–Kiu representation $\mathcal{E}$. Define furthermore*

$$[n] = \begin{cases} 1, & \text{if } n \text{ is odd}, \\ 2, & \text{if } n \text{ is even}. \end{cases}$$

Then, for each $x \neq 0$, the process

$$(\xi(t), J(t)) = \big(\Re(\mathcal{E}_t^{(x)}), \big[\Im(\mathcal{E}^{(x)})_t/\pi + 1_{\{x>0\}}\big]\big)$$

defined in Proposition 3 is a Markov additive process with state space $E = \{1, 2\}$, and X under $\mathbb{P}_x$ has the representation

$$X_t = x \exp\big(\xi(\tau(t)) + i\pi(J(\tau(t)) + 1)\big), \text{for } \ 0 \leq t < T_0,$$

where we note that $(\xi(0), J(0))$ *is equal to* $(0,1)$ *if* $x > 0$, *or* $(0,2)$ *if* $x < 0$. *Furthermore, the time-change* τ *has the representation*

$$\tau(t) = \inf\Big\{s > 0 : \int_0^s \exp(\alpha\xi(u))\,\mathrm{d}u > t|x|^{-\alpha}\Big\}, \qquad \text{for } t < T_0, \tag{3.6}$$

in terms of the real-valued process ξ.

Note that the MAP (ξ, J) under $\mathbf{P}_1$ corresponds to the rssMp X started at a point $x > 0$, and the MAP under $\mathbf{P}_2$ corresponds to the rssMp started at a point $x < 0$.

Furthermore, we observe from the form (3.6) of the time-change τ that under $\mathbb{P}_x$, for any $x \neq 0$, the following identity holds for T_0, the hitting time of zero:

$$|x|^{-\alpha} T_0 \overset{\text{Law}}{=} \int_0^\infty e^{\alpha\xi(u)}\,\mathrm{d}u.$$

Implicit in this statement is that the MAP on the right-hand side has law $\mathbf{P}_1$ if $x > 0$, and law $\mathbf{P}_2$ if $x < 0$.

(3.3) Exponential functionals of MAPs. We start by describing the existence of the *leading eigenvalue* of the matrix F, which will play a key role in our analysis of MAPs. This is sometimes also called the *Perron-Frobenius eigenvalue*; see [2, §XI.2c] and [33, Proposition 2.12].

PROPOSITION (3.7). *Suppose that* $z \in \mathbb{C}$ *is such that* $F(z)$ *is defined. Then, the matrix* $F(z)$ *has a real simple eigenvalue* $\kappa(z)$, *which is larger than the real part of all its other eigenvalues. Furthermore, the corresponding right-eigenvector* $v(z)$ *may be chosen so that* $v_i(z) > 0$ *for every* $i \in E$, *and normalised such that*

$$\pi v(z) = 1 \tag{3.8}$$

where π *is the equilibrium distribution of the chain* J.

This leading eigenvalue features in the following probabilistic result, which identifies a martingale (also known as the Wald martingale) and change of measure analogous to the exponential martingale and Esscher transformation of a Lévy process; cf. [2, Proposition XI.2.4, Theorem XIII.8.1].

PROPOSITION (3.9). *Let*

$$M(t,\gamma) = e^{\gamma\xi(t) - \kappa(\gamma)t} \frac{v_{J(t)}(\gamma)}{v_{J(0)}(\gamma)}, \text{ for } t \geq 0,$$

for some γ *such that the right-hand side is defined. Then,*

i) M$(\cdot,\gamma)$ *is a unit-mean martingale with respect to* $(\mathcal{G}_t)_{t\geq 0}$ *under any initial distribution of* $(\xi(0), J(0))$.

ii) Define the change of measure

$$\left.\frac{\mathrm{d}\mathbf{P}^{(\gamma)}}{\mathrm{d}\mathbf{P}}\right|_{\mathcal{G}_t} = M(t,\gamma).$$

Under $\mathbf{P}^{(\gamma)}$, *the process* ξ *is still a Markov additive process, and it has the following characteristics, for each* $i, j \in E$:

- $\mathbf{P}^{(\gamma)}(U_{ij} \in \mathrm{d}x) = \dfrac{e^{\gamma x}}{G_{ij}(\gamma)}\mathbf{P}(U_{ij} \in \mathrm{d}x)$, *and hence* $G_{ij}^{(\gamma)}(z) = \dfrac{G_{ij}(z+\gamma)}{G_{ij}(\gamma)}$,
- $q_{ij}^{(\gamma)} = \dfrac{v_j(\gamma)}{v_i(\gamma)} q_{ij} G_{ij}(\gamma)$ *and*

– $\psi_i^{(\gamma)}(z) = \psi_i(z+\gamma) - \psi_i(\gamma)$.

Furthermore,

$$F^{(\gamma)}(z) = (\mathrm{diag}(v_i(\gamma),\, i \in E))^{-1}[F(z+\gamma) - \kappa(\gamma)\mathrm{Id}]\mathrm{diag}(v_i(\gamma),\, i \in E),$$

and hence,

$$\kappa^{(\gamma)}(z) = \kappa(z+\gamma) - \kappa(\gamma).$$

Making use of this, the following proposition with properties of κ are often used in the literature.

PROPOSITION (3.10). *Suppose that F is defined in some open interval D of $\mathbb{R}$. Then, the leading eigenvalue κ of F is smooth and convex on D.*

In Section 2.2, we studied the exponential functional of Lévy processes, now we are interested in describing some results that are useful to compute the law of an integrated exponential functional associated to Markov additive processes.

For a MAP ξ, let

$$I(-\xi) = \int_0^\infty \exp(-\xi(t))\,\mathrm{d}\,t.$$

One way to characterise the law of $I(-\xi)$ is via its Mellin transform, which we write as $\mathcal{M}(s)$. This is the vector in $\mathbb{R}^N$ whose ith element is given by

$$\mathcal{M}_i(s) = \mathbf{E}_i[I(-\xi)^{s-1}],\ \text{for } i \in E.$$

We will shortly express a functional equation for $\mathcal{M}$, analogous to the functional equation for the Mellin transform for the exponential functional of Lévy processes which we saw in Section 2.2. For Lévy processes, proofs of the result can be found in [15], [45] and [55].

We make the following assumption, which is analogous to the Cramér condition for a Lévy process; recall that κ is the leading eigenvalue of the matrix F, as discussed in Section 3.1.

Definition (3.11) (Cramér condition for a Markov additive process). There exists $z_0 < 0$ such that $F(s)$ exists on $(z_0, 0)$, and some $\theta \in (0, -z_0)$, called the *Cramér number*, such that $\kappa(-\theta) = 0$.

Since the leading eigenvalue κ is smooth and convex where it is defined, it follows also that $\kappa(-s) < 0$ for $s \in (0,\theta)$. In particular, this renders the matrix $F(-s)$ negative definite, and hence invertible. Furthermore, it follows that $\kappa'(0-) > 0$, and hence (see [2, Corollary XI.2.7] and [33, Lemma 2.14]) that ξ drifts to $+\infty$ independently of its initial state. This implies that $I(-\xi)$ is an a.s. finite random variable.

PROPOSITION (3.12). *Suppose that ξ satisfies the Cramér condition (Assumption 3.11) with Cramér number $\theta \in (0,1)$. Then, $\mathcal{M}(s)$ is finite and analytic when $\Re(s) \in (0, 1+\theta)$, and we have the following vector-valued functional equation:*

$$\mathcal{M}(s+1) = -s(F(-s))^{-1}\mathcal{M}(s), \qquad \textit{for } s \in (0,\theta).$$

Received June 17, 2013

Final version received October 31, 2013

JUAN CARLOS PARDO,
CIMAT A.C.
CALLE JALISCO S/N.
C.P. 36240,
GUANAJUATO, MEXICO.
jcpardo@cimat.mx

VÍCTOR RIVERO,
CIMAT A.C.
CALLE JALISCO S/N.
C.P. 36240,
GUANAJUATO, MEXICO.
rivero@cimat.mx

REFERENCES

[1] S. ASMUSSEN, *Ruin probabilities*, vol. 2 of Advanced Series on Statistical Science & Applied Probability, World Scientific Publishing Co. Inc., River Edge, NJ, 2000.

[2] ———, *Applied probability and queues*, vol. 51 of Applications of Mathematics (New York), Springer-Verlag, New York, second ed., 2003. Stochastic Modelling and Applied Probability.

[3] M. BARCZY AND J. BERTOIN, *Functional limit theorems for Lévy processes satisfying Cramér's condition*, Electron. J. Probab., **16**, (2011), no. 73, 2020–2038.

[4] J. BERTOIN, *Lévy processes*, vol. 121 of Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 1996.

[5] J. BERTOIN AND M.-E. CABALLERO, *Entrance from* 0+ *for increasing semi-stable Markov processes*, Bernoulli, **8** (2), (2002), 195–205.

[6] J. BERTOIN, A. LINDNER, AND R. MALLER, *On continuity properties of the law of integrals of Lévy processes*, in Séminaire de probabilités XLI, vol. 1934 of Lecture Notes in Math., Springer, Berlin, 2008, 137–159.

[7] J. BERTOIN AND M. YOR, *On subordinators, self-similar Markov processes and some factorizations of the exponential variable*, Electron. Comm. Probab., **6**, (2001), 95–106 (electronic).

[8] ———, *The entrance laws of self-similar Markov processes and exponential functionals of Lévy processes*, Potential Anal., **17** (4), (2002), 389–400.

[9] ———, *Exponential functionals of Lévy processes*, Probab. Surv., **2**, (2005), 191–212.

[10] R. M. BLUMENTHAL, *Excursions of Markov processes*, Probability and its Applications, Birkhäuser Boston Inc., Boston, MA, 1992.

[11] M. E. CABALLERO AND L. CHAUMONT, *Weak convergence of positive self-similar Markov processes and overshoots of Lévy processes*, Ann. Probab., **34** (3), (2006), 1012–1034.

[12] M. E. CABALLERO, J. C. PARDO, AND J. L. PÉREZ, *Explicit identities for Lévy processes associated to symmetric stable processes*, Bernoulli, **17** (1), (2011), 34–59.

[13] M. E. CABALLERO AND V. RIVERO, *On the asymptotic behaviour of increasing self-similar Markov processes*, Electron. J. Probab., **14**, (2009), 865–894.

[14] N. CAI AND S. KOU, *Pricing Asian options under a general jump diffusion model*, Oper. Res., **60** (1), (2012), 64–77.

[15] P. CARMONA, F. PETIT, AND M. YOR, *On the distribution and asymptotic results for exponential functionals of Lévy processes*, in Exponential functionals and principal values related to Brownian motion, Bibl. Rev. Mat. Iberoamericana, Rev. Mat. Iberoamericana, Madrid, 1997, 73–130.

[16] L. CHAUMONT, A. KYPRIANOU, J. C. PARDO, AND V. RIVERO, *Fluctuation theory and exit systems for positive self-similar Markov processes*, Ann. Probab., **40** (1), (2012), 245–279.

[17] L. CHAUMONT, A. E. KYPRIANOU, AND J. C. PARDO, *Some explicit identities associated with positive self-similar Markov processes*, Stochastic Process. Appl., **119** (3), (2009), 980–1000.

[18] L. CHAUMONT, H. PANTÍ, AND V. RIVERO, *The Lamperti representation of real-valued self-similar Markov processes*, Bernoulli, (2013).

[19] L. CHAUMONT AND J. C. PARDO, *The lower envelope of positive self-similar Markov processes*, Electron. J. Probab., **11**, (2006), no. 49, 1321–1341.
[20] L. CHAUMONT AND V. RIVERO, *On some transformations between positive self-similar Markov processes*, Stochastic Process. Appl., **117** (12), (2007), 1889–1909.
[21] O. CHYBIRYAKOV, *The Lamperti correspondence extended to Lévy processes and semi-stable Markov processes in locally compact groups*, Stochastic Process. Appl., **116** (5), (2006), 857–872.
[22] R. A. DONEY AND A. E. KYPRIANOU, *Overshoots and undershoots of Lévy processes*, Ann. Appl. Probab., **16** (1), (2006), 91–106.
[23] M. D. DONSKER, *An invariance principle for certain probability limit theorems*, Mem. Amer. Math. Soc.„ **1951** (6), (1951), 12.
[24] L. DÖRING AND M. BARCZY, *A jump type SDE approach to positive self-similar Markov processes*, Electron. J. Probab., **17**, (2012), no. 94, 39.
[25] A. DVORETZKY AND P. ERDÖS, *Some problems on random walk in space*, in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950, Berkeley and Los Angeles, 1951, University of California Press, 353–367.
[26] P. EMBRECHTS AND M. MAEJIMA, *Selfsimilar processes*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, 2002.
[27] P. J. FITZSIMMONS, *On the existence of recurrent extensions of self-similar Markov processes*, Electron. Comm. Probab., **11**, (2006), 230–241.
[28] B. FRISTEDT, *The behavior of increasing stable processes for both small and large times*, J. Math. Mech., **13**, (1964), 849–856.
[29] B. HAAS, *Asymptotic behavior of solutions of the fragmentation equation with shattering: an approach via self-similar Markov processes*, Ann. Appl. Probab., **20** (2), (2010), 382–429.
[30] B. HAAS AND V. RIVERO, *Quasi-stationary distributions and Yaglom limits of self-similar Markov processes*, Stochastic Process. Appl., **122** (12), (2012), 4054–4095.
[31] ———, *Factorizations of generalized Pareto random variables using exponential functionals of Lévy processes*, 2013.
[32] F. HUBALEK AND A. KUZNETSOV, *A convergent series representation for the density of the supremum of a stable process*, Electron. Commun. Probab., **16**, (2011), 84–95.
[33] J. IVANOVS, *One-sided Markov Additive Processes and Related Exit Problems*, BOXPress, 2011.
[34] A. KHINTCHINE, *Sur la croissance locale des processus stochastiques homogènes à accroissements indépendants*, Bull. Acad. Sci. URSS. Sér. Math. [Izvestia Akad. Nauk SSSR], **1939**, (1939), 487–508.
[35] S. W. KIU, *Semistable Markov processes in* $\mathbf{R}^n$, Stochastic Process. Appl., **10** (2), (1980), 183–191.
[36] A. KUZNETSOV, *On extrema of stable processes*, Ann. Probab., **39** (3), (2011), 1027–1060.
[37] A. KUZNETSOV, *On the distribution of exponential functionals for Lévy processes with jumps of rational transform*, Stochastic Process. Appl., **122** (2), (2012), 654–663.
[38] A. KUZNETSOV, A. E. KYPRIANOU, J. C. PARDO, AND A. R. WATSON, *The hitting time of zero for a stable process*, ArXiv e-prints, (2012).
[39] A. KUZNETSOV AND J. C. PARDO, *Fluctuations of stable processes and exponential functionals of hypergeometric Lévy processes*, Acta Appl. Math., **123**, (2013), 113–139.
[40] A. E. KYPRIANOU, *Introductory lectures on fluctuations of Lévy processes with applications*, Universitext, Springer-Verlag, Berlin, 2006.
[41] A. E. KYPRIANOU AND J. C. PARDO, *Continuous-state branching processes and self-similarity*, J. Appl. Probab., **45** (4), (2008), 1140–1160.
[42] A. E. KYPRIANOU, J. C. PARDO, AND V. RIVERO, *Exact and asymptotic n-tuple laws at first and last passage*, Ann. Appl. Probab., **20** (2), (2010), 522–564.
[43] J. LAMPERTI, *Semi-stable stochastic processes*, Trans. Amer. Math. Soc., **104**, (1962), 62–78.
[44] ———, *Semi-stable Markov processes. I*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, **22**, (1972), 205–225.
[45] K. MAULIK AND B. ZWART, *Tail asymptotics for exponential functionals of Lévy processes*, Stochastic Process. Appl., **116** (2), (2006), 156–177.
[46] M. MOTOO, *Proof of the law of iterated logarithm through diffusion equation*, Ann. Inst. Statist. Math., **10**, (1958), 21–28.
[47] J. C. PARDO, *On the future infimum of positive self-similar Markov processes*, Stochastics, **78** (3), (2006), 123–155.

[48] ———, *The upper envelope of positive self-similar Markov processes*, J. Theoret. Probab., **22** (2), (2009), 514–542.
[49] J. C. PARDO, P. PATIE, AND M. SAVOV, *A Wiener-Hopf type factorization for the exponential functional of Lévy processes*, J. Lond. Math. Soc. (2), **86** (3), (2012), 930–956.
[50] J. C. PARDO, V. RIVERO, AND K. VAN SCHAIK, *On the density of exponential functionals of Lévy processes*, Bernoulli, (2013).
[51] P. PATIE, *Law of the absorption time of some positive self-similar Markov processes*, Ann. Probab., **40** (2), (2012), 765–787.
[52] P. PATIE AND M. SAVOV, *Exponential functional of Lévy processes: Generalized Weierstrass products and Wiener–Hopf factorization*, C. R. Math. Acad. Sci. Paris, **351** (9-10), (2013), 393–396.
[53] V. RIVERO, *A law of iterated logarithm for increasing self-similar Markov processes*, Stoch. Stoch. Rep., **75** (6), (2003), 443–472.
[54] ———, *Recurrent extensions of self-similar Markov processes and Cramér's condition*, Bernoulli, **11** (3), (2005), 471–509.
[55] ———, *Recurrent extensions of self-similar Markov processes and Cramér's condition. II*, Bernoulli, **13** (4), (2007), 1053–1070.
[56] ———, *Tail asymptotics for exponential functionals of Lévy processes: the convolution equivalent case*, Ann. Inst. Henri Poincaré Probab. Stat., **48** (4), (2012), 1081–1102.
[57] K.-I. SATO, *Lévy processes and infinitely divisible distributions*, vol. 68 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 1999. Translated from the 1990 Japanese original, Revised by the author.
[58] V. VIGON, *Simplifiez vos Lévy en titillant la factorisation de Wiener-Hopf*, PhD thesis, Université Louis Pasteur, 2002.
[59] J. VUOLLE-APIALA, *Itô excursion theory for self-similar Markov processes*, Ann. Probab., **22** (2), (1994), 546–565.
[60] J. VUOLLE-APIALA AND S. E. GRAVERSEN, *Duality theory for self-similar processes*, Ann. Inst. H. Poincaré Probab. Statist., **22** (3), (1986), 323–332.
[61] Y. XIAO, *Asymptotic results for self-similar Markov processes*, in Asymptotic methods in probability and statistics (Ottawa, ON, 1997), North-Holland, Amsterdam, 1998, 323–340.

Bol. Soc. Mat. Mexicana (3) Vol. 19, 2013

CONVEX RISK MEASURES: A SELECTION OF PROPERTIES AND ITS APPLICATIONS

LEONEL PÉREZ HERNÁNDEZ AND ERICK TREVIÑO AGUILAR

ABSTRACT. Fundamental problems in financial markets involve the understanding of risks at different levels. In this note we start with the axiomatic framework of a convex measure of risk and illustrate some of the new aspects being under research. We focus on a selection of properties capturing more aspects of real risks and allowing a more accurate modeling. We address the question of the minimality of the penalty function associated to a convex measure of risk, which for instance plays a central role in robust optimization problems. We also explain how our general result specializes to a Lévy filtration. It is interesting that many models based on Lévy processes include non bounded financial payoffs and we review recent advances in the representation of conditional risk measures for non bounded variables. We also discuss the key property of time consistency which is crucial in the dynamical minimization of risks. As an application illustrating the results being presented, we review the valuation of American options and the identification of times of maximal risk

1. Introduction

A fundamental step impulsing the development of the theory of convex risk measures is the axiomatic formulation building upon intuitive properties delineated by Ph. Artzner, F. Delbaen, J. Eber and D. Heath [3]. The financial principle 'diversification does not increase risk' transforms into a convexity property, and by methods of convex analysis Föllmer and Schied [20, 21] and Frittelli and Rosazza Gianin [22] clarify the general form of convex risk measures by means of *robust numerical representations* and *penalty functions*. Artzner et al. [3] work in a finite probability space. Representations of risk measures in a general probability space were obtained by Delbaen [8] for subadditive, first degree positive homogeneous risk measures, and by [20, 21] and [22] in the general convex case. These seminal papers initiated a new theory for the quantification of risks.

Efforts have been made to include more aspects of risk. Our goal in this note is to illustrate through a selection of results some recent developments in the theory and its applications.

The paper is organized as follows. In Section 2 we present results on risk measures. In Subsection 2.1 we introduce the axioms underlying the concept of a convex measure of risk and immediately become clear the role that minimal penalties will play. In Subsection 2.2 we discuss the minimality property of the penalty function in the context of Lévy filtrations. It turns out to be quite relevant, and it is a desirable property that is hard to prove in general. For instance, in the study of robust portfolio optimization the property is crucial; see [37] and [26]. In

2010 *Mathematics Subject Classification*: 60H30, 91B25, 91B30.

Keywords and phrases: American options, convex risk measures, Lévy processes, minimal penalty functions, time consistency.

Subsection 2.3 we present a result clarifying the general structure of a conditional risk measure quantifying non bounded payoffs. In subsection 2.4 we recall a key result on time consistent risk measures due to Föllmer and Penner [17] which is crucial for the next part. In Section 3 we present the solution to the valuation problem of American options as well as the existence of times of maximal risk.

2. Convex measures of risk

(2.1) Axiomatic setting. Let us first recall from Föllmer and Schied [20] the definition and some basic properties of convex risk measures.

Definition (2.1). Let $\mathcal{X}$ be a linear space of bounded functions containing the constants. A mapping $\rho : \mathcal{X} \to \mathbb{R}$ is called a *monetary measure of risk* if it satisfies the following conditions for all $X, Y \in \mathcal{X}$.

1. Monotonicity: If $X \leq Y$, then $\rho(X) \geq \rho(Y)$.
2. Cash invariance: If $m \in \mathbb{R}$, then $\rho(X+m) = \rho(X) - m$.

Definition (2.2). A monetary risk measure $\rho : \mathcal{X} \to \mathbb{R}$ is called a *convex measure of risk* if it satisfies

- Convexity: $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$, for $0 \leq \lambda \leq 1$.

Definition (2.3). A convex measure of risk $\rho : \mathcal{X} \to \mathbb{R}$ is called a *coherent measure of risk* if it satisfies

- Positive homogeneity: If $\lambda \geq 0$, then $\rho(\lambda X) = \lambda\rho(X)$.

Given any class $\mathcal{A}$ of subsets of Ω, a set function $\mathbb{Q} : \mathcal{A} \to \mathbb{R}$ is said to be *finitely additive* if $\mathbb{Q}(\emptyset) = 0$, $\sup_{A\in\mathcal{A}} |\mathbb{Q}(A)| < \infty$, and $\mathbb{Q}(A \cup B) = \mathbb{Q}(A) + \mathbb{Q}(B)$, for $A, B \in \mathcal{A}$ disjoint. We say that a set function $\mathbb{Q} : \mathcal{F} \to [0,1]$ is a *probability content* if it is finitely additive and $\mathbb{Q}(\Omega) = 1$. The set of *probability contents* on this measurable space is denoted by $\mathcal{Q}_{cont}$.

From the general theory of static convex risk measures (see in [19, Chapter 4]), we know that any map $\psi : \mathcal{Q}_{cont} \to \mathbb{R} \cup \{+\infty\}$, with $\inf_{\mathbb{Q}\in\mathcal{Q}_{cont}} \psi(\mathbb{Q}) \in \mathbb{R}$, induces a static convex measure of risk as a mapping $\rho : \mathfrak{M}_b \to \mathbb{R}$ given by

$$\rho(X) := \sup_{\mathbb{Q}\in\mathcal{Q}_{cont}} \{\mathbb{E}_{\mathbb{Q}}[-X] - \psi(\mathbb{Q})\}. \tag{2.4}$$

Here $\mathfrak{M}$ denotes the class of measurable functions and $\mathfrak{M}_b$ the subclass of bounded measurable functions. The function ψ will be referred as a *penalty function*. Föllmer and Schied [20, Theorem 3.2], and Frittelli and Rosazza Gianin [22, Corollary 7] proved that any convex risk measure is essentially of this form. More precisely, a convex measure of risk ρ on the space of bounded functions $\mathfrak{M}_b(\Omega, \mathcal{F})$ has the representation

$$\rho(X) = \sup_{\mathbb{Q}\in\mathcal{Q}_{cont}} \left\{\mathbb{E}_{\mathbb{Q}}[-X] - \psi^*_\rho(\mathbb{Q})\right\}, \tag{2.5}$$

where

$$\psi^*_\rho(\mathbb{Q}) := \sup_{X\in\mathcal{A}_\rho} \mathbb{E}_{\mathbb{Q}}[-X], \tag{2.6}$$

and $\mathcal{A}_\rho := \{X \in \mathfrak{M}_b : \rho(X) \leq 0\}$ is the *acceptance set* of ρ. The penalty ψ^*_ρ is called the *minimal penalty function* associated to ρ. Furthermore, for the minimal

penalty function, the next biduality relation is satisfied

$$\psi_\rho^*(\mathbb{Q}) = \sup_{X\in\mathfrak{M}_b(\Omega,\mathcal{F})} \{\mathbb{E}_\mathbb{Q}[-X] - \rho(X)\}, \quad \forall \mathbb{Q}\in\mathcal{Q}_{cont}. \tag{2.7}$$

We now introduce a probability space $(\Omega,\mathcal{F},\mathbb{P})$. We denote by $L^0(\mathcal{F})$ the linear space of equivalence classes of random variables which are measurable with respect to $\mathcal{F}$. For p satisfying $1\le p<+\infty$ the linear space

$$L^p(\mathcal{F}),$$

is the set of elements X in $L^0(\mathcal{F})$ such that $E_\mathbb{P}[|X|^p]<+\infty$. We denote by $L^\infty(\mathcal{F})$ the linear space of elements in $L^0(\mathcal{F})$ which are essentially bounded with respect to $\mathbb{P}$. The set of probability measures defined in the measurable space $(\Omega,\mathcal{F})$ which are absolutely continuous with respect to $\mathbb{P}$ is denoted by $\mathcal{P}$.

We consider convex risk measures defined on the Banach space $\mathcal{X}=L^\infty(\mathcal{F})$. A convex risk measure ρ on the Banach space of bounded measurable function on $(\Omega,\mathcal{F})$ may be viewed as a convex risk measure on $L^\infty(\mathcal{F})$ if it respects the $\mathbb{P}$-null sets, i.e.,

$$\rho(X)=\rho(Y), \quad \text{if } X=Y \quad \mathbb{P}-a.s.$$

The next theorem clarifies the structure of a convex risk measure on $L^\infty(\mathcal{F})$, by extending Delbaen's representation theorem for coherent measures of risk to the general convex case; see [8, Theorem 3.2].

THEOREM (2.8). *Suppose $\mathcal{X}=L^\infty(\mathcal{F})$, and $\rho:\mathcal{X}\to\mathbb{R}$ is a convex measure of risk. Then the following properties are equivalent.*

1. *There is a "penalty function" $\alpha:\mathcal{P}\to(-\infty,\infty]$ such that*

$$\rho(X)=\sup_{P\in\mathcal{P}}(E_P[-X]-\alpha(P)), \text{ for all } X\in\mathcal{X}, \tag{2.9}$$

   with

$$\alpha(P):=\sup_{X\in\mathcal{X}}\big(E_P[-X]-\rho(X)\big).$$

2. *ρ possesses the Fatou property: If the sequence $(X_n)_{n\in\mathbb{N}}\subset\mathcal{X}$ is uniformly bounded, and X_n converges to some $X\in\mathcal{X}$ in probability, then $\rho(X)\le\liminf_n\rho(X_n)$.*
3. *If the sequence $(X_n)_{n\in\mathbb{N}}\subset\mathcal{X}$ decreases to $X\in\mathcal{X}$, then $\rho(X_n)\to\rho(X)$.*

In the coherent case, the representation (2.9) reduces to the representation

$$\rho(X)=\sup_{Q\in\mathcal{Q}}E_Q[-X], \text{ for all } X\in\mathcal{X}. \tag{2.10}$$

for the family $\mathcal{Q}=\{Q\in\mathcal{P}\mid\alpha(Q)=0\}$.

Proof. See Theorem 6 in Föllmer and Schied [21], parts *1, 3, 4*. □

For a systematic exposition of risk measures and a comparative analysis of its properties see e.g., Denuit and Dhaene [10], Föllmer and Schied [19] and Hult et al [27].

(2.2) Minimal penalty functions. Let $\mathcal{Q}(\Omega,\mathcal{F})$ be the family of probability measures on the measurable space $(\Omega,\mathcal{F})$. We denote by $\mathcal{Q}_{\approx}(\mathbb{P})$ the subclass of equivalent probability measure. Of course, $\mathcal{Q}_{\approx}(\mathbb{P}) \subset \mathcal{P} \subset \mathcal{Q}(\Omega,\mathcal{F})$. Among the measures of risk, the class of them that are concentrated on the set of probability measures $\mathcal{Q} \subset \mathcal{Q}_{cont}$ are of special interest.

When we deal with a set of measures $\mathcal{K} \subset \mathcal{P}$ it is necessary to make reference to some topological concepts, meaning that we are considering the corresponding set of densities and the strong topology in $L^1(\mathbb{P})$.

LEMMA (2.11). *For a function $\psi : \mathcal{K} \subset \mathcal{P} \to \mathbb{R}_+ \cup \{+\infty\}$ with $\inf_{\mathbb{Q}\in\mathcal{K}} \psi(\mathbb{Q}) \in \mathbb{R}$ define the extension $\psi(\mathbb{Q}) := \infty \ \forall \mathbb{Q} \in \mathcal{Q}_{cont} \setminus \mathcal{K}$ with $\mathcal{K}$ a convex closed set. Also define the function Ψ with domain in L^1 as*

$$\Psi(Z) := \begin{cases} \psi(\mathbb{Q}) & \text{if } Z = d\mathbb{Q}/d\mathbb{P} \\ \infty & \text{otherwise} \end{cases}.$$

Then for the convex measure of risk $\rho(X) := \sup_{\mathbb{Q}\in\mathcal{Q}_{cont}} \{\mathbb{E}_{\mathbb{Q}}[-X] - \psi(\mathbb{Q})\}$ associated to ψ holds:

(a) If ρ has as minimal penalty ψ^*_ρ the function ψ (i.e. $\psi = \psi^*_\rho$), then Ψ is a proper convex function and lower semicontinuous w.r.t. the (strong) L^1-topology or equivalently w.r.t. the weak topology $\sigma(L^1, L^\infty)$.

(b) If Ψ is a proper convex function and lower semicontinuous w.r.t. the (strong) L^1-topology or equivalently w.r.t. the weak topology $\sigma(L^1, L^\infty)$, then $\psi \mathbf{1}_{\mathcal{P}} = \psi^*_\rho \mathbf{1}_{\mathcal{P}}$.

See [25] for details.

2.2.1. *Density processes and penalty functions.* We want to indicate now how to take advantage of Lemma (2.11) in a concrete probability space. We say that $L := \{L_t\}_{t\in\mathbb{R}_+}$ is a Lévy process for this probability space if it is an adapted càdlàg process with independent stationary increments starting at zero. The filtration considered is $\mathbb{F} := \{\mathcal{F}^{\mathbb{P}}_t(L)\}_{t\in\mathbb{R}_+}$, the completion of its natural filtration. The jump measure of L is denoted by $\mu : \Omega \times (\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{B}(\mathbb{R}_0)) \to \mathbb{N}$ where $\mathbb{R}_0 := \mathbb{R} \setminus \{0\}$. The dual predictable projection of this measure, also known as its Lévy system, satisfies the relation $\mu^{\mathcal{P}}(dt,dx) = dt \times \nu(dx)$, where $\nu(\cdot) := \mathbb{E}[\mu([0,1] \times \cdot)]$ is the intensity or Lévy measure of L. Denote by $\mathcal{A}_{loc}$ (resp. $\mathcal{A}^+_{loc}$) the collection of adapted processes with locally integrable variation (resp. adapted locally integrable increasing processes). For a càdlàg process X we denote by $X_- := (X_{t-})$ the left hand limit process, where $X_{0-} := X_0$ by convention, and by $\triangle X = (\triangle X_t)$ the jump process $\triangle X_t := X_t - X_{t-}$. Further, we denote by $\mathcal{P}_d \subset \mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+)$ the predictable σ-algebra and by $\widetilde{\mathcal{P}} := \mathcal{P}_d \otimes \mathcal{B}(\mathbb{R}_0)$. With some abuse of notation, we write $\theta_1 \in \widetilde{\mathcal{P}}$ when the function $\theta_1 : \Omega \times \mathbb{R}_+ \times \mathbb{R}_0 \to \mathbb{R}$ is $\widetilde{\mathcal{P}}$-measurable and $\theta \in \mathcal{P}_d$ for predictable processes. For a semimartingale U let

$$\begin{aligned} \mathcal{L}(U^c) := \quad & \{\theta \in \mathcal{P}_d : \exists \{\tau_n\}_{n\in\mathbb{N}} \text{ sequence of stopping times with } \tau_n \uparrow \infty \\ & \text{and } \mathbb{E}\left[\int_0^{\tau_n} \theta^2 d[U^c]\right] < \infty \ \forall n \in \mathbb{N}\} \end{aligned} \tag{2.12}$$

be the class of predictable processes $\theta \in \mathcal{P}_d$ integrable with respect to the continuous part U^c in the sense of local martingale, and by

$$\Lambda(U^c) := \left\{ \int \theta_0 dU^c : \theta_0 \in \mathcal{L}(U^c) \right\}$$

the linear space of processes which admits a representation as the stochastic integral with respect to U^c. For the Lévy process L with jump measure μ, denote by

$$\mathcal{G}(\mu) \equiv \left\{ \theta_1 \in \widetilde{\mathcal{P}} : \left\{ \sqrt{\sum_{s \le t} \{\theta_1(s, \triangle L_s)\}^2 \mathbf{1}_{\mathbb{R}_0}(\triangle L_s)} \right\}_{t \in \mathbb{R}_+} \in \mathcal{A}^+_{loc} \right\}. \tag{2.13}$$

the domain of the functional $\theta_1 \to \int \theta_1 d\left(\mu - \mu^{\mathcal{P}}\right)$. We use the notation $\int \theta_1 d(\mu - \mu^{\mathcal{P}})$ to write the value of this functional in θ_1. It is important to point out that this integral functional is not, in general, the integral with respect to the difference of two measures. For a detailed exposition on these topics see He, Wang and Yan [24] or Jacod and Shiryaev [29], which are our basic references.

We say that the semimartingale U has the *weak property of predictable representation* when

$$\mathcal{M}_{loc,0} = \Lambda(U^c) + \left\{ \int \theta_1 d\left(\mu_U - \mu_U^{\mathcal{P}}\right) : \theta_1 \in \mathcal{G}(\mu_U) \right\}, \tag{2.14}$$

where the previous sum is the linear sum of the vector spaces, and $\mathcal{M}_{loc,0}$ is the linear space of local martingales starting at zero. Given an absolutely continuous probability measure $\mathbb{Q} \ll \mathbb{P}$ in a filtered probability space, where a semimartingale with the weak predictable representation property is defined, the structure of the density process has been studied extensively by several authors; see Theorem 14.41 in He, Wang and Yan [24] or Theorem III.5.19 in Jacod and Shiryaev [29]. Denote by $D_t := \mathbb{E}\left[\frac{d\mathbb{Q}}{d\mathbb{P}} \middle| \mathcal{F}_t \right]$ the càdlàg version of the density process. For the increasing sequence of stopping times $\tau_n := \inf\{t \ge 0 : D_t < \frac{1}{n}\}$ $n \ge 1$ and $\tau_0 := \sup_n \tau_n$ we have $D_t(\omega) = 0$ $\forall t \ge \tau_0(\omega)$ and $D_t(\omega) > 0$ $\forall t < \tau_0(\omega)$, i.e.

$$D = D\mathbf{1}_{[\![0,\tau_0[\![} \tag{2.15}$$

and the process

$$\frac{1}{D_{s-}} \mathbf{1}_{[\![D_- \neq 0]\!]} \text{ is integrable w.r.t. } D, \tag{2.16}$$

where we abuse of the notation by setting $[\![D_- \neq 0]\!] := \{(\omega, t) \in \Omega \times \mathbb{R}_+ : D_{t-}(\omega) \neq 0\}$. Both conditions (2.15) and (2.16) are necessary and sufficient in order a semimartingale to be an *exponential semimartigale*, i.e. $D = \mathcal{E}(Z)$ the Doléans-Dade exponential of another semimartingale Z. In that case we have

$$\tau_0 = \inf\{t > 0 : D_{t-} = 0 \text{ or } D_t = 0\} = \inf\{t > 0 : \triangle Z_t = -1\}. \tag{2.17}$$

The following Lemma deals with such a characterization for the case of a Lévy-process, which as known fulfills the weak property of predictable representation for its completed natural filtration.

LEMMA (2.18). *For any absolutely continuous probability measure $\mathbb{Q} \ll \mathbb{P}$ there are coefficients $\theta_0 \in \mathcal{L}(W)$ and $\theta_1 \in \mathcal{G}(\mu)$, such that*

$$\frac{d\mathbb{Q}_t}{d\mathbb{P}_t} = \frac{d\mathbb{Q}_t}{d\mathbb{P}_t} \mathbf{1}_{[\![0,\tau_0[\![} = \mathcal{E}\left(Z^\theta\right)(t), \tag{2.19}$$

where $Z^\theta_t \in \mathcal{M}_{loc}$ is the local martingale given by

$$Z^\theta_t := \int_{]0,t]} \theta_0 dW + \int_{]0,t] \times \mathbb{R}_0} \theta_1(s,x)\left(\mu(ds,dx) - ds\ \nu(dx)\right). \tag{2.20}$$

The coefficients θ_0 and θ_1 are $dt\times\mathbb{P}$-a.s and $\mu^{\mathcal{P}}_{\mathbb{P}}(ds,dx)\times\mathbb{P}$-a.s. unique on $[\![0,\tau_0]\!]$ and $[\![0,\tau_0]\!]\times\mathbb{R}_0$ respectively .The coefficients can be choosen with $\theta_0=0$ on $]\!]\tau_0,\infty[\![$ and $\theta_1=0$ on $]\!]\tau_0,\infty[\![\times\mathbb{R}$.

Now, we shall introduce a family of penalty functions for the density processes described in Lemma (2.18), for the absolutely continuous measures $\mathbb{Q}\in\mathcal{P}$.

Let $h:\mathbb{R}_+\to\mathbb{R}_+$ and $h_0,h_1\ \mathbb{R}\to\mathbb{R}_+$ be convex functions with $0=h(0)=h_0(0)=h_1(0)$. Further define the penalty function

$$\begin{aligned}\vartheta(\mathbb{Q}) \quad &: \quad =\mathbb{E}_{\mathbb{Q}}\left[h\left(\int_0^{T\wedge\tau_0}\left(h_0(\theta_0(t))+\int_{\mathbb{R}_0}h_1(\delta(t,x)\theta_1(t,x))\nu(dx)\right)dt\right)\right]\mathbf{1}_{\mathcal{P}}(\mathbb{Q})\\ &\qquad +\infty\times\mathbf{1}_{\mathcal{Q}_{cont}\setminus\mathcal{P}}(\mathbb{Q}),\end{aligned}\tag{2.21}$$

where θ_0, θ_1 are the processes associated to $\mathbb{Q}$ from Lemma (2.18), and $\delta(t,x):\mathbb{R}_+\times\mathbb{R}_0\to\mathbb{R}_+$ is an arbitrary fixed nonnegative function $\delta(t,x)\in\mathcal{G}(\mu)$. Since $\theta_0\equiv 0$ on $[\![\tau_0,\infty[\![$ and $\theta_1\equiv 0$ on $[\![\tau_0,\infty[\![\times\mathbb{R}_0$ we have from the conditions imposed to h,h_0, and h_1

$$\begin{aligned}\vartheta(\mathbb{Q}) \quad &= \quad \mathbb{E}_{\mathbb{Q}}\left[h\left(\int_0^{T}\left(h_0(\theta_0(t))+\int_{\mathbb{R}_0}h_1(\delta(t,x)\theta_1(t,x))\nu(dx)\right)dt\right)\right]\mathbf{1}_{\mathcal{P}}(\mathbb{Q})\\ &\qquad +\infty\times\mathbf{1}_{\mathcal{Q}_{cont}\setminus\mathcal{P}}(\mathbb{Q}).\end{aligned}\tag{2.22}$$

Further, define the convex measure of risk

$$\rho(X):=\sup_{\mathbb{Q}\in\mathcal{P}}\left\{\mathbb{E}_{\mathbb{Q}}[-X]-\vartheta(\mathbb{Q})\right\}.\tag{2.23}$$

Notice that ρ is a normalized and sensitive measure of risk.

The next theorem establishes the minimality of the penalty function introduced above for the risk measure ρ. Its proof is based on the sufficient conditions given in Theorem 2.11 and details might be found in [25].

THEOREM (2.24). *The penalty function ϑ defined in (2.21) equals the restriction on $\mathcal{P}$ of the minimal penalty function of the convex risk measure ρ given by (2.23).*

(2.3) Conditional measures of risk in spaces beyond $L^\infty(\mathcal{F})$. The Banach space of essentially bounded random variables represents the payoff of financial positions with *limited size of a loss*. In many situations this is an unrealistic assumption, e.g., losses produced by high leverage.

The quantification of financial payoff's as measurable random variables without any integrability condition is only possible if we allow the value $+\infty$ so that there are positions which are going to be unacceptable, no matter how we reduce its size through a marginal reserve. More precisely: If the space $(\Omega,\mathcal{F},\mathbb{P})$ is atomless, then there is no normalized, real-valued convex risk measure ρ defined in the space of measurable random variables $L^0(\mathcal{F})$. The result is proved by Delbaen [8], Theorem 5.1 for coherent risk measures. The proof for the convex case is similar.

However, if we consider the spaces $L^p(\mathcal{F})$, which satisfy $L^\infty(\mathcal{F})\subset L^p(\mathcal{F})\subset L^0(\mathcal{F})$, we may insist on real-valued quantifications of risk and still obtain interesting classes of measures of risk. Quantifying risks beyond $L^\infty(\mathcal{F})$ has been the subject of recent research; see e.g., Cheridito and Li [7], Filipović and Svindland [14], Kaina and Rüschendorf [30].

In a period of time, say one year, new information arrives. Thus, a question of interest is, how to construct measures of risk whose quantifications incorporate the benefit of new information. This question has motivated the development of conditional convex risk measures and more in general of dynamical risk measures. The discussion of this new class and the crucial robust representations have been considered in the literature. A systematic study of dynamical convex risk measures in $L^\infty(\mathcal{F})$ is presented by Föllmer and Penner [17].

Thus, risk measures quantifying random variables beyond the space of bounded payoff's $L^\infty(\mathcal{F})$ and explicitly taking into account the arrival of new information, together with an analogous robust representation, is interesting and required by realistic applications.

In order to study the general structure of such risk measures we go into more details.

Let us start with the definition of a conditional convex risk measure in the setting of Detlefsen and Scandolo [11].

Definition (2.25). Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. A map $\rho : L^\infty(\mathcal{F}) \to L^\infty(\mathcal{G})$ is called a *conditional convex risk measure* if it satisfies the following properties for all $X, Y \in L^\infty(\mathcal{F})$

1. Conditional cash invariance: For all $Z \in L^\infty(\mathcal{G})$, $\rho(X+Z) = \rho(X) - Z$.
2. Monotonicity: $X \le Y$ implies $\rho(X) \ge \rho(Y)$.
3. Conditional convexity: For all $\lambda \in L^\infty(\mathcal{G})$, $0 \le \lambda \le 1$:

$$\rho(\lambda X + (1-\lambda)Y) \le \lambda\rho(Y) + (1-\lambda)\rho(Y).$$

4. Normalization $\rho(0) = 0$.

A conditional convex risk measure is called a conditional coherent risk measure if it has in addition the conditional positive homogeneity property

$$\rho(\lambda X) = \lambda\rho(X), \text{ for } \lambda \in L^\infty(\mathcal{G}) \text{ with } \lambda \ge 0.$$

For the structure of this class of risk measures see Detlefsen and Scandolo [11], Theorem 1. We see an extension to conditional risk measures in the following setting. We denote by $\overline{L}^0(\mathcal{G})$ the family of $\mathcal{G}$-measurable random variables taking values in $\mathbb{R} \cup \{+\infty, -\infty\}$.

Definition (2.26). A conditional convex risk measure $\rho : L^p(\mathcal{F}) \to \overline{L}^0(\mathcal{G})$ is *real-valued* if it takes values in $L^1(\mathcal{G})$. More precisely, for each $X \in L^p(\mathcal{F})$ we have $\rho(X) \in L^1(\mathcal{G})$.

We distinguish a special class of absolutely continuous probability measures:

$$\mathcal{Q}^q := \left\{ Q \in \mathcal{P} \mid \frac{dQ}{d\mathbb{P}} \in L^q(\mathbb{P}) \right\}. \tag{2.27}$$

Definition (2.28). Let $\mathcal{Q} \subset \mathcal{Q}^q$ be a class of absolutely continuous probability measures. A *penalty function* is a correspondence of the form

$$\alpha : \mathcal{Q} \to \overline{L}^0_+(\mathcal{G}).$$

The pair $(\mathcal{Q}, \alpha)$ represents the convex risk measure ρ if

$$\rho(X) = \operatorname{ess\,sup}_{Q \in \mathcal{Q}} \{E_Q[-X \mid \mathcal{G}] - \alpha(Q)\}, \mathbb{P} - a.s, \text{ for each } X \in L^p(\mathcal{F}). \tag{2.29}$$

In this case, we say that the conditional convex risk measure ρ is representable and (2.29) defines a robust representation.

Let us introduce the class

$$\mathcal{Q}^q_{e,loc} := \{Q \in \mathcal{Q}^q \mid E_{\mathbb{P}}\left[\frac{dQ}{dP} \mid \mathcal{G}\right] > 0, \mathbb{P}-a.s.\}.$$

Definition (2.30). Let $\alpha : \mathcal{Q}^q_{e,loc} \to \overline{L}^0_+(\mathcal{G})$ be a penalty function. We say that α is a *coercive penalty function* if there exist real constants a, b with $b > 0$ and

$$E_{\mathbb{P}}[\alpha(Q)] \geq a + b E_{\mathbb{P}}\left[\frac{1}{E_{\mathbb{P}}\left[\frac{dQ}{d\mathbb{P}} \mid \mathcal{G}\right]} E^{\frac{1}{q}}_{\mathbb{P}}\left[\left(\frac{dQ}{d\mathbb{P}}\right)^q \mid \mathcal{G}\right]\right], Q \in \mathcal{Q}^q_{e,loc}. \tag{2.31}$$

The next result clarifies the general structure of a real-valued conditional convex risk measure.

THEOREM (2.32). 1. *Let $\alpha : \mathcal{Q}^q_{e,loc} \to \overline{L}^0_+(\mathcal{G})$ be a penalty function. Assume there exists $Q^0 \in \mathcal{Q}^q_{e,loc}$ such that $\alpha(Q^0) \in L^\infty(\mathcal{G})$. Let us define a mapping ρ by*

$$\rho(X) := \operatorname{ess\,sup}_{Q \in \mathcal{Q}^q_{e,loc}} \{E_Q[-X \mid \mathcal{G}] - \alpha(Q)\}, X \in L^p(\mathcal{F}).$$

Then ρ is a real-valued conditional risk measure, if α is coercive.

2. *Conversely, let ρ be a real-valued conditional convex risk measure. If the pair $(\mathcal{Q}^q_{e,loc}, \alpha)$ represents the convex risk measure ρ, then the penalty function α must be coercive.*

In a first step it is proved that real-valued conditional convex risk measures always admit a numerical representation in terms of a nice class of "locally equivalent" probability measures. The extended Namioka-Klee Theorem, due to Biagini and Frittelli [6], is crucial in this part. In a second step, duality of continuous operators in spaces L^p is important. The robust representation, together with an invariance property characterize conditional convex risk measures defined in a space $L^\infty(\mathcal{F})$ which can be extended to a space $L^p(\mathcal{F})$, and at the same time continue to be real-valued. See [40, Theorems 2.7, 2.9] for details.

In particular the measure of risk AVaR, can be extended from their natural domain $L^\infty(\mathcal{F})$ to a space $L^p(\mathcal{F})$, For details see [40, Example 5.3]. Let us illustrate how we can construct a conditional real-valued version of AVaR in a space $L^p(\mathcal{F})$.

Let λ be a $\mathcal{G}$-measurable random variable such that

$$0 < \lambda \leq 1, \mathbb{P}-a.s.$$

The conditional Average Value at Risk at the stochastic level λ is defined by the representation:

$$AVaR_\lambda(X) = \operatorname{ess\,sup}_{Q \in \mathcal{Q}_\lambda} E_Q[-X \mid \mathcal{G}], \mathbb{P}-a.s. \tag{2.33}$$

where $\mathcal{Q}_\lambda$ is the set of all absolutely continuous probability measures whose $\mathbb{P}$-density $\frac{dQ}{d\mathbb{P}}$ is $\mathbb{P}$-a.s. bounded by $\frac{1}{\lambda}$ and moreover

$$E_{\mathbb{P}}\left[\frac{dQ}{d\mathbb{P}} \mid \mathcal{G}\right] = 1, \mathbb{P}-a.s.$$

The numerical representation (2.33) involves the minimal penalty function

$$\alpha^{\min}(Q) = \begin{cases} 0, & \text{if } Q \in \mathcal{Q}_\lambda, \\ +\infty, & \text{if } Q \notin \mathcal{Q}_\lambda, \end{cases} \tag{2.34}$$

which is coercive if

$$\frac{1}{\lambda} \in L^q(\mathcal{G}),$$

where q is the conjugate number of p. In this case, the conditional Average Value at Risk defines a real-valued conditional risk measure in $L^p(\mathcal{F})$.

(2.4) Dynamical convex risk measures. Time consistency is a principle for decision making in an intertemporal setting. An insight on how to consistently update the quantification of an uncertain lottery taking into account new information is provided by Epstein and Schneider [13]. Building upon the atemporal multiple-priors model of Gilboa and Schmeidler [23], Epstein and Schneider axiomatize preferences which are able to summarize those dynamic behaviors partially committing to a future course of action. Their robust representation, Theorem 3.2, clarifies the general structure of dynamically consistent preferences.

We will see in the context of risk, a version of a time consistency axiom and the corresponding numerical representation.

Take a filtration in discrete time of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, that is, a family $\{\mathcal{F}_t\}_{t=0,\dots,T}$ of σ-algebras satisfying $\mathcal{F}_{t-1} \subset \mathcal{F}_t \subset \mathcal{F}$. Denote by L_t^∞ the class of random variables of L^∞ which are measurable with respect to $\mathcal{F}_t$.

Definition (2.35). A sequence of conditional risk measures $\{\rho_t\}_{t=0,\dots,T}$ is called *time-consistent*, if for any $X, Y \in L^\infty$ and for $t = 0, \dots, T$ the following condition holds:

$$\rho_{t+1}(X) = \rho_{t+1}(Y), \mathbb{P}-a.s. \Longrightarrow \rho_t(X) = \rho_t(Y), \mathbb{P}-a.s.$$

The following family of probability measures plays a crucial role in characterizing time consistency

$$\mathcal{Q}^* := \{Q \in \mathcal{P} \mid Q \approx \mathbb{P} \text{ and } \alpha_0^{\min}(Q) < \infty\},$$

where

$$\alpha_t^{\min}(Q) = \operatorname{esssup}_{X \in L^\infty} \{E_Q[-X \mid \mathcal{F}_t] - \rho_t(X)\}.$$

The following characterization of time consistency is due to Föllmer and Penner [17, Theorem 4.5].

THEOREM (2.36). *Let $\{\rho_t\}_{t=0,1\dots,T}$ be a sequence of conditional convex risk measures such that each ρ_t is continuous from above, and assume that $\mathcal{Q}^*$ is non empty. Then, the following conditions are equivalent:*

1. *$\{\rho_t\}_{t=0,1\dots,T}$ is time consistent*
2. *For all $Q \in \mathcal{Q}^*$ and all $X \in L^\infty$, the process*

$$V_t^Q(X) := \rho_t(X) + \alpha_t^{\min}(Q),$$

is a Q-supermartingale.

In each case, the dynamic risk measure admits a robust representation in terms of the set $\mathcal{Q}^*$.

In Section 3.2 we present an application to time-consistency in the context of American options. We conclude this section with the class of conditional entropic risk measures. Given an "index of risk aversion" $\gamma > 0$, they are defined by

$$\rho_t^\gamma(X) := \frac{1}{\gamma} \log E_{\mathbb{P}}[e^{-\gamma X} \mid \mathcal{F}_t].$$

The entropic risk measure is time consistent: Take two random variables X and Y such that

$$\rho_t^\gamma(X) = \rho_t^\gamma(Y).$$

Then

$$E_{\mathbb{P}}[e^{-\gamma X} \mid \mathcal{F}_t] = E_{\mathbb{P}}[e^{-\gamma Y} \mid \mathcal{F}_t].$$

As a consequence

$$E_{\mathbb{P}}[e^{-\gamma X} \mid \mathcal{F}_{t-1}] = E_{\mathbb{P}}[e^{-\gamma Y} \mid \mathcal{F}_{t-1}],$$

due to the "tower property" of conditional expectations. Then we get

$$\rho_{t-1}^\gamma(X) = \rho_{t-1}^\gamma(Y),$$

which is precisely the property of time consistency. The entropic risk measure is introduced by Föllmer and Schied [18] in the static case and its conditional version is discussed by Detlefsen and Scandalo [11]. This class is characterized as the only dynamic risk measure which is law invariant, time consistent and relevant, see Kupper and Schachermayer [34].

3. American options

The role of derivatives in real economics have been advanced e.g., through the Arrow-Debreu securities. Futures are seen as contracts improving the "efficiency" of the real-economic market by easing exchanges of goods. For example, Ross [35] studies the completion of the market by call and put options. Interestingly, in recent research, some papers start by the completion of a market with some exogenous contracts for the valuation of complex derivatives. We can also think of derivatives written on commodities which many times are used as a form of insurance.

An American call option is a financial contract giving the right to buy an amount of an underlying asset at a prespecified exercise price. The holder of the contract has the right to exercise in a period of time.

The two sides of the contract are confronted to several intertemporal decisions: Is the price fair?, from the point of view of the buyer, given new information of how market is evolving, when is the best time to exercise? From the point of view of the seller, how should the hedging portfolio be readjusted?

Profit and Losses for both parts are risky due to the unpredictability of the market. Time consistency is crucial at different levels depending on replicability of the contract. If it is replicable, only the party not being able to follow a time consistent discipline, suffers a loss. This is what we learn from an analysis based on no arbitrage prices.

If the contract is non replicable, the typical situation in incomplete markets, the analysis is far more complex. Time-consistency continues to be crucial and lies as one of our most basic assumptions: The agents take prices in view of future actions and they have the discipline to commit to a plan. Indeed, as we are going

to see, it is crucial to determine rules of exercise and how to adjust the hedging portfolio.

In complete financial markets, the valuation of American options is characterized through its Snell envelope with respect to the unique equivalent martingale measure of the underlying price process. The Snell envelope is a supermartingale with a Doob-Meyer decomposition into a martingale and a non increasing predictable process. The representation of the martingale part as stochastic integral with respect to the underlying price process yields a hedging strategy for the seller of the American option. The stopping problem giving rise to the Snell envelope, determines an optimal exercise for the buyer. The expected reward of optimal stopping characterizes a 'fair' price; see Bensoussan [4] and Karatzas [31].

In incomplete markets, the analysis becomes more complex and has motivated the development of new techniques. In this more general setting, the class of martingale measures is infinite and the valuation of an American option is solved through robust versions of the Snell envelope, *the lower- and the upper- Snell envelopes*. The upper Snell envelope has been studied by El Karoui and Quenez [12], Kramkov [33], Föllmer and Kramkov [16], and Föllmer and Kabanov [15]. It has the remarkable property of having a decomposition into a stochastic integral part and a non increasing optional process, this is the celebrated *optional decomposition theorem*. The stochastic integral dominates the underlying payoff in a minimal way. It characterizes an strategy of a "complete insurance" for the seller and it is called a superhedging strategy. It characterizes the upper bound of no arbitrage prices.

The lower Snell envelope for American options has been systematically studied by Föllmer and Schied [19]. It characterizes the lower bound of no arbitrage prices.

Our goal in this section is to review the solution of the valuation of American options in a general semimartingale model. Our exposition is based on the discussion of chapter six in Föllmer and Schied [19] and the presentation in continuous time of [38].

Take a stochastic base $(\Omega, \mathcal{F}, \mathbb{F} = \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$, with finite time horizon. We model the discounted price of an asset in a **financial market** by an $\mathbb{F}$-adapted semimartingale $X := \{X_t\}_{0\le t\le T}$, defined in the domain $\Omega \times [0,T]$, whose trajectories are right continuous and have finite left limits (càdlàg). We assume the market is free of arbitrage opportunities in the sense that the set of equivalent local martingale measures defined by

$$\mathcal{M} := \{P \sim \mathbb{P} \mid X \text{ is a local martingale under } P\}, \tag{3.1}$$

is nonempty. For the precise formulation of the relationship between the notion of an arbitrage free market and the family of martingale measures we refer to Delbaen and Schachermayer [9] and references therein. For any martingale measure $P \in \mathcal{M}$, we denote by $E_P[\cdot]$ the corresponding P-expectation. Denote by $\mathcal{T}$ the family of $\mathbb{F}$-stopping times with values in $[0,T]$. We model the payoff of an American option by an $\mathbb{F}$-adapted non negative process $H := \{H_t\}_{0\le t\le T}$ with càdlàg trajectories. It must satisfy the following integrability condition

$$\sup_{P\in\mathcal{M}} \sup_{\theta\in\mathcal{T}} E_P[H_\theta] < \infty. \tag{3.2}$$

An admissible strategy is a pair (c,ξ) where $c \in \mathbb{R}_+$ is a positive constant, and $\xi := \{\xi_t\}_{0\le t\le T}$ is a $\mathbb{F}$-predictable process $\xi : \Omega\times[0,T] \to \mathbb{R}$ such that the value process $V_t^{c,\xi} := c + \int_0^t \xi_s dX_s$ is well defined and nonnegative. In this case we say that ξ is a c-admissible strategy and the family of c-admissible strategies is denoted by Ad_c.

REMARK (3.3). *Admissible strategies as considered here are sufficient for our purposes of exposition. This is a subtle issue. For example, the optimal portfolio allocation with respect to a real-valued utility function in the entire interval* $(-\infty,+\infty)$ *requires a larger class of admissible strategies, see e.g.,* [36], [5].

It is natural to ask whether it is possible to hedge the risk in an American option completely. This leads us to the concept of a superhedging strategy. A superhedging strategy for H is a pair $(c,\xi) \in \mathbb{R}_+ \times Ad_c$ such that

$$V^{c,\xi} \ge H.$$

Definition (3.4). The American option H is called *replicable* if there exists a superhedging strategy (c,ξ) for some $c \in \mathbb{R}_+$, and a stopping time $\tau \in \mathcal{T}$, such that $V_{t\wedge\tau}^{c,\xi}$ is a martingale for each $P \in \mathcal{M}$ and

$$V_\tau^{c,\xi} = H_\tau.$$

In this case, we say that the strategy (c,ξ) replicates the American option H.

We will see that superhedging strategies exist, and we are interested in the minimal capital that allows to construct such strategies.

Definition (3.5). The *superhedging cost* of the American option H is defined by

$$\inf\{c \ge 0 \mid \exists \xi \in Ad_c, (c,\xi) \text{ is a superhedging strategy}\}.$$

A superhedging strategy (c_0,ξ) with c_0 being equal to the superhedging cost is called minimal.

The superhedging cost of the American option H is characterized in the following important result.

THEOREM (3.6). *The superhedging cost of the American option H is equal to*

$$U_0^{\uparrow} = \sup_{P\in\mathcal{M}} \sup_{\theta\in\mathcal{T}} E_P[H_\theta], \tag{3.7}$$

and there exists a minimal superhedging strategy $(U_0^{\uparrow},\xi)$.

The proof requires a special uniform decomposition clarified in the *optional decomposition theorem* in the following form.

THEOREM (3.8). *Let* $\{U_t\}_{0\le t\le T}$ *be a positive càdlàg* $\mathcal{M}$*–supermartingale with*

$$U_0 = \sup_{P\in\mathcal{M}} \sup_{\theta\in\mathcal{T}} E_P[U_\theta] < \infty.$$

Then, there exists $\xi \in Ad_{U_0}$*, and an increasing optional process* $\{C_t\}_{0\le t\le T}$ *with* $C_0 = 0$*, such that*

$$U_t = U_0 + \int_0^t \xi_s dX_s - C_t, \quad \text{for all } t \in [0,T].$$

For the original papers see El Karoui and Quenez [12], Kramkov [33], Föllmer and Kramkov [16], and Föllmer and Kabanov [15].

(3.1) Arbitrage free prices and the lower Snell envelope. A superhedging strategy in the underlying market acts as total insurance of an American option. The superhedging cost is the minimal capital in order to construct such a portfolio. We are going to see that this is a price itself, in a precise sense, only if H is replicable. In the general case, it determines an upper bound for no arbitrage prices. We start by recalling the structure of arbitrage free prices, building upon the theory developed in [19].

Definition (3.9). A real number c is called an *arbitrage free price* for H if the following two conditions are satisfied.

- There exists a stopping time $\tau \in \mathcal{T}$ and a martingale measure $P \in \mathcal{M}$ such that $c \leq E_P[H_\tau]$.
- For any stopping time $\tau' \in \mathcal{T}$ there exists $P' \in \mathcal{M}$ such that $c \geq E_{P'}[H_{\tau'}]$.

The set of all arbitrage free prices for H is denoted $\Pi(H)$, and we set

$$\pi_{\inf}(H) := \inf \Pi(H) \text{ and } \pi_{\sup}(H) := \sup \Pi(H).$$

According to this definition, given $c \in \Pi(H)$, the following inequality holds

$$\sup_{\theta \in \mathcal{T}} \inf_{P \in \mathcal{M}} E_P[H_\theta] \leq c \leq \sup_{\theta \in \mathcal{T}} \sup_{P \in \mathcal{M}} E_P[H_\theta]. \tag{3.10}$$

The right-hand term equals $\pi_{\sup}(H)$ and is finite, due to our condition (3.2). We have just seen in Theorem (3.6) that $\pi_{\sup}(H) = U_0^{\uparrow}$. Thus, the upper bound is sharp and the proof makes a crucial use of the optional decomposition Theorem (3.8). The lower bound is also sharp as we quote in the next result.

THEOREM (3.11). *Assume that H is upper semicontinuous in expectation from the left with respect to any probability measure $P \in \mathcal{M}$. Then, the set of arbitrage free prices $\Pi(H)$ is an interval with infimum*

$$\pi_{\inf}(H) = \inf_{P \in \mathcal{M}} \sup_{\theta \in \mathcal{T}} E_P[H_\theta] = \sup_{\theta \in \mathcal{T}} \inf_{P \in \mathcal{M}} E_P[H_\theta] \tag{3.12}$$

and supremum

$$\pi_{\sup}(H) = \sup_{P \in \mathcal{M}} \sup_{\theta \in \mathcal{T}} E_P[H_\theta] = \sup_{\theta \in \mathcal{T}} \sup_{P \in \mathcal{M}} E_P[H_\theta]. \tag{3.13}$$

The result is proved in discrete time by [19, Theorem 6.33]. In continuous time, the characterization (3.13) has been proved in the literature above cited. The identity (3.12) is a consequence to the minimax identity:

$$\sup_{\theta \in \mathcal{T}} \inf_{P \in \mathcal{M}} E_P[H_\theta] = \inf_{P \in \mathcal{M}} \sup_{\theta \in \mathcal{T}} E_P[H_\theta]. \tag{3.14}$$

It has been established in discrete time by Föllmer and Schied [19, Theorem 6.42], and in Karatzas and Kou [32, Proposition 5.14], in a Brownian filtration. In a general model in continuous time the minimax identity has been proved by [39, Corollary 3.2]. The proof involves the following important concept.

Definition (3.15). The *lower Snell envelope* associated to H with respect to $\mathcal{M}$, is an $\mathbb{F}$-adapted process which we denote by $U^{\downarrow}$ (following the notation of [19]), such that the equality

$$U_t^{\downarrow} = \operatorname{ess\,inf}_{P \in \mathcal{M}} \operatorname{ess\,sup}_{\theta \in \mathcal{T}[t,T]} E_P[H_\theta \mid \mathcal{F}_t], \tag{3.16}$$

holds $\mathbb{P}$-a.s., for all $t \in [0, T]$.

The minimax identity (3.14) is consequence to the existence of a saddle point:

PROPOSITION (3.17). *For ρ a stopping time, the random variable*

$$\tau_\rho^\downarrow := \operatorname{ess\,inf}_{Q\in\mathcal{M}} \tau_\rho^Q, \tag{3.18}$$

is a stopping time and is optimal in the following sense

$$Z_\rho^\downarrow = \operatorname{ess\,inf}_{Q\in\mathcal{M}} E_Q[H_{\tau_\rho^\downarrow} \mid \mathcal{F}_\rho]. \tag{3.19}$$

In particular for $\tau_0^\downarrow$:

$$\inf_{Q\in\mathcal{M}} E_Q[H_{\tau_0^\downarrow}] = \sup_{\theta\in\mathcal{T}} \inf_{Q\in\mathcal{M}} E_Q[H_\theta]. \tag{3.20}$$

For a proof see [39, Proposition 3.1]. The following result characterizes the replicability of H in terms of its upper bound $\pi_{\sup}(H)$.

PROPOSITION (3.21). *Let us assume the conditions of Theorem (3.11). Then, the following conditions are equivalent*

1. *H is replicable.*
2. *There exists $P_0 \in \mathcal{M}$ such that $U_0^{P_0} = U_0^\uparrow$.*

In this case, we have that $U_0^P = U_0^\uparrow$, for arbitrary $P \in \mathcal{M}$.

REMARK (3.22). *Proposition (3.21) is an important characterization. The result was first obtained by Ansel and Stricker [2] and Jacka [28] for European options. The form presented in Proposition (3.21) appears in* [38, Proposition 1.21]. *The proof makes a crucial application of the optional decomposition theorem.*

REMARK (3.23). *An analogous characterization of replicability in terms of the lower price $\pi_{\inf}(H)$ has been recently established in discrete time by Acciaio and Svindland* [1, Theorem 2.3]. *They make a crucial application of the stopping time $\tau_0^\downarrow$. The extension to continuous time involves the Snell envelope in continuous time. A first step in this direction is the construction of a right-continuous modification; see* [39, Theorem 2.4].

(3.2) Times of maximal risk. As an application to the time consistency property we report a result on the existence of a stopping time of maximal risk as presented in [41].

Let us consider a discrete time setting with dates $t = 0,\dots,T$ and take a time consistent risk measure $\Phi = \{\rho_0,\dots,\rho_T\}$ with the Fatou property. We say that τ_t is of t-maximal risk if

$$\rho_t(-H_{\tau_t}) = U_t^\Phi := \operatorname{ess\,sup}_{\theta\in\mathcal{T}[t,T]} \rho_t(-H_\theta).$$

In the next result we show that times of maximal risk exists by making a crucial use of time consistency

THEOREM (3.24). *The upper Snell envelope satisfies the backwards representation*

$$U_t^\Phi = H_t \vee \rho_t(-U_{t+1}^\Phi). \tag{3.25}$$

The stopping time defined by

$$\tau_t := \inf\{s \geq t \mid H_s = U_s^\Phi\},$$

is of t-maximal risk.

Proof. Let θ be a stopping time with $t \leq \theta \leq T$. Then

$$\rho_t(-H_\theta) = 1_{\{\theta=t\}} H_t + 1_{\{\theta>t\}} \rho_t(-H_\theta),$$

due to the localization property [11, Proposition 1]. This identity clearly implies the following inequality

$$\rho_t(-H_\theta) \leq H_t \vee \rho_t(-U^\Phi_{t+1}).$$

Thus

$$U^\Phi_t \leq H_t \vee \rho_t(-U^\Phi_{t+1}).$$

Now we prove the converse. There exists a sequence of stopping times $\{\theta^n\}_{n=1}^\infty$ with $t+1 \leq \theta^n \leq T$ such that

$$\rho_{t+1}(-H_{\theta^n}) \to U^\Phi_{t+1}.$$

Then:

$$U^\Phi_t \geq \rho_t(-H_{\theta^n}) = \rho_t(-\rho_{t+1}(-H_{\theta^n})),$$

due to the time-consistency of the risk measure ρ. Hence

$$U^\Phi_t \geq \liminf_{n\to\infty} \rho_t(-\rho_{t+1}(-H_{\theta^n})) \geq \rho_t(-U^\Phi_{t+1}),$$

since the conditional risk measure has the Fatou property.

Clearly we have $\mathbb{P}(\tau_t \leq T) = 1$, since $U^\Phi_T = H_T$, since we are considering normalized risk measures. Note that $\tau_T = T$. By way of induction, we get

$$U^\Phi_t = H_T \vee \rho_t(-H_{\tau_{t+1}}),$$

due to the recursive formula (3.25). To conclude the proof, we show

$$H_t \vee \rho_t(-H_{\tau_{t+1}}) = \rho_t(-H_{\tau_t}).$$

The following relationships are clear

$$\tau_t = 1_{\{\tau_t=t\}} t + 1_{\{\tau_t>t\}} \tau_{t+1}.$$

Now, let us define $A := \rho_t(-H_{\tau_{t+1}})$. Then

$$H_t \vee \rho_t(-H_{\tau_{t+1}}) = H_t 1_{\{H_t \geq A\}} + A 1_{\{H_t<A\}}.$$

Moreover

$$\rho_t(-H_{\tau_t}) = H_t 1_{\{\tau_t=t\}} + \rho_t(-H_{\tau_{t+1}}) 1_{\{\tau_t>t\}}.$$

We then conclude with the set equalities

$$\{\tau_t = t\} = \{H_t = U^\Phi_t\} = \{H_t \geq \rho_t(-H_{\tau_{t+1}})\}. \qquad \square$$

Received May 30, 2013

Final version received October 17, 2013

LEONEL PÉREZ HERNÁNDEZ
DEPARTMENT OF ECONOMICS AND FINANCE
UNIVERSITY OF GUANAJUATO
lperezhernandez@yahoo.com

ERICK TREVIÑO AGUILAR
DEPARTMENT OF ECONOMICS AND FINANCE
UNIVERSITY OF GUANAJUATO
erick.trevino@ugto.org

REFERENCES

[1] B. ACCIAIO AND G. SVINDLAND, *On the lower arbitrage bound of american contingent claims*, Mathematical Finance, DOI:10.: (2012), 1–9.

[2] J. P. ANSEL AND C. STRICKER, *Couverture des actifs contingents et prix maximum*, Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques **30 (2)**, (1994), 303–315.

[3] P. ARTZNER, F. DELBAEN, J. M. EBER, AND D. HEATH, *Coherent measures of risk*, Mathematical Finance **9** (3), (1999), 203–228.

[4] A. BENSOUSSAN, *On the theory of Option Pricing*, Acta Applicandae Mathematicae **2** (2), (1984), 139–158.

[5] S. BIAGINI AND A. CERNY, *Admissible strategies in semimartingale portfolio selection*, SIAM Journal on Control and Optimization **49** (1), (2011), 42–72.

[6] S. BIAGINI AND M. FRITTELLI, *On the extension of the Namioka-Klee theorem and on the Fatou property for risk measures*, Optimality and risk: Modern Trends in Mathematical Finance, The Kabanov Festschrift, 2009.

[7] P. CHERIDITO AND T. LI, *Risk measures on Orlicz hearts*, Mathematical Finance **19** (2), (2009), 189–214.

[8] F. DELBAEN, *Coherent measures of risk on general probability spaces*, Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann, (2002), 1–37.

[9] F. DELBAEN AND W. SCHACHERMAYER, *A general version of the fundamental theorem of asset pricing*, Mathematische Annalen **300** (1), (1994), 463–520.

[10] M. DENUIT AND J. DHAENE, Actuarial Theory for Dependent Risks Measures, Orders and Models. Wiley, 2005.

[11] K. DETLEFSEN AND G. SCANDOLO, *Conditional and dynamic convex risk measures*, Finance and Stochastics **9** (4), (2005), 539–561,

[12] N. EL KAROUI AND M. C. QUENEZ, *Dynamic programming and pricing of contingent claims in an incomplete market*, SIAM Control and Optimization **33** (1), (1995), 29–66.

[13] L. EPSTEIN AND M. SCHNEIDER, *Recursive multiple priors*, Journal of Economic Theory **113** (1), (2003), 1–31.

[14] D. FILIPOVIĆ AND G. SVINDLAND, The canonical model space for law-invariant convex risk measures is L^1, Mathematical Finance, forthcoming, 2008.

[15] H. FÖLLMER AND Y. KABANOV, *Optional decomposition and Lagrange multipliers*,Finance and Stochastics **2** (1), (1998), 69–81.

[16] H. FÖLLMER AND D. O. KRAMKOV, *Optional decompositions under constraints*, Probability Theory and Related Fields **109** (1), (1997), 1–25.

[17] H. FÖLLMER AND I. PENNER, *Convex risk measures and the dynamics of their penalty functions*, Statistics and Decisions **24** (1), (2006), 61–96.

[18] H. FÖLLMER AND A. SCHIED, Stochastic Finance, An Introduction in Discrete Time, Walter de Gruyter, Berlin, 1st edition, 2002.

[19] ——, Stochastic Finance, An Introduction in Discrete Time, Walter de Gruyter, Berlin, 2nd edition, 2004.

[20] ——, *Robust representation of convex measures of risk*, Advances in Finance and Stochastics. Essays in Honor of Dieter Sondermann **6** (4), (2002), 39–56.

[21] ——, *Convex measures of risk and trading constraints*, *Finance and Stochastics* **6** (4), (2002), 429–447.

[22] M. FRITTELLI AND E. ROSAZZA GIANIN, *Putting order in risk measures*, Journal of Banking and Finance **26** (7), (2002), 1473–1486.

[23] I. GILBOA AND D. SCHMEIDLER, *Maxmin expected utility with non-unique prior*, Journal of Mathematical Economics **18** (2), (1989), 141–153.

[24] S. W. HE, J. G. WANG, AND J. A. YAN, Semimartingale theory and stochastic calculus, Beijing, Science Press, New York, 1992.

[25] D. HERNÁNDEZ-HERNÁNDEZ AND L. PÉREZ-HERNÁNDEZ, *Characterization of the minimal penalty of a convex risk measure with applications to lévy processes*, arXiv 1205.3827, 2012.

[26] ——-, *Robust utility maximization for lévy processes: Penalization and solvability*, arXiv 1206.0715, 2012.

[27] H. HULT ET AL, Risk and Portfolio Analysis. Principles and Methods, Springer, 2012.

[28] S. D. JACKA, *A martingale representation result and an application to incomplete financial markets*, Mathematical Finance **2** (4), (1992), 239–250.

[29] J. JACOD AND A. SHIRYAEV, Limit Theorems for Stochastic Processes, Springer, 2nd edition, 2003.
[30] M. KAINA AND L. RÜSCHENDORF, *On convex risk measures on L^p-spaces*, Mathematical methods of operations research **69** (3), (2008), 475–495.
[31] I. KARATZAS, *On the pricing of American options*, Applied Mathematics and Optimization **17** (1), (1988), 37–60.
[32] I. KARATZAS AND S. G. KOU, *Hedging American contingent claims with constrained portfolios*, Finance and Stochastics **2** (3), (1998), 215–258.
[33] D. O. KRAMKOV, *Optional decomposition of supermartingales and hedging contingent claims in incomplete security markets*, Probability Theory and Related Fields **105** (4), (1996), 459–479.
[34] M. KUPPER AND W. SCHACHERMAYER, *Representation results for law invariant time consistent functions*, Mathematics and Financial Economics **2** (3), (2009), 189–210.
[35] S. A. ROSS, *Options and efficiency*, The Quarterly Journal of Economics **90** (1), (1976), 75–89.
[36] W. SCHACHERMAYER, *Optimal investment in incomplete markets when wealth may become negative*, Ann. Appl. Probab. **11**, (2001), 694–734.
[37] A. SCHIED, *Optimal investments for risk- and ambiguity-averse preferences: A duality approach*, Finance and Stochastics **11** (1), (2007), 107–129.
[38] E. TREVINO-AGUILAR, American options in incomplete markets: Upper and lower Snell Envelopes and robust partial Hedging, PhD thesis, Humboldt Universität zu Berlin, 2008.
[39] ——-, *Optimal stopping under model uncertainty and the regularity of lower Snell envelopes*, Journal of Quantitative Finance **12** (6), (2011), 865–871.
[40] ——-, *Real-Valued conditional convex risk measures in L^p*, Proceedings of the X-Symposium on probability and stochastic processes, first French-Mexican meeting, ESAIM Proceedings **31**, (2011), 101–115.
[41] ——-, *Monitoring dates of maximal risk*, Arxiv.org/0902.2756, 2009.

Bol. Soc. Mat. Mexicana (3) Vol. 19, 2013

TWO APPLICATIONS OF PERMUTATION TESTS IN BIOSTASTICS

LUIS LEÓN-NOVELO, KAISA M. KEMPPAINEN, ALEXANDRIA ARDISONNE, AUSTIN DAVIS-RICHARDSON, JENNIE FAGEN, KELSEY GANO, ERIC W. TRIPLETT, AND TEDDY STUDY GROUP

ABSTRACT. We show two examples of how we answer biological questions by converting them into statistical hypothesis testing problems. We consider gene abundance data, and apply permutation tests. Though these tests are simple, they allow us to test biologically relevant hypotheses. Here we present the analysis of data rising from two studies on Type 1 Diabetes. In the first study [3] are interested in comparing the gut bacterial biodiversity in children at risk and not at risk of developing diabetes. In the second study, [4] compare the gut bacterial biodiversity of children in six different sites in USA and Europe. The statistical analyses presented here are parts of the "statistical methods" in two papers mentioned above. Here we offer a detailed explanation of the "Statistical Methods" addressed to readers with a statistics background.

1. Introduction

The main focus of this paper is to present in more detail the "statistical methodology" of two relevant studies [3] and [4] in Type 1 Diabetes (T1D). A deeper explanation of the data set and biological interpretation can be found there. Here we aim to explain the biological question and the structure of the data that it is analyzed there in. In [3], the researchers found that children who developed T1D later in life have a more different gut bacterial biodiversity than children who do not develop T1D. In [4],the gut diversity of individuals in different geographic locations is measured across time. The researchers are interested in testing if the gut bacteria diversity evolves similarly at these sites. In Sections 2 and 3 we present the first and second examples respectively. There is also a small discussion section at the end.

2. First example

The study of [3] is motivated by several animal studies suggesting that rats that develop T1D have a significantly different gut bacteria than rats that are resistant to the disease. They were interested in seeing if the same applies to human beings. The researchers wanted to determine whether the children in the control group have a gut bacteria population (microbiome) more similar to one another than the corresponding microbiomes of the case group.

2010 *Mathematics Subject Classification*: 62-07.

Keywords and phrases: analysis of covariance, permutation test, Shannon diversity index, type 1 diabetes, unifrac distance.

(2.1) Data. The data come from eight Finnish children participating in the Diabetes Prediction and Prevention study [8, 5, DIPP]. For each child, three stool samples were obtained at three time points, obtaining a total of 24 separate samples. There are four cases and four controls, the cases are children that became autoimmune and developed T1D. This is a paired design, each child in the case group was matched with a child in the control group of the same age and T1D-susceptibility genotype that did not develop autoimmunity or T1D during the study. Though the matching is used in other statistical analyses in [3], in the analysis discussed here this matching is ignored.

The stool samples of the 8 individuals provide the data for this paper. High-throughput, 16S ribosomal ribonucleic acid (16S rRNA) sequencing was performed on the stool samples. 16S rRNA gene sequencing is a widely used technology that allows the classification of bacteria. The 16S rRNA is a highly conserved gene found in all bacteria that contains hypervariable regions. The nucleic acid sequence of these regions is unique to different species of bacteria and can be used to identify bacteria. The set of 16S rRNA sequences that are amplified from a sample during high-throughput sequencing are clustered into Operational Taxonomic Units (OTUs) based on sequence similarity. Different OTUs correspond to different groups of bacteria that share a certain level of 16S rRNA sequence similarity. Every OTU has a name and the data used here are counts of OTUs in each sample. Table 2.1 shows some rows of the data set corresponding to the first and third individuals at time 1. We will use the data in this table as an example below.

First individual	Counts
First individual	
01_FS63YEP02GADJS	7
01_FS63YEP02GWM9M	9
01_FS63YEP02HBHGD	2
01_FS63YEP02HLYA0	1
01_FS63YEP02JJMWG	2
Third individual	
20_FS63YEP02FGL4W	2
20_FS63YEP02FLYRR	1
20_FS63YEP02JVBDD	1
20_FXCV9AW02H57E6	5
20_FXCV9AW02IFECK	95
20_FXCV9AW02J3H7K	1
20_FXCV9AW02JTQOI	21

Table 1. A subsample of the data corresponding to the first and third individuals at time 1.

The first row of the Table 2.1 indicates that the OTU "FS63YEP02GADJS" was observed 7 times in the first individual at time point 1. Two different OTUs correspond to two different bacteria, but two OTUs may be similar. The similarity is a measure based on the sequences and it is measured in a 0-100% scale. A similarity matrix can then be built, and with it a phylogenetic tree is built. This is a

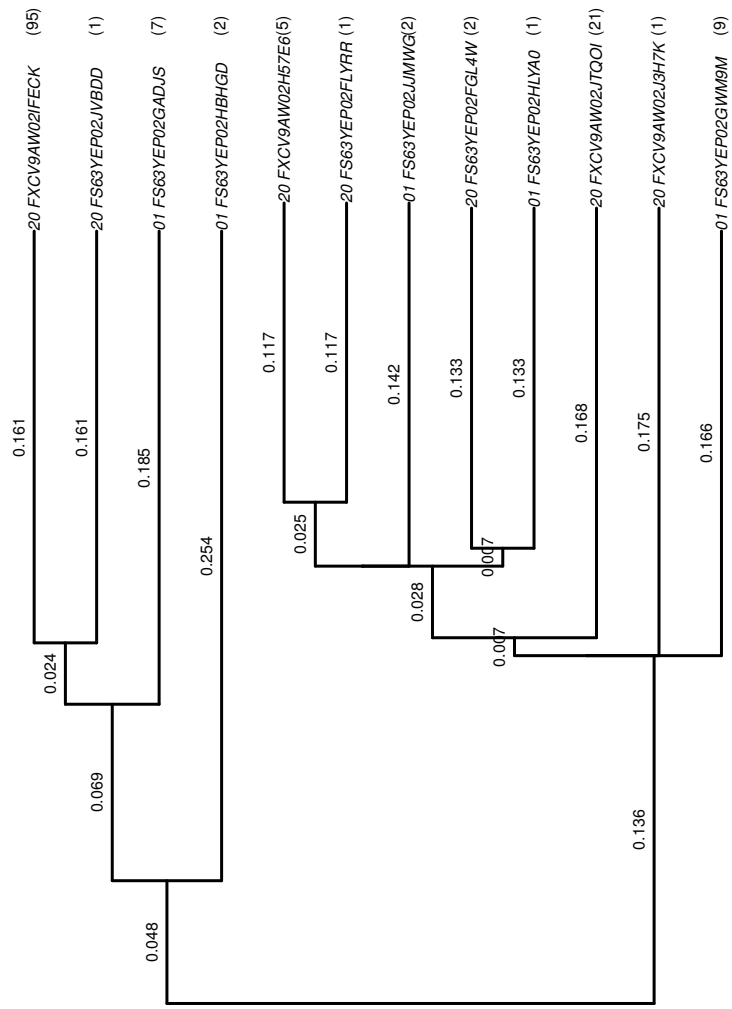


Figure 1. Phylogenetic tree corresponding to the OTUs in Table 1. The counts are given between parentheses.

weighted tree in the Theory of Graphics sense, where every leaf corresponds to an OTU. Every branch of the tree has a weight associated to it. The tree is built in such a way that the similarity between two leaves is the sum of the weight of the branches we have to pass in order to go from the leaf to the root of the smallest subtree containing both leaves. Figure 1 shows the phylogenetic tree corresponding to the OTU in Table 1. For example, the similarity between the first and third most upper leaves (labeled 20_FXCV9AW02IFECK and 01_FS63YEP02GADJS) tree is $0.161 + 0.024 = 0.185$.

(2.2) Statistical Analysis. In this subsection we explain how we used the Unifrac distance [6] and a permutation test inspired in the P-test of [7] to compare

the gut diversity of the case and control groups. In particular, we are interested in testing if the gut microbiome of the control groups are more similar to one another than the microbiomes in the case group.

The phylogenetic tree considers the similarity between OTUs but not the OTU counts. The weighted version of the UNIFRAC distance [6] incorporates the OTU count and the OTU similarity information. Using the phylogenetic tree with the OTUs of two individuals, we can compute the weighted UNIFRAC distance between the two bacteria populations. It is defined as

$$u = \sum_{i=1}^{n} b_i \left| \frac{A_i}{A_T} - \frac{B_i}{B_T} \right| \tag{2.1}$$

Here, n is the total number of branches in the tree, b_i is the length of branch i, A_i and B_i are the number of descendants of branch i from communities A and B respectively, $i = 1,\ldots,n$. A_T and B_T are the total number of sequences from communities A and B respectively. To adjust for different sample sizes, A_i and B_i are divided by A_T and B_T. In our toy example (See Figure 1), if A and B represent, respectively, the community coming from the first and third individuals, $n = 22$, $A_T = 21$, $B_T = 126$,

$$\begin{aligned} u =& 0.048 \left| \frac{9}{21} - \frac{96}{126} \right| + 0.069 \left| \frac{7}{21} - \frac{96}{126} \right| + 0.024 \left| \frac{0}{21} - \frac{96}{126} \right| \\ &+ 0.161 \left| \frac{0}{21} - \frac{95}{126} \right| + \cdots + 0.166 \left| \frac{9}{21} - \frac{0}{126} \right| \\ =& 0.45 \end{aligned}$$

The more different the two populations are, the larger the unifrac distance is. [7] proposes a permutation test called "P test" to determine if two populations are significantly (*i.e.*, statistically) different. Under the null hypothesis, *i.e.* under the assumption that both bacteria populations are equal, the population labels of the OTUs are exchangeable. The test consists in permuting the population labels in the phylogenetic tree, with the permuted labels, compute $u_\star$ according to (2.1). We repeat the process M times to obtain $u_\star^{(1)},\ldots,u_\star^{(M)}$. The p-value is the proportion of times that $u > u_\star$, in math, p-value$= \sum_{m=1}^{M} 1(u > u_\star^{(m)})/M$, where $1(A)$ is the indicator function of the event A. Rejecting the null hypothesis is claiming that the bacteria populations are different. In other words, the difference between the OTU counts are not due to random chance. It is worth it to mention that the Unifrac distance measures the difference in the diversity of two populations. This is, a large unifrac distance indicated that one population is more diverse than the other. Two populations can be equally diverse but completely different. This is, two samples or populations could be statistically similar even though they might not contain any common bacteria.

As mentioned earlier, [3] were interested in knowing if the controls have a more similar gut bacteria population to one another than the case individuals. To test so, in [3], we performed the following permutation test, inspired in the P-test:

1. Denote with a_i, o_j the case i and control j individuals respectively, $i,j = 1,2,3,4$; and $u(\square,\circ)$ the unifrac distance between the individuals $\square$ and $\circ$. For each one of the six possible pairs of individuals in the case group we

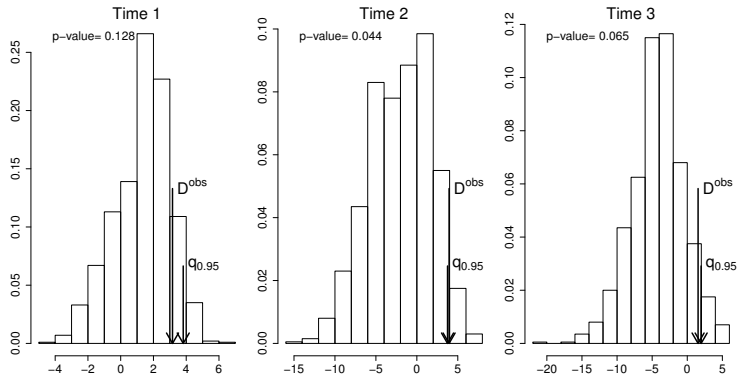


Figure 2. Histogram of simulated $D^\star$s along with the 95% percentile, D^{obs} and p-value for each one of the times in the study. See

compute the unifrac distances and sum them to obtain

$$S_{\mathrm{a}}^{obs} = u(\mathrm{a}_1,\mathrm{a}_2)+u(\mathrm{a}_1,\mathrm{a}_3)+u(\mathrm{a}_1,\mathrm{a}_4)+u(\mathrm{a}_2,\mathrm{a}_3)+u(\mathrm{a}_2,\mathrm{a}_4)+u(\mathrm{a}_3,\mathrm{a}_4)$$

Similarly, we define S_{o}^{obs} as the sum of the unifrac distances of each of the six possible pairs of individuals in the control group. Our test statistic is

$$D^{obs} = S_{\mathrm{a}}^{obs} - S_{\mathrm{o}}^{obs}.$$

2. For each phylogenetic tree, we randomly permute the labels of the individuals (children) and, as we computed D^{obs} but now considering these permuted-label trees, we compute $D^\star$.
3. We repeat step 2 a total of $M = 10^5$ times to get a sample of differences, *i.e.*, of $D^\star$: $D_1^\star,\ldots,D_M^\star$
4. We compute the p-value as the proportion of $D^\star$s greater than D^{obs}. In math,

$$p-\text{value} = (1/M)\sum_{m=1}^{M} 1(D^{obs} > D_m^\star).$$

Since the Unifrac distance is a measurement of how far apart two microbiomes (*i.e.*, bacteria populations) of two individuals are, a large value of D^{obs} would suggests that the case microbiomes are more different to one another than the control microbiomes. Equivalently, D^{obs} large is evidence that the control microbiomes are more similar to each other than the case microbiomes. The question becomes now what large means. If all the populations were equal, the population labels in every phylogenetic tree would be exchangeable. Following the idea of the P-test, simulated samples from the null distribution of the statistic D are obtained by permuting the population labels in every phylogenetic tree.

Figure 2 shows the histogram of the $D^\star$s for each time. The short and long arrows indicate the 95% quantile and D^{obs} respectively. The long arrow is at the right of the 0.95 quatile (short arrow) just at time 2. That is, we are able to claim that at time 2 the population of microrganisms in the control group are more similar to one another than in the case group. The data suggest that the same is true at time 1 and 3 but is not conclusive.

3. Second Example

The second example is part of the statistical analysis in the paper [4]. Inspired by the findings of [3] and others, researchers suspect that the gut microbiome has a role in the development of T1D. The composition of the gut microbiome of children in six different locations was analyzed across time. The sampling units are children genetically at high risk for T1D but currently free of islet autoantibodies or disease. The TEDDY group gathered data about the the composition of the gut microbiome of children across time in six different locations. The sampling units are children genetically at high risk for type 1 diabetes but currently free of islet autoantibodies or disease. In the current manuscript, we explain and reproduce the statistical analysis that yields them to conclude that the microbiome diversity across time differs in the six study sites.

(3.1) Data. [4] analyzed stool samples taken monthly from children starting at age four months old until they turned 19 months old. The data correspond to 90 children, 15 from each one of the six different participating sites: Finland, Germany, Sweden, Washington state, Colorado, and Georgia/Florida. As in the example in Section 2, high-throughput 16S rRNA sequencing was performed on these stool samples. The data consists of a table of genus-level OTU counts for every stool sample (not shown). Since we are interested in the bacterial diversity the biologists work with the Shannon Diversity Index [10, SDI],

$$\text{SDI} = \sum_{i=1}^{R} p_i \log p_i,$$

where i indexes the different OTUs in the sample, p_i is the proportion of OTUs i in the sample, and R is the total number of different OTUs in the sample. The more diverse the bacteria population is, the larger the SDI is. For our purposes the data were reduced to a sequence of SDI measurements across different time points for every child. These sequences are shown in Figure 3. Every line represents the SDI of a child across time. Visually, we cannot appreciate any clear difference among the SDI curves across the sites, except, probably, Sweden where the SDI seems to have less variance. [4] speculate that the reason for this may be that the Sweden children are the least exposed to antibiotics of all the sites in the study. Since there are few stool samples for the youngest and oldest ages, we have removed from this analysis the data corresponding to ages under 100 days and over 550 days.

(3.2) Statistical Analysis. The aim of this statistical analysis is to test if the curves of the SDI are statistically different or not. In order to do so we need to introduce a statistical model. We consider the following mixed model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma t_k + \delta_i t_k + \eta t_k^2 + \epsilon_{ijk}, \tag{3.1}$$

where, y_{ijk} represents the k-th measurement of the SDI for child j at site i, μ is the over all mean, α_i is the fixed site effect (for estimation purposes we impose $\sum_i \alpha_i = 0$), $\beta_j \sim N(0, \sigma^2_{child})$ is the child-specific random effect, t_k is the child age in days (treated as a continuous variable standardized to have sample mean and variance equal to 0 and 1 respectively) when the k-measurement was taken, the fixed effect δ_i is the interaction coefficient between days and site (also assuming $\sum_i \delta_i = 0$), η is also a fixed effect, and $\epsilon_{ijk} \sim N(0, \sigma^2_\epsilon)$ is a random error. In the

context of the model, testing if the SDI curves are statistically significant reduces to test

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_6 = 0 \quad vs \quad H_1 : \sum_i \alpha_i^2 > 0 \tag{3.2}$$

Strictly speaking, in order to test if the curves are the same for all sites, we should not only test that all α_i are zero but also that all δ_is are equal to zero. Nevertheless, rejecting the null hypothesis as stated above would make us conclude not all curves are equal.

We can think of the model in (3.1) as an Analysis of covariance ANCOVA [2, See p. 62 on] where the covariate is time (in days). Fitting the model in a statistical package is straightforward. We used the function "aov" in the lme4 R package [9] by [1]. The R code is,

```
> model=aov(Shannon \sim Site+Error(1/Patient)+Time+Time:Site+Time2)
```

where $Shannon$ is the SDI, $Site$ takes one of the six possible locations, $Time$ is the standardized time in days and $Time2 = Time^2$ fits the model. An F test of (3.2) is straightforward, (R code `>summary(model))`

	Df	Sum Sq	Mean Sq	F value	Pr(> F)
Site	5	30.98	6.20	25.58	$< 2 \times 10^{-16}$
Time	1	20.42	20.42	84.29	$< 2 \times 10^{-16}$
Time2	1	0.00	0.00	0.01	0.9217
Site:Time	5	2.93	0.59	2.42	0.0341
Residuals	1099	266.17	0.24		

As one can see Site is highly significant and we reject the null in (3.2). We notice that the site time interaction ($Site : Time$) is significant. One of the assumptions of ANCOVA is that the "covariate is not related with the treatment" [2, See p. 63 on]. This is not the case if there is an interaction between Site and treatment. Additionally, the Kolmogorov-Smirnov test rejects (not shown) the normality of the residuals. The ANCOVA is known to be robust against the normality assumption of the errors as long as the symmetry of the errors follows. We test the symmetry of the residuals following [11], explained below. Let $\mathbf{x} = x_1, \ldots, x_n$ denote the sample of size n from a distribution with mean μ, median ν and standard deviation σ. Let $\bar{x}, M$ and s denote the sample mean, median and standard deviation. They propose a Bootstrap test of symmetry consisting in

1. Define $C_{obs} = (\bar{x} - M)/s$, the sample version of the measure of skewness $(\mu - \nu)/\sigma$
2. Define the symmetrized empirical distribution $\hat{F}^s$ as the CDF that gives $1/(2n)$ to all possible values in the sample and to all points in $\{2M - x_1, \ldots, 2M - x_n\}$
3. Obtain T bootstrap samples $\hat{F}^s$ of size n. For each bootstrap sample $\mathbf{x}^\star$ compute its sample mean, median and standard deviation $\bar{\mathbf{x}}^\star, M^\star, s^\star$ and obtain $C^\star = (\bar{\mathbf{x}}^\star - M^\star)/s^\star$
4. The bootstrap p-value is then the proportion of $|C^\star|$s greater that $|C_{obs}|$. Here $|x|$ denotes the absolute value of x.

This test applied to the residuals (with $T = 10^6$ boot strap samples in step 3) yields a boot-strap p-value of $< 10^{-6}$. The residuals and then the error terms are not symmetric.

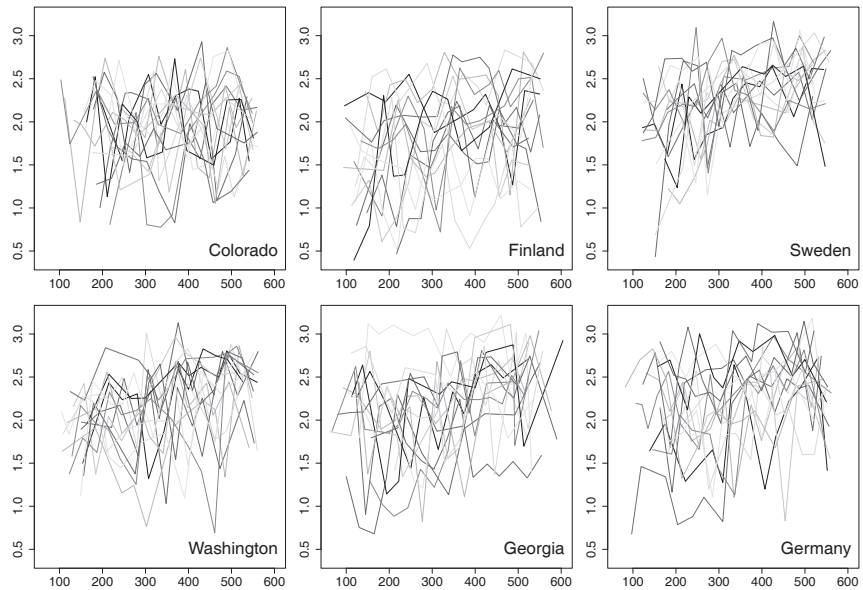


Figure 3. Shannon Diversity Index per child through time (days) in the six different study sites. Every line represents the measurement of the Shannon Diversity Index of a child across time. For visual purposes, the line joints the time/SDI points of the child it represens.

To deal with the violation of the assumptions of the ANCOVA model, we apply a permutation test. Large values of F indicate that the null hypothesis is false. We apply the following simple permutation test permuting the labels of Site.

1. Compute F_{obs}, the F statistic in the ANCOVA table testing for the model in (3.1) testing (3.2).
2. Permute the "Site" label, and compute the F statistic T times to obtain a sample $F^{\star(1)},\ldots,F^{\star(M)}$
3. The p-value of this permutation test is the percentage of $F^\star$s greater than F_{obs}

Applying this permutation test (with $T = 10^5$) we obtain a p-value $< 10^{-5}$. The data provide enough evidence to reject the null hypothesis in (3.2). This is, the SDI curves are not all equal in the six sites.

4. Discussion

Through this paper we have shown two examples of the application of simple permutation tests to answer relevant biological questions. These examples are the product of joint work with researchers at the University of Florida and are part of the "Statistics Methods" section of two biology papers. The main merit of these analyses is the collaboration between biologists and statisticians to formulate the biological problem in statistical terms. In the first example a permutation test allowed us to answer a relevant question without the need to depend on model assumptions. In the second example, a model is required. A standard F-test is applied to the parameters of the model. If the data followed the assumptions

of this F-test (normality or at least symmetry of the random errors), the F-test would be valid. This is not the case, we obtain evidence against the symmetry of the distribution of the random errors trough a boot-strap test of symmetry applied to the residuals of the model. The F-test is not valid. Nevertheless, we are able to take advantage of this F-test by incorporating it into the scheme of a permutation test. With this we avoid more complicated models in order to get an answer to the biological question.

Acknowledgments

Luis León Novelo was supported by NIH 1R01GM081704. Luis León Novelo is thankful to Professor George Casella for his work and advise in the statistical analysis presented here.

TEDDY Study Acknowledgments

The TEDDY Study Group (See appendix)

Funded by U01 DK63829, U01 DK63861, U01 DK63821, U01 DK63865, U01 DK63863, U01 DK63836, U01 DK63790, UC4 DK63829, UC4 DK63861, UC4 DK63821, UC4 DK63865, UC4 DK63863, UC4 DK63836, and UC4 DK95300 and Contract No. HHSN267200700014C from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Institute of Child Health and Human Development (NICHD), National Institute of Environmental Health Sciences (NIEHS), Juvenile Diabetes Research Foundation (JDRF), and Centers for Disease Control and Prevention (CDC).

Appendix

The Teddy Study Group

Colorado Clinical Center: Marian Rewers, M.D., Ph.D., PI[1,4,6,10,11], Katherine Barriga[12], Kimberly Bautista[12], Judith Baxter[9,12,15], George Eisenbarth, M.D., Ph.D., Nicole Frank[2], Patricia Gesualdo[2,6,12,14,15], Michelle Hoffman[12,13,14], Lisa Ide, Rachel Karban12, Edwin Liu, M.D.[13], Jill Norris, Ph.D.[2,3,12], Kathleen Waugh[6,7,12,15], Adela Samper-Imaz, Andrea Steck, M.D.3, University of Colorado, Anschutz Medical Campus, Barbara Davis Center for Childhood Diabetes.

Georgia/Florida Clinical Center: Jin-Xiong She, Ph.D.,PI[1,3,4,11,†], Desmond Schatz, M.D.[*4,5,7,8], Diane Hopkins[12], Leigh Steed[12,13,14,15], Jamie Thomas[*6,12], Katherine Silvis[2], Michael Haller, M.D.[*14], Meena Shankar[*2], Melissa Gardiner, Richard McIndoe, Ph.D., Haitao Liu, M.D.[†], John Nechtman[†], Ashok Sharma, Joshua Williams, Gabriela Foghis, Stephen W. Anderson, M.D.[^] Medical College of Georgia, Georgia Regents University, ∗University of Florida, †Jinfiniti Biosciences LLC, Augusta, GA, ∧Pediatric Endocrine Associates, Atlanta, GA.

Germany Clinical Center: Anette G. Ziegler M.D.,PI[1,3,4,11], Andreas Beyerlein Ph.D.[2], Ezio Bonifacio Ph.D.[*5], Lydia Henneberger[2,12], Michael Hummel M.D.[13], Sandra Hummel Ph.D.[2], Kristina Foterek[ϒ2], Mathilde Kersting Ph.D.[ϒ2], Annette Knopff[7], Sibylle Koletzko, M.D.[¶13], Stephanie Krause, Claudia Peplow[12], Maren Pflüger Ph.D.[6], Roswith Roth Ph.D.[9], Julia Schenkel[2,12], Joanna Stock[9,12], Elisabeth Strauss[12], Katharina Warncke M.D.[14], Christiane Winkler Ph.D.[2,12,15], Forschergruppe Diabetes e.V. at Helmholtz Zentrum München, ∗ Center for

Regenerative Therapies, TU Dresden, ¶Dr. von Hauner Children's Hospital,Department of Gastroenterology, Ludwig Maximillians University Munich, ϒ Research Institute for Child Nutrition, Dortmund.

Finland Clinical Center: Olli G. Simell, M.D., Ph.D., PI[ϒ∧1,4,11,13], Heikki Hyöty, M.D., Ph.D.[*±6], Jorma Ilonen, M.D., Ph.D.[ϒ¶3], Mikael Knip, M.D., Ph.D.[*±], Maria LŽnnrot, M.D., Ph.D.[*±6], Elina Mantymaki[ϒ∧], Juha Mykkänen, Ph.D.[∧ϒ3], Kirsti Nanto-Salonen, M.D., Ph.D.[ϒ∧12], Tiina Niininen[±*12], Mia Nyblom[*±], Anne Riikonen[*±2], Minna Romo[ϒ∧], Barbara Simell[ϒ∧9,12,15], Tuula Simell, Ph.D.[ϒ∧9,12], Ville Simell[∧ϒ13], Maija SjŽberg[ϒ∧12,14], Aino Stenius[μ□12], Jorma Toppari, M.D., Ph.D., Eeva Varjonen[ϒ∧12], Riitta Veijola, M.D., Ph.D. [μ□14], Suvi M. Virtanen, M.D., Ph.D.[*±§2]. ϒUniversity of Turku, ∗University of Tampere, μUniversity of Oulu, ∧Turku University Hospital, ± Tampere University Hospital, □Oulu University Hospital, §National Institute for Health and Welfare, Finland, ¶University of Kuopio.

Sweden Clinical Center: Åke Lernmark, Ph.D., PI [1,3,4,5,6,8,10,11,15], Daniel Agardh, M.D., Ph.D.[13], Carin Andrén-Aronsson[2,13], Maria Ask, Jenny Bremer, Corrado Cilio Ph.D., M.D.[5], Emilie Ericson-Hallström[2], Lina Fransson, Thomas Gard, Joanna Gerardsson, Gertie Hansson[12,14], Monica Hansen, Susanne Hyberg, Fredrik Johansen, Berglind Jonasdottir M.D., Ulla-Marie Karlsson, Helena Larsson M.D., Ph.D. [6,14], Barbro Lernmark, Ph.D.[9,12], Maria Markan, Theodosia Massadakis, Jessica Melin[12], Maria Månsson-Martinez, Anita Nilsson, Kobra Rahmati, Falastin Salami, Monica Sedig Järvirova, Sara Sibthorpe, Birgitta Sjöberg, Ulrica Swartling, Ph.D. [9,12], Erika Trulsson, Carina Törn, Ph.D. [3,15], Anne Wallin, Åsa Wimar[12], Sofie Åberg. Lund University.

Washington Clinical Center: William A. Hagopian, M.D., Ph.D., PI[1,3,4,5,6,7,11,13,14], Xiang Yan, M.D., Michael Killian[6,7,12,13], Claire Cowen Crouch[12,14,15], Kristen M. Hay[2], Stephen Ayres, Carissa Adams, Brandi Bratrude, David Coughlin, Greer Fowler, Czarina Franco, Carla Hammar, Diana Heaney, Patrick Marcus, Arlene Meyer, Denise Mulenga, Elizabeth Scott, Jennifer Skidmore[2], Joshua Stabbert, Viktoria Stepitova, Nancy Williams. Pacific Northwest Diabetes Research Institute.

Pennsylvania Satellite Center: Dorothy Becker, M.D., Margaret Franciscus[12], MaryEllen Dalmagro-Elias[2], Ashi Daftary, M.D. ChildrenÕs Hospital of Pittsburgh of UPMC.

Data Coordinating Center: Jeffrey P. Krischer, Ph.D.,PI [1,4,5,10,11], Michael Abbondondolo, Sarah Austin, Rasheedah Brown[12,15], Brant Burkhardt, Ph.D.[5,6], Martha Butterworth[2], David Cuthbertson, Christopher Eberhard, Steven Fiske[9], Veena Gowda, David Hadley, Ph.D.[3,13], Page Lane, Hye-Seung Lee, Ph.D.[3,6,13,15], Shu Liu, Xiang Liu, Ph.D.[2,12], Kristian Lynch, Ph.D.[6,9], Jamie Malloy, Cristina McCarthy[12,15], Wendy McLeod[2,5,6,13,15], Laura Smith, Ph.D.[9,12], Susan Smith[12,15], Roy Tamura, Ph.D.[2], Ulla Uusitalo, Ph.D.[2,15], Kendra Vehik, Ph.D. [4,5,9,14,15], Earnest Washington, Jimin Yang, Ph.D., R.D.[2,15]. University of South Florida.

Project scientist: Beena Akolkar, Ph.D.[1,3,4,5,6,7,10,11], National Institutes of Diabetes and Digestive and Kidney Diseases.

Other contributors: Kasia Bourcier, Ph.D.[5], National Institutes of Allergy and Infectious Diseases. Thomas Briese, Ph.D.[6,15], Columbia University, Suzanne

Bennett Johnson, Ph.D.[9,12], Florida State University, Steve Oberste, Ph.D.[6], Centers for Disease Control and Prevention, Eric Triplett, Ph.D. [6], University of Florida.

Autoantibody Reference Laboratories: Liping Yu, M.D.[5], Dongmei Miao, M.D.^, Polly Bingley, M.D., FRCP*[5], Alistair Williams*, Kyla Chandler*, Saba Rokni*, Anna Long Ph.D.*, Joanna Boldison*, Jacob Butterly*, Jessica Broadhurst*, Gabriella Carreno*, Rachel Curnock*, Peter Easton*, Ivey Geoghan*, Julia Goode*, James Pearson*, Charles Reed*, Sophie Ridewood*, Rebecca Wyatt*. ∧Barbara Davis Center for Childhood Diabetes, University of Colorado Denver, ∗School of Clinical Sciences, University of Bristol UK.

Cortisol Laboratory: Elisabeth Aardal Eriksson, M.D., Ph.D., Ewa Lönn Karlsson. Department of Clinical Chemistry, Linköping University Hospital, Linköping, Sweden.

Dietary Biomarkers Laboratory: Iris Erlund, Ph.D.[2], Irma Salminen, Jouko Sundvall, Jaana Leiviskä, Mari Lehtonen, Ph.D. National Institute for Health and Welfare, Helsinki, Finland.

HbA1c Laboratory: Randie R. Little, Ph.D., Alethea L. Tennill. Diabetes Diagnostic Laboratory, Dept. of Pathology, University of Missouri School of Medicine.

HLA Reference Laboratory: Henry Erlich, Ph.D.[3], Teodorica Bugawan, Maria Alejandrino. Department of Human Genetics, Roche Molecular Systems.

Metabolomics Laboratory: Oliver Fiehn, Ph.D., Bill Wikoff, Ph.D., Tobias Kind, Ph.D., Mine Palazoglu, Joyce Wong, Gert Wohlgemuth. UC Davis Metabolomics Center.

Microbiome and Viral Metagenomics Laboratory: Joseph F. Petrosino, Ph.D.[6] Alkek Center for Metagenomics and Microbiome Research, Department of Molecular Virology and Microbiology, Baylor College of Medicine.

OGTT Laboratory: Santica M. Marcovina, Ph.D., Sc.D. Northwest Lipid Metabolism and Diabetes Research Laboratories, University of Washington.

Repository: Heather Higgins, Sandra Ke. NIDDK Biosample Repository at Fisher BioServices.

RNA Laboratory and Gene Expression Laboratory: Jin-Xiong She, Ph.D.,PI[1,3,4,11], Richard McIndoe, Ph.D., Haitao Liu, M.D., John Nechtman, Yansheng Zhao, Na Jiang, M.D. Jinfiniti Biosciences, LLC.

SNP Laboratory: Stephen S. Rich, Ph.D.3, Wei-Min Chen, Ph.D.[3], Suna Onengut-Gumuscu, Ph.D.[3], Emily Farber, Rebecca Roche Pickin, Ph.D., Jordan Davis, Dan Gallo. Center for Public Health Genomics, University of Virginia.

Committees: [1]Ancillary Studies,[2]Diet, [3]Genetics, [4]Human Subjects/Publicity/Publications, [5]Immune Markers, [6]Infectious Agents, [7]Laboratory Implementation, [8]Maternal Studies, [9]Psychosocial, [10]Quality Assurance, [11]Steering, [12]Study Coordinators, [13]Celiac Disease, [14]Clinical Implementation, [15]Quality Assurance Subcommittee on Data Quality.

Received May 30, 2013

Final version received October 24, 2013

LUIS LEÓN-NOVELO,
DEPARTMENT OF MATHEMATICS,
UNIVERSITY OF LOUISIANA AT *Lafayette*,
LAFAYETTE, LA, USA
luis@louisina.edu

KAISA M. KEMPPAINEN, kkemppainen@ufl.edu
ALEXANDRIA ARDISSONE, lexi88@ufl.edu
AUSTIN DAVIS-RICHARDSON, adavisr@ufl.edu
JENNIE FAGEN, jenn1eruth@ufl.edu
KELSEY GANO, ganokelsey@gmail.com
ERIC TRIPLETT, ewt@ufl.edu
TEDDY PUBLICATIONS PUBLICATIONS, TeddyPublications@epi.usf.edu
DEPARTMENT OF MICROBIOLOGY AND CELL SCIENCE,
INSTITUTE OF FOOD AND AGRICULTURAL SCIENCES,
UNIVERSITY OF FLORIDA.
GAINESVILLE, FL, USA.

REFERENCES

[1] D. BATES AND M. MAECHLER, *lme4: Linear mixed-effects models using S4 classes*, 2010. R package version 0.999375-37.

[2] G. CASELLA, *Statistical design*, Springer New York, 2008.

[3] A. GIONGO, K. A. GANO, D. B. CRABB, N. MUKHERJEE, L. L. NOVELO, G. CASELLA, J. C. DREW, J. ILONEN, M. KNIP, H. HYÖTY, ET AL., *Toward defining the autoimmune microbiome for type 1 diabetes*, The ISME journal, **5** (1), (2010), 82–91.

[4] K. M. KEMPPAINEN, A. ARDISSONE, A. DAVIS-RICHARDSON, J. FAGEN, K. GANO, L. LEON-NOVELO, G. CASELLA, O. SIMELL, A. G. ZIEGLER, M. J. REWERS, A. LENMARK, W. HAGOPIAN, J. SHE, J. P. KRISCHER, B. AKOLKAR, D. SCHATZ, M. ATKINSON, E. W. TRIPLETT, AND THE TEDDY STUDY GROUP, *An analysis of gut microbial diversity of non-autoimmune subjects at high-risk for type 1 diabetes*, Submitted, , (2013).

[5] A. KUPILA, P. MUONA, T. SIMELL, P. ARVILOMMI, H. SAVOLAINEN, A.-M. HÄMÄLÄINEN, S. KORHONEN, T. KIMPIMÄKI, M. SJÖROOS, J. ILONEN, ET AL., *Feasibility of genetic and immunological prediction of type i diabetes in a population-based birth cohort*, Diabetologia, **44** (3), (2001), 290–297.

[6] C. LOZUPONE AND R. KNIGHT, *Unifrac: a new phylogenetic method for comparing microbial communities*, Applied and environmental microbiology, **71** (12), (2005), 8228–8235.

[7] A. P. MARTIN, *Phylogenetic approaches for describing and comparing the diversity of microbial communities*, Applied and environmental microbiology, **68** (8), (2002), 3673–3682.

[8] S. NEJENTSEV, M. SJÖROOS, T. SOUKKA, M. KNIP, O. SIMELL, T. LÖVGREN, AND J. ILONEN, *Population-based genetic screening for the estimation of type 1 diabetes mellitus risk in finland: selective genotyping of markers in the hla-dqb1, hla-dqa1 and hla-drb1 loci*, Diabetic medicine, **16** (12), (1999), 985–992.

[9] R DEVELOPMENT CORE TEAM, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2010. ISBN 3-900051-07-0.

[10] C. E. SHANNON AND W. WEAVER, *A mathematical theory of communication*, 1948.

[11] T. ZHENG AND J. L. GASTWIRTH, *On bootstrap tests of symmetry about an unknown median*, Journal of data science: JDS, **8** (3), (2010), 413.

Bol. Soc. Mat. Mexicana (3) Vol. 19, 2013

LÉVY-DRIVEN PROCESSES IN BAYESIAN NONPARAMETRIC INFERENCE

LUIS E. NIETO-BARAJAS

ABSTRACT. In this article we highlight the important role that Lévy processes have played in the definition of priors for Bayesian nonparametric inference. We review some of the main properties and characterization of Lévy processes, and show several applications where these processes have demonstrated good properties, when they are used as nonparametric prior distributions for Bayesian inference. These applications include: inference for mixed Poisson processes; survival analysis; and density estimation.

1. Introduction

Statistical inference is the process of making decisions about a phenomenon that is subject to variability. Decisions are based on a series of realizations of the phenomenon, known as sample. Inference procedures can be classified into two paradigms: the frequentist, which makes exclusive use of the data, usually maximizing the probability of having observed them under certain model assumptions; and the Bayesian, which treats all decisions under the axiomatic decision theory and quantifies all knowledge and uncertainty of the phenomenon through probability distributions. Within this Bayesian paradigm, it is also possible to incorporate prior knowledge about the phenomenon in the decision making process.

Regardless of the paradigm you choose to make decisions, there are a series of assumptions you need to make in order to characterize the phenomenon of interest. These are called model assumptions. If the number of parameters used to describe the model is finite then we say that the model is parametric. On the other hand, if the number of parameters is infinite the model is termed nonparametric.

Bayesian nonparametric theory handles statistical inference by assuming a nonparametric model and making decisions via the Bayesian paradigm. Since the Bayesian decision theory requires to express prior knowledge on the unknown quantities, a Bayesian nonparametric prior is a probability measure on an infinite dimensional space. What makes a prior to be nonparametric was clearly stated by [8] in his Annals of Statistics paper, "a nonparametric prior must have large support in the sense that any fixed probability measure can be arbitrarily approximated by a probability measure generated by the prior."

In notation, let $X_1,\ldots,X_n$ be a sample of random variables (r.v.) such that, conditionally on a fixed cumulative distribution function (c.d.f.) F defined on $(\mathcal{X},\mathcal{B})$, the r.v.'s are independent and identically distributed (i.i.d.), that is, $X_i|F \overset{\text{iid}}{\sim} F$,

2010 *Mathematics Subject Classification*: 62F15, 62G07, 62N01.

Keywords and phrases: Bayesian nonparametrics, density estimation, increasing additive process, Markov process, mixed Poisson process, survival analysis.

where $\mathcal{X} \subset \mathbb{R}$ is the sample space and $\mathcal{B}$ the Borel's σ-algebra. Being nonparametric means that the law F that describes the behaviour of the X_i's is all unknown. To place our prior knowledge about F, we rely on stochastic processes whose paths are c.d.f.'s. In notation, $F \sim \mathcal{P}$, where $\mathcal{P}$, defined on $(\mathcal{F}, \mathcal{A})$, is the nonparametric prior, with $\mathcal{F}$ the set of all c.d.f.'s and $\mathcal{A}$ an appropriate σ-albebra of subsets of $\mathcal{F}$. If we think on the probability measure induced by F, then we say that this is a random probability measure.

In this paper we will review some Bayesian nonparametric priors which are Lévy-driven. For this we mean priors that are based on a Lévy or increasing additive process. In Section 2 we define Lévy processes and present some of their properties. In Section 3 we describe different nonparametric priors that can be defined in terms of Lévy processes. We briefly discuss posterior inference in Section 4 and show some illustrations in Section 5.

Before we proceed we introduce notation: $\text{Ga}(\alpha, \beta)$ denotes the gamma density with mean α/β; $\text{Po}(c)$ is a Poisson density with mean c; $\text{N}(\mu, \sigma^2)$ is a normal density with mean μ and variance σ^2.

2. Lévy processes

A stochastic process can be thought of as a family of random variables linked via a parameter which takes values on a specific domain. According to [7], a stochastic process is the mathematical abstraction of an empirical process whose development is governed by probabilistic laws. Let $\{Z(x); x \in \mathcal{X}\}$ denote a stochastic process with domain $\mathcal{X}$ and range or state space $\mathcal{Z}$. For simplicity let us assume that $\mathcal{X} \subset \mathbb{R}$ and $\mathcal{Z} \subset \mathbb{R}$.

An independent increments process, or additive process, satisfies the condition that for $x_1 < x_2 < \cdots < x_k \in \mathcal{X}$, the increments defined as $Z(x_1), Z(x_2) - Z(x_1), \ldots, Z(x_n) - Z(x_{n-1})$ are independent. The two principal members of this class are the Wiener and Poisson processes. According to, e.g., [10], every stochastically continuous process with independent increments can be represented as the sum of a Wiener process and an integral of Poisson processes.

A stochastic process $Z(x)$ with independent increments is said to be homogeneous if the increments are stationary, that is, if the distribution of $Z(x+t) - Z(x)$ only depends on t. Lévy processes are homogeneous independent increments processes, and are usually referred as random walks in continuous time.

Lévy processes can have non-monotonic paths, however, the Lévy processes used for defining priors in Bayesian nonparametric inference are bound to be nondecreasing, nonnegative and with piecewise constant paths. They are known as pure jump Lévy processes. Moreover, the homogeneity constraint that characterises a Lévy process is relaxed to nonhomogeneous cases in the definition of nonparametric priors. Strictly speaking such processes are not Lévy anymore and a more appropriate name would be increasing additive processes, but we will still refer to them as nonhomogeneous Lévy processes.

The probability law of a pure jump (homogeneous or nonhomogeneous) Lévy process is characterized by its Laplace transform and is given by

$$\text{E}\left\{e^{-\phi Z(x)}\right\} = \exp\left\{-\int_{-\infty}^{x}\int_{0}^{\infty}\left(1 - e^{-\phi v}\right)\nu(\text{d}v, \text{d}s)\right\}, \tag{2.1}$$

where $\nu(dv,ds) = \rho(dv|s)\alpha(ds)$ is called the Lévy intensity, $\rho(\cdot|s)$ is a measure on $\mathbb{R}^+$ that controls the jump sizes for every location s, and $\alpha(\cdot)$ is a measure on $\mathbb{R}$ that determines the jump locations. The Lévy intensity must satisfy the condition

$$\int_A \int_0^\infty \min\{v,1\}\nu(dv,ds) < \infty,$$

for any bounded $A \subset \mathcal{X}$. If the measure ρ is independent of the locations s, i.e., $\rho(dv|s) = \rho(dv)$, the process $Z(x)$ is homogeneous.

There are several Lévy intensities that satisfy the previous condition, most of them can be seen as particular cases of the following two nonhomogeneous Lévy intensities:

(i) Generalized gamma: $\rho(dv|s) = \Gamma(1-\epsilon)^{-1}v^{-(1+\epsilon)}e^{-\beta(s)v}dv$, with $\epsilon \in \{(0,1) \cup \{-1\}\}$, and
(ii) Log-beta: $\rho(dv|s) = (1-e^{-v})^{-1}e^{-\beta(s)v}dv$,

with a non-homogeneous parameter function $\beta(s) \geq 0$ for all $s \in \mathbb{R}$, together with a measure $\alpha(s)$ on $\mathbb{R}$. In particular, for case (i) and with $\epsilon = -1$, the Lévy measure becomes finite, i.e., the number of jumps in a finite interval is finite, whereas infinite Lévy measures have an infinite number of jumps in a finite interval.

A Lévy process can be generalized to include fixed jump locations $\chi_1, \chi_2, \ldots$, with independent nonnegative jump sizes $Z\{\chi_1\}, Z\{\chi_2\}, \ldots$ (also independent of the rest of the process), then a general Lévy process becomes

$$Z(x) = Z_c(x) + \sum_j Z\{\chi_j\}I(\chi_j \leq x),$$

where, $Z\{x\} = Z(x) - Z(x-)$ and $Z_c(x)$ is a Lévy process without fixed points of discontinuity, also known as "continuous" part, whose law is characterized by (2.1). Although $Z_c(x)$ is called "continuous", $Z_c(x)$ is almost surely discrete, so it can be represented as $Z_c(x) = \sum_j J_j^c I(\chi_j^c \leq x)$ [9, e.g.], where $\{J_j^c\}$ are random jump sizes and $\{\chi_j^c\}$ are random locations.

If we think of μ as the measure induced by the Lévy process $Z(x)$. That is, for a set $A \subset \mathcal{X}$, say $A = (a_0, a_1]$ for $a_0, a_1 \in \mathcal{X}$, then $\mu(A) := Z(a_1) - Z(a_0)$. Then, in measure theory, μ is called completely random measure. These measures are important since they can be generalized to more general complete and separable metric spaces. We refer the reader to [6] for details.

3. Lévy-driven priors

(3.1) Definition. In general, a Lévy-driven process is any process defined as a function of a Lévy process. The specific form of the Lévy-driven processes used to construct Bayesian nonparamtric priors are Lévy-driven mixtures of a kernel $k(x,s)$ with weights (or mixing measure) given by a Lévy process $Z(s)$. In notation, a Lévy-driven process $W(x)$ has the form

$$W(x) = \int k(x,s)Z(ds) \tag{3.1}$$

As a Lévy process, the Lévy-driven process defined in (3.1) is a Markov process. Depending on the choice of the kernel k, a Lévy-driven process can have piecewise constant paths, smooth increasing paths, or non-monotonic paths. These behaviours can be seen in Figure 1, where we show random paths of Lévy-driven

processes for different choices of the kernel k. Note that the jump sizes and locations were kept the same across the different panels in Figure 1 to better appreciate the influence of the kernel. The general condition we require on a kernel $k(x,s)$ is to be nonnegative for all x and s. Further conditions need to be imposed according to the specific use of the process to properly define a prior. These will be described later in this section.

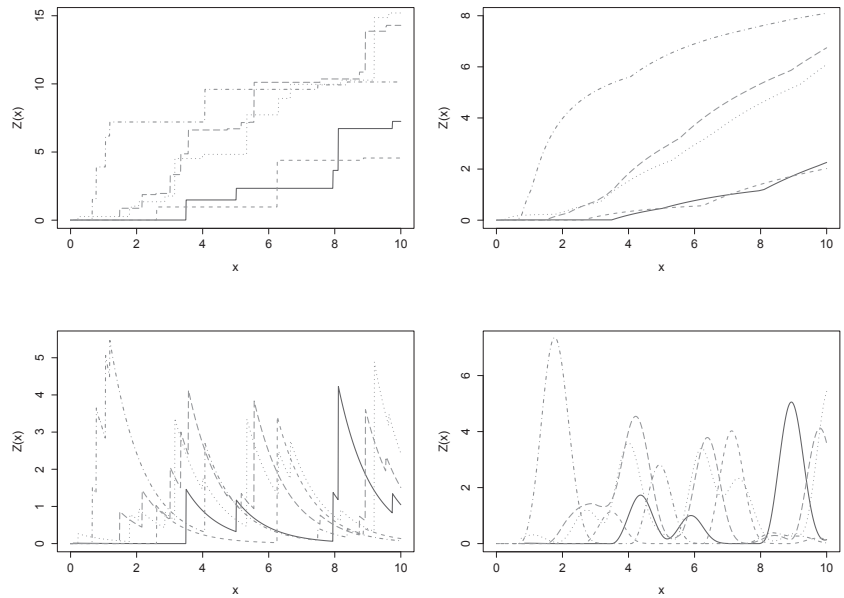


Figure 1. Random Lévy-driven paths for different kernels: $k(x,s) = I(s \le x)$ (top left); $k(x,s) = \{1-(s/x)\}I(s \le x)$ (top right); $k(x,s) = \exp\{-(x-s)\}I(s \le x)$ (bottom left); and $k(x,s) = 3(x-s)^2\exp\{-(x-s)^3\}I(s \le x)$ (bottom right). In each panel the lines represent different random realizations of the process $W(x)$.

(3.2) A limit theorem. There are several ways of characterizing a Lévy-driven process. In this section we describe a way of obtaining a gamma Lévy-driven process as a limit of autoregressive gamma processes in discrete time.

[23] established a connection between Gibbs and autoregressive processes. Considering this connection and by taking a suitable limit, they showed that a particular autoregressive gamma process converges to a Lévy-driven process. We sketch the derivation here.

[20] introduced a Gibbs type gamma process in discrete time, $\{\lambda_k\}$, for $k = 1,2,\ldots$. This process is defined in terms of a set of latent variables $\{u_k\}$ as: $\lambda_1 \sim \text{Ga}(\alpha,1)$ with $u_k|\lambda_k \sim \text{Po}(c\lambda_k)$ and $\lambda_{k+1}|u_k \sim \text{Ga}(\alpha+u_k, 1+c)$. This process has conditional expectation

$$\text{E}(\lambda_{k+1}|\lambda_k) = \frac{\alpha + c\lambda_k}{1+c}, \tag{3.2}$$

and the process is strictly stationary with marginal distribution $\lambda_k \sim \text{Ga}(\alpha, 1)$.

Starting from the Gibbs construction, [23] derived the innovation term for an autoregressive process to be stationary gamma. In particular they obtained that by taking $\lambda_k = \rho\lambda_{k-1} + \rho\zeta_k$, with $\rho = c/(1+c)$, $\zeta_k|\xi_k \sim \text{Ga}(\xi_k, 1)$, $\xi_k|\gamma_k \sim \text{Po}(\gamma_k/c)$ and $\gamma_k \sim \text{Ga}(\alpha, 1)$, implies that $\{\lambda_k\}$ is a strictly stationary process with $\text{Ga}(\alpha, 1)$ marginals and with conditional expected value as that of the Gibbs type process, given in (3.2).

To define a process in continuous time, we define a partition for each n, via $0 = x_{n,0} < x_{n,1} < x_{n,2} < \cdots$, with $x_{n,k} = k/n$ for $k = 0, 1, \ldots$, and make c_n depend on n via $c_n = cn$ for some $c > 0$. We thus define a piecewise constant process $W_n(x)$ as $W_n(0) = \lambda_{n,0}$ and, for $x > 0$

$$W_n(x) = \lambda_{n,k}\, I\left(x_{n,k-1} < x \le x_{n,k}\right),$$

where $\lambda_{n,0} \sim \text{Ga}(\alpha, 1)$ and $\{\lambda_{n,k}\}$ is either the Gibbs or the autoregressive process. This continuous time process $W_n(x)$ is strictly stationary with marginal distribution $\text{Ga}(\alpha, 1)$ for all $x \ge 0$.

[23] showed that the autocorrelation function of process $W_n(x)$, regardless of the choice of $\{\lambda_{n,k}\}$ to be of Gibbs type or autoregressive, converges to $e^{-x/c}$, which is the autocorrelation function of a Lévy-driven process of the form (3.1). However, the only process that does converge to a Lévy-driven is $W_n(x)$ defined in terms of the autoregressive process. In this case, the limiting Lévy-driven process, say $W(x)$, is a shot-noise process of the form

$$W(x) = \int_0^x e^{-(x-s)/c} Z(\mathrm{d}s),$$

where $Z(s)$ is a Lévy process with a fixed jump at zero and finite Lévy measure for the continuous part $\nu(\mathrm{d}v, \mathrm{d}s) = (\alpha/c)e^{-v}\mathrm{d}v\mathrm{d}s$. This implies that $W(x) \sim \text{Ga}(\alpha, 1)$ marginally for all $x \ge 0$. Bottom left panel in Figure 1 presents random paths from a shot-noise process. They have non-monotonic paths and present sudden increments (shots) with exponential decays.

(3.3) Nonparametric priors. As was mentioned in Section 1, a nonparametric prior is a probability measure $\mathcal{P}$ on the space $\mathcal{F}$ of all cumulative distribution functions. Broadly speaking, we can think of a nonparametric prior as a probability measure on the space of probability models. Therefore, we can place the prior on densities, cumulative distributions, survival functions, hazard rates, or any other function that characterizes the behaviour of the observable random variables. In survival analysis, for instance, it is customary to place the prior on the space of survival or hazard functions. For density estimation, the nonparametric prior is usually placed on the density or cumulative distribution functions.

A Lévy-driven process prior is thus, a suitable transformation of a Lévy-driven process, together with some constraints on the kernel k, such that the paths of the transformed process satisfy some propriety conditions over the space of functions where the prior is defined.

A nonparametric prior on the space of cumulative distribution functions can be obtained via a normalization of a Lévy-driven process W of the form

$$F(x) = \frac{W(x)}{\widetilde{W}}, \tag{3.3}$$

where $\widetilde{W} = \lim_{x\to\infty} W(x)$. These type of priors are called *normalized random measures driven by increasing additive processes* and were introduced by [19]. The denominator $\widetilde{W}$ plays the role of the (random) total mass induced by the process $W(x)$, and makes the paths of the new $F(x)$ process to be constrained to lie in the interval $[0,1]$.

For prior (3.3) to be well defined we require that the kernel k and the Lévy measure ν satisfy the following conditions: (a) $k(x,s)$ must be nondecreasing and right continuous as a function of x, and $\lim_{x\to-\infty} k(x,s) = 0$; (b) $\int_{\mathbb{R}}\int_{\mathbb{R}^+}[1-\exp\{-\phi v\tilde{k}(s)\}]\nu(\mathrm{d}v,\mathrm{d}s) < \infty$, for $\phi > 0$ and with $\tilde{k}(s) = \lim_{x\to\infty} k(x,s)$; and (c) $\nu(\mathbb{R}^+\times\mathbb{R}) = \infty$. The Lévy measures (i) and (ii) defined in Section 2 satisfy these properties.

Two particular choices of the kernel become relevant. If $k(x,s) = I_{(-\infty,x]}(s)$, then the process $F(x)$ reduces to a normalized Lévy (increasing additive) process whose paths are almost surely (a.s.) discrete. See [24] for the distribution of means of these processes. On the other hand, if $k(x,s)$ is absolutely continuous as a function of x, for every $s\in\mathbb{R}$, then the process $F(x)$ has absolutely continuous paths (a.s.) with respect to the Lebesgue measure. In this case, the process $F(x)$ admits a density function of the form

$$f(x) = \int k'(x,s)\frac{Z(\mathrm{d}s)}{\widetilde{W}},$$

where $k'(x,s) = \frac{\mathrm{d}}{\mathrm{d}x}k(x,s)$.

If we re-write expression (3.3) in terms of the Lévy process $Z(s)$ and the limit kernel $\tilde{k}(s)$ we obtain

$$F(x) = \frac{\int k(x,s)Z(\mathrm{d}s)}{\lim_{x\to\infty}\int k(x,s)Z(\mathrm{d}s)} = \int \frac{k(x,s)}{\tilde{k}(s)}\frac{\tilde{k}(s)Z(\mathrm{d}s)}{\int\tilde{k}(s)Z(\mathrm{d}s)} = \int k^*(x,s)P(\mathrm{d}s),$$

where $k^*(x,s) = k^*(x,s)/\tilde{k}(s)$ is a normalized (probability) kernel, and $P(\mathrm{d}s) = \tilde{k}(s)Z(\mathrm{d}s)/\int\tilde{k}(s)Z(\mathrm{d}s)$ is a normalized weighted or perturbed Lévy process. These processes were used by [18] to study the impact of the weighting function $\tilde{k}(s)$ in the posterior inference for density estimation. [18] showed that the perturbation function $\tilde{k}$ has little or almost null effect in the posterior density estimates, what makes these type of models to be not sensitive to prior misspecifications.

If $\tilde{k}(s)$ is constant for all s, then the prior (3.3) reduces to the class of mixtures of normalized Lévy processes, better known as mixtures of normalized random measures. In particular, [15] and [16] study the properties of mixtures of normalized inverse Gaussian processes and mixtures of normalized generalized gamma processes, respectively. The posterior distribution of these priors in given in [14]. Additionally, [1] present a review of these processes when used for density estimation and clustering. Via a simulation study the authors show that these kind of processes are a good default choice for density estimation. They also created computational algorithms for fitting these models and made them available though the R package `BNPdensity`. In Section 5 we will illustrate its use with a real dataset.

In the study of nonnegative random variables, as is the case for survival analysis, nonparametric priors are placed on the survival function S, which is defined as the complement to one of a cumulative distribution function F, that is, $S(x) = 1-F(x)$, for $x\geq 0$. In this case, a negative exponential transformation of a

Lévy-driven process W is commonly used. In notation,

$$S(x) = e^{-W(x)}. \tag{3.4}$$

This transformation can be seen as another way of "normalization" of the process $W(x)$, since it transforms the unbounded W process to the unit interval. However, differing from (3.3), here the process $W(x)$ plays the role of a cumulative hazard function, so strictly speaking, the Lévy-driven process places a prior on the space of (cumulative) hazard functions. Priors defined by (3.4) have been studied by several authors, including [21]. This prior is properly defined as long as the kernel $k(x,s)$ is nondecreasing and right continuous as a function of x, and $\lim_{x\to 0} k(x,s) = 0$, for all $s \geq 0$; moreover, the Lévy measure ν has to be defined on $\mathbb{R}^+ \times \mathbb{R}^+$. This is achieved by taking the jump locations of the process from an $\alpha(s)$ measure defined on $[0,\infty)$.

Generalization of prior (3.4) to include covariates has been proposed in [22]. There, the way that covariates are included is via a proportional hazards model [4]. Assuming that the covariates are time dependent, say $\mathbf{u}_i(x)' = (u_{1i}(x),\ldots,u_{qi}(x))$, with q the number of covariates, then the semiparametric prior becomes

$$S_i(x) = \exp\left\{-\int_0^x e^{\boldsymbol{\theta}'\mathbf{u}_i(s)} W(\mathrm{d}s)\right\}, \tag{3.5}$$

for individual $i \in \{1,\ldots,n\}$, where $\boldsymbol{\theta}' = (\theta_1,\ldots,\theta_q)$ is a vector of covariate coefficients.

Within the case of nonnegative random variables, other transformations of Lévy-driven processes can also be used to define nonparametric priors. For example, [17] proposed the semiparametric transformation

$$S_i(x) = \left\{1 + \frac{\lambda_i}{\theta_i} W(x)\right\}^{-\theta_i}, \tag{3.6}$$

for $i = 1,\ldots,n$, where λ_i and θ_i are nonnegative parameters interpreted as the short– ($\lim_{x\to 0}$) and long– ($\lim_{x\to\infty}$) term hazard ratios between individual i and a baseline individual j such that $\lambda_j = \theta_j = 1$. In this case, the Lévy-driven process $W(x)$ turns out to be the odds function $\{1 - S_j(x)\}/S_j(x)$ of the baseline individual j. Broadly speaking, transformation (3.6) can be seen as another "normalization" of the process $W(x)$. Conditions on the kernel k and Lévy process Z are the same as those for prior (3.4).

The semiparametric transformation (3.6) is interesting since it includes several models as particular cases: the proportional hazards model [4], when $\lambda_i = \theta_i$; the proportional odds model [2], when $\theta_i = 1$; and crossing hazards, when $(\lambda_i - 1)(\theta_i - 1) < 0$. Again, covariates can be included in the model via logarithmic links of the form $\lambda_i = \exp(\boldsymbol{\delta}'\mathbf{u}_i)$ and $\theta_i = \exp(\boldsymbol{\gamma}'\mathbf{u}_i)$, with $\mathbf{u}_i$ a fixed time vector of q covariates for individual i.

In a totally different context, nonparametric priors can also be used to make inference on the functional parameters of stochastic processes. To be specific, consider a nonhomogeneous Poisson process, that is, $X(t) \sim \mathrm{Po}(\Lambda(t))$, with $\Lambda(t)$ the cumulative intensive (mean) function of the process. Then, a Lévy-driven prior as in (3.1) can be used as a prior distribution for $\Lambda(t)$. Alternatively, consider a more general mixed Poisson, Cox, or doubly stochastic processes [11] $X(t)$ such that $X(t)|\Lambda \sim \mathrm{Po}(\Lambda(t))$, and $\Lambda(t) \sim \text{Lévy}(\alpha(\cdot),\beta(\cdot))$, where $\alpha(\cdot)$ (unknown) and $\beta(\cdot)$

(known) are the functions that characterize the Lévy measure described in Section 2. Thus, we can use a Lévy-driven prior to express our prior knowledge on $\alpha(s)$ and make inference. These models were studied in detail by [12].

4. Posterior inference

Within the Bayesian paradigm, statistical inference is carried out through the posterior distribution of the unknown quantities. This is the result of updating the prior distributions (prior knowledge) with the observed data via the Bayes' Theorem. In the context of survival analysis, data are not always observed completely ($X_i = x_i$), and only partial information is available, for example, in the form of right censored observations ($X_i > x_i$). Both types of observations are useful information that need to be taken into account in the Bayesian updating process. Let us denote by A_i either, an exact or right censored, observation.

Recall that our functional parameter $W(x)$ is a function of the Lévy process $Z(x)$, and that the law of Z is characterized by its Laplace transform (2.1). Therefore, posterior distribution of W will be characterized by the posterior Laplace transform of Z. That is, we are interested in finding

$$\mathrm{E}\left\{ e^{-\int \phi(s) Z(\mathrm{d}s)} \middle| \mathbf{A} \right\} = \frac{\mathrm{E}\left\{ e^{-\int \phi(s) Z(\mathrm{d}s)} P(\mathbf{A}|Z) \right\}}{\mathrm{E}\{P(\mathbf{A}|Z)\}}, \tag{4.1}$$

where $\mathbf{A} = (A_1, \ldots, A_n)$ denotes the observed data such that $P(\mathbf{A}|Z) = \prod_{i=1}^{n} P(A_i|Z)$, and with $P(A_i|Z)$ becoming the density or the survival function according to whether A_i denotes an exact or a right censored observation, respectively.

Expression (4.1) is nothing but the Bayes' Theorem written in terms of expected values of a specific negative exponential function. Since the law of an stochastic process is characterized by its Laplace transform, expression (4.1) can be seen as the Bayes' Theorem for stochastic processes.

Finding the posterior Laplace transform for the priors described in Section 3 is not an easy task. It requires to be familiar with a bunch of useful mathematical identities. The inclusion of latent variables is also useful, and in many cases, necessary to obtain analytical expressions. If the Lévy-driven process W is embedded into a semiparametric model, as is the case for priors (3.5) and (3.6), posterior characterization of the driving process Z is done conditionally on the other parameters, and a Gibbs sampler [25, e.g.] will be needed to produce numerical results. Details about the posterior characterization of all priors mentioned in Section 3 can be found in their respective references.

5. Illustrations

(5.1) Density estimation. To illustrate the use of Lévy-driven processes in density estimation, we consider the Mexican stamp thickness data [13]. This dataset contains 486 thickness measurements (in millimeters) of the 1872 Hidalgo stamp issue of Mexico. This issue consisted of stamps printed with the image of the national independence hero Miguel Hidalgo y Costilla, lasted in circulation until 1874 and included stamps of five denominations (6, 12, 25, 50 and 100 cents). The data are available as the object *stamp* in the R package `locfit`.

We will take a nonparametric prior of the form (3.3) with an absolutely continuous kernel k such that the prior admits a density. In particular we take

$k'(x,s) = \mathrm{N}(x|s_1, s_2^2)$, that is, a normal density for x with mean s_1 and standard deviation s_2. This choice of kernel implies that $\widetilde{k}(s) = 1$ for all s. The Lévy intensity measure is characterized by the generalized gamma ρ measure (i) with $\epsilon = 0.4$, $\beta(s) = 0$, and α measure given by $\alpha(\mathrm{d}s) = \mathrm{N}(s_1|0.09, 0.0005)\mathrm{d}s_1 \times \mathrm{Ga}(s_2|a_0, a_1)\mathrm{d}s_2$.

Posterior density estimates are shown in Figure 2. Top two graphs are obtained when taking $(a_0, a_1) = (1,2)$, whereas for the last two we take $(a_0, a_1) = (1,1)$. The underlying histograms, in the first and third panels, were produced with the raw data with 30 and 20 bins respectively. We deliberately chose a differentiated number of bins to justify the third mode around value 0.07 (shown in the top panel). The difference in the posterior density estimates is due to the prior choice on the standard deviation s_2, with a prior favouring smaller values on the first panel and larger values on the third panel. In the second and fourth panels of Figure 2, we show the posterior distribution for the total number of groups induced by the mixture, with the second panel favouring three groups and the fourth panel with a distinctive mode in two groups. [13] suggested three mixture components when estimating the density for these data. Our model suggests that two mixture components are also adequate.

(5.2) Survival analysis. We now illustrate the use of a Lévy-driven process in survival analysis. We consider the well known Stanford heart transplant data. There are several version of these data, one of them is that studied in [5] and consists on survival information of 103 patients who were accepted into the heart transplant program. Patients were accepted into the program and when a donor heart became available, medical judgement was used to select the receiving candidate. Among the 103 patients, 69 received transplants, and from them 24 were still alive at the end of the study. The data are available as the object *heart* in the R package `survival`.

The reason why this dataset has been so famous is because patients change treatment status during the course of the study, and thus defining a time dependent covariate. If we denote by w_i the waiting time from acceptance to the day of transplant, for those lucky enough to have a matching donor, then $u_i(t) = I(t \geq w_i)$ is a time dependent indicator variable which takes the value of one or zero according to whether the patient has or has not received a transplant by time t.

To analyse these data we will use model (3.5) and concentrate on the hazard rates which take the form

$$h_i(x) = -\frac{\mathrm{d}}{\mathrm{d}x}\log\{S_i(x)\} = e^{\boldsymbol{\theta}'\mathbf{u}_i(x)}\int_0^x k'(x,s)Z(\mathrm{d}s).$$

To define the prior we use a kernel $k(x,s) = \left\{1 - e^{-a(x-s)^b}\right\} I(s \leq x)$. This corresponds to a location Weibull c.d.f. Here b is a smoothing parameter and a determines the rate of decay, so in particular we take $b = 2$ and a hyper prior $a \sim \mathrm{Ga}(1/2, 1/2)$. The Lévy intensity measure is characterized by the generalized gamma ρ measure (i), with $\epsilon = -1$ so that the measure is discrete, $\beta(s) = 1$ and α measure given by $\alpha(\mathrm{d}s) = \mathrm{Ga}(s|1, 0.001)\mathrm{d}s$. Finally, the prior for θ was a $\mathrm{N}(0, 10)$. The model was implemented in Fortran and is available upon request from the author.

Posterior hazard rate estimates (posterior means) are shown in Figure 3. The solid thin line correspond to the hazard rate for a patients who did not receive

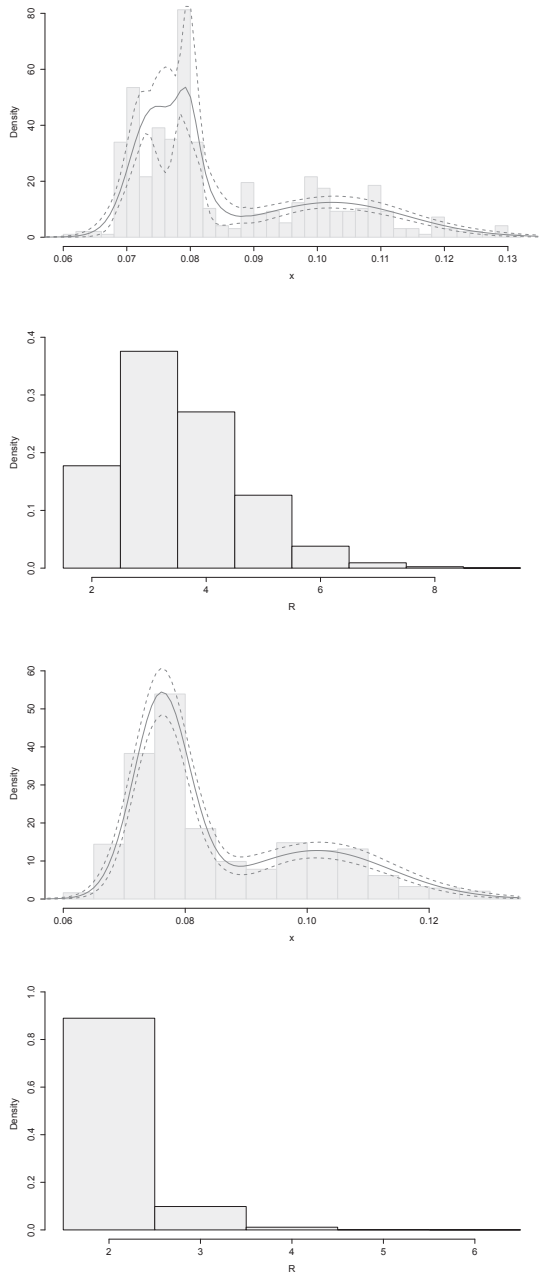


Figure 2. Mexican stamp thickness data. First and thrid panels: histogram of the raw data, density estimate (solid line) and 95% credible intervals (dashed lines). Second and fourth panels: Posterior distribution of the number of mixture components. Top two: $(a_0, a_1) = (1, 2)$; and last two: $(a_0, a_1) = (1, 1)$.

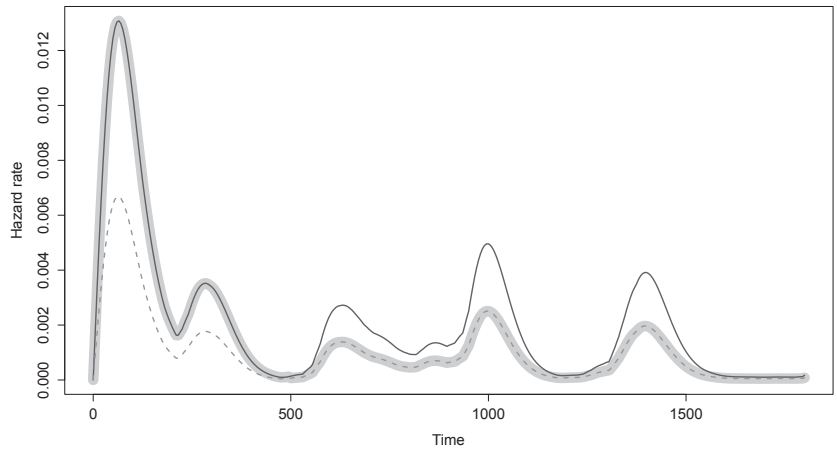


Figure 3. Hazard rate estimates for stanford heart transplant data. With no transplant made (solid thin line), transplant made at time zero (dashed line), and transplant made at time 500 days (thick grey line).

transplant, i.e. $w_i = \infty$. The dashed line corresponds to a patient who received a heart transplant immediately after being accepted in the program, i.e. $w_i = 0$. The effect of a heart transplant can be clearly seen by a lower hazard rate for the patient who did receive a transplant. In fact, this reduction can be quantified by the parameter θ which has a posterior mean of -0.68 and a 95% credible interval $(-1.09, -0.24)$. These values imply a 50% reduction (in average) in the risk of dying after the transplant. Moreover, the thick grey line in Figure 3 corresponds to the hazard rate estimate of a patient who received a heart transplant 500 days after being accepted into the program ($w_i = 500$). In the figure, we can see that this hypothetical patient starts with the no transplant group (higher) hazard rate and at time 500 it changes to the transplant group (lower) hazard rate. This clearly shows the implication for a patient when changing treatment group during the course of the study.

In summary, Lévy-driven priors are useful stochastic processes who have shown to be analytically tractable and are very flexible to produce good results in a wide range of Bayesian nonparametric inference problems.

Acknowledgments

This research was supported by grant I130991-F from the National Council for Science and Technology of Mexico (CONACYT).

Received May 14, 2013

Final version received October 24, 2013

LUIS E. NIETO-BARAJAS
DEPARTMENT OF STATISTICS, ITAM, MEXICO
lnieto@itam.mx

REFERENCES

[1] E. BARRIOS, A. LIJOI, L.E. NIETO-BARAJAS, AND I. PRÜNSTER, *Modeling with normalized random measure mixture models*, Statistical Science, (2013), 313–334. In press. DOI: 10.1214/13-STS416, (2013).

[2] S. BENNET, *Analysis of survival data by the proportional odds model*, Statistics in Medicine **2**, (1983), 273–277.

[3] A. BRIX, *Generalized gamma measures and shot-noise Cox processes*, Advances in Applied Probability **31**, (1999), 629–953.

[4] D.R. COX, , *Regression models and life tables (with discussion)*, Journal of the Royal Statistical Society, Series B **34**, (1999), 187–220.

[5] J. CROWLEY AND M. HU, *Covariance analysis of heart transplant survival data*, Journal of the American Statistical Association **72**, (1977), 27–36.

[6] D.J. DALEY AND D. VERE-JONES, An introduction to the theory of point processes. Volume II: General theory and structure, Springer, New York, 2008.

[7] J.L. DOOB, Stochastic processes, Wiley, New York, (1953).

[8] T.S. FERGUSON, *A Bayesian analysis of some nonparametric problems*, Annals of Statistics **1**, (1973), 209–230.

[9] T.S. FERGUSON AND M.J. KLASS, *A representation of independent increment processes without Gaussian components*, Annals of Mathematical Statistics **43**, (1972), 1634–1643.

[10] I.I. GIKHMAN AND A.V. SKOROKHOD, Introduction to the theory of random processes, Dover, New York, 1969.

[11] J. GRANDELL, Mixed Poisson processes, Chapman and Hall, London, 1997.

[12] E. GUTIÉRREZ-PEÑA AND L.E. NIETO-BARAJAS, *Bayesian nonparametric inference for mixed Poisson processes (with discussion)*, Bayesian Statistics **7**, (2003), 163–179.

[13] A.J. IZENMAN AND C.J. SOMMER, *Philatelic mixtures and multimodal densities*, Journal of the American Statistical Association **83**, (1988), 941–953.

[14] L.F. JAMES, A. LIJOI, AND I. PRÜNSTER, *Posterior analysis for normalized random measures with independent increments*, Scandinavian Journal of Statistics **36**, (2009), 76–97.

[15] A. LIJOI, R. MENA, AND I. PRÜNSTER, *Hierarchical mixture modelling with normalized inverse Gaussian priors*, Journal of the American Statistical Association **100**, (2005), 1278–1291.

[16] A. LIJOI, R. MENA, AND I. PRÜNSTER, *Controlling the reinforcement in Bayesian nonparametric mixture models*, Journal of the Royal Statistical Society, Series B **69**, (2007), 715–740.

[17] L.E. NIETO-BARAJAS, *Bayesian semiparametric analysis of short- and long-term hazard ratios with covariates*, Computational Statistics and Data Analysis **71**, (2014), 477-490. In press. DOI: 10.1016/j.csda.2013.03.012,

[18] L.E. NIETO-BARAJAS AND I. PRÜNSTER, *A sensitivity analysis for Bayesian nonparametric density estimators*, Statistica Sinica **19**, (2009), 685–705.

[19] L.E. NIETO-BARAJAS, I. PRÜNSTER, AND S.G. WALKER, *Normalized random measure driven by increasing additive processes*, Annals of Statistics **32**, (2004), 2343–2360.

[20] L.E. NIETO-BARAJAS AND S.G. WALKER, *Markov beta and gamma processes for modeling hazard rates*, Scandinavian Journal of Statistics **29**, (2002), 413–424.

[21] L.E. NIETO-BARAJAS AND S.G. WALKER, *Bayesian nonparametric survival analysis via Lévy driven Markov processes*, Statistica Sinica **14**, (2004), 1127–1146.

[22] L.E. NIETO-BARAJAS AND S.G. WALKER, *A semiparametric Bayesian analysis of survival data based on Lévy-driven processes*, Lifetime data analysis **11**, (2005), 529–543.

[23] L.E. NIETO-BARAJAS AND S.G. WALKER, *Gibbs and autoregressive Markov processes*, Statistics and Probability Letters **77**, (2007), 1479–1485.

[24] E. REGAZZINI, A. LIJOI, AND I. PRÜNSTER, *Distributional results for means of normalized random measures with independent increments*, Annals of Statistics **31**, (2003), 560–585.

[25] A.F.M. SMITH AND G.O. ROBERTS, *Bayesian computations via the Gibbs sampler and related Markov chain Monte Carlo methods*, Journal of the Royal Statistical Society, Series B **55**, (1993), 3–23.

Bol. Soc. Mat. Mexicana (3) Vol. 19, 2013

NONPARAMETRIC TESTS METHODS FOR STATISTICAL INFERENCE IN POINT PROCESSES

CARLOS DÍAZ-AVALOS

ABSTRACT. Modeling and inference for point processes is a topic that has been investigated broadly in the last years. Application fields such as forestry, epidemiology and ecology have been the main engine driving such rised interest. Subjects of interest in point processes modeling range from pattern detection, modeling clustering and interaction, as well as spatial prediction. Separability of some dimensional components is a common assumption in the context of multidimensional point processes. This hypothesis is especially convenient since each component of a separable process may be modeled and estimated individually, and this greatly facilitates model building, fitting, and assessment. Also, the inclusion of spatially varying covariates in the models for the intensity function is becoming of particular interest. Little attention has been paid to testing formally either the assumption of separability or the significance of covariates. Testing separability and significance of covariates is important if one seeks model adequately the data and to explain which covariates have an effect in the spatial distribution of the point pattern observed. We present the results of the use of non parametric significance tests, using kernel estimation and statistical tests based on discrepancies between the null and the alternative model. Also, two application examples are also presented to illustrate the use of the tests and the conclusions that may be drawn form them.

1. Introduction

Point processes are useful statistical tools that can be applied in a variety of scientific fields such as forestry, epidemiology or ecology. A space-time point process $\mathcal{X}(t)$ may be defined as a finite random subset of a given bounded space-time region $\mathcal{S} \times \mathcal{T} \subset \mathbb{R}^2 \times \mathbb{R}$. A realization of such process is a space-time point pattern $(\mathbf{x}, t) = \{(x_1, t_1), \ldots, (x_n, t_n)\}$ of n space-time points or *events* contained in $\mathcal{S} \times \mathcal{T}$ (Moller and Waagepetersen, 2006). In many studies there is a property $m \in \mathcal{M}$ that can be associated to each point of $\mathcal{X}(t)$ called the *mark*, and we will refer to such proces as a marked space-time point process in $\mathcal{S} \times \mathcal{T} \times \mathcal{M}$. For example, when $\mathcal{X}(t)$ is is used to model an epidemic outbreake, the mark of a given event may be the severity of the disease. This idea may be extended to incorporate more components, giving place to the so called multi dimensional point processes.

Modeling and inference for spatial and spatio-temporal point processes is a subject investigated broadly in the last years. The wide range of application fields has been the main engine driving such raised interest. In paricular, spatial Poisson Processes play an important role in both statistical theory (Daley and Vere-Jones, 1973), and applications (Diggle, 2003). Spatio-temporal marked point

2010 *Mathematics Subject Classification*: 62H11, 62P12, 62M30.

Keywords and phrases: conditional intensity function, separability, covariates, multidimensional spatial point processes, nonparametric estimation, significance, forest fires.

process models have been increasingly used in a wide variety of environmental applications to represent observations of events such as earthquakes, wildfires, sightings of rare species or incidence of epidemics (see Comas and Mateu, 2011; Comas et al.,2009, Diggle, 2003; Guttorp, 1995, Juan et al., 2012, Ripley, 1976, Schoenberg et al., 2002).

Multi-dimensional point processes are typically modeled by specifying the conditional intensity function (CIF) of the process, which may be defined as the limiting expected rate of occurrence of points per space-time-mark volume conditional on $\mathcal{H}_t$, the history of the process prior to time t. When this function exists, the conditional intensity $\lambda(t,\mathbf{x},m)|H_t)$ may be defined as the limiting expected rate of occurrence of points per space-time-mark volume conditional on the history H_t, i.e.

$$\lambda(t,\mathbf{x},m)|H_t) = \lim_{|B|\to 0} \frac{E[N(B)|H_t]}{|B|} \tag{1.1}$$

where $B = (t,t+\Delta t)\times(\mathbf{x},\mathbf{x}+\Delta\mathbf{x})\times(m,m+\Delta m)$, $N(B)$ is the total number of points in B, and $|B|$ is the volume of B.

Modeling the joint distribution of spatial locations, occurrence times, marks and covariates (if any) is challenging, and only a few models are flexible enough to be applied over a wide range of applications. However, existing models commonly in use in such applications almost invariably have a conditional intensity that has a product form, or that, in the terminology of Cressie (1993), is separable. Despite the importance of this assumption, the separability of such processes is rarely rigorously scrutinized, except some notable exceptions (Schoenberg, 2004, Assunçao and Maia, 2007). Similarly, the inclusion of spatially varying covariates in models for the intensity function is becoming of particular interest because this allows modeling non constant trends in the conditional intensity function and because the estimated coefficients associated to such covariates often have a meaning in the context of ecological, epidemiological or other fields of application.

The separability hypothesis is especially convenient since each component of a separable process may be modeled and estimated individually. A space-time spatial point process $\mathcal{X}(t)$ is separable if its conditional intensity function can be expressed as

$$\lambda(x,t) = \lambda_1(x)\lambda_2(t) \tag{1.2}$$

where the term $\lambda_1(x)$ is usually modeled as $\lambda_1(x) = \exp\{\mathbf{c}'\beta\}$, with $\mathbf{c}\in\mathcal{C}$ a vector of spatially varying covariates and β a vector of coefficients.

The concept of separability may be extended to multi-dimensional point processes $\mathcal{X}\in\mathcal{S}\times\mathcal{T}\times\mathcal{M}\times\mathcal{C}$. For instance, for a multi-dimensional point process with time-varying covariates $\{c_j(t,x), j=1,\ldots,k\}$ is completely separable, i.e., the spatial-temporal coordinates, the mark, and the k covariates are all separable from each other, if its conditional intensity function can be written as

$$\lambda^S(t,\mathbf{x},m;c_1(t,\mathbf{x}),\ldots,c_k(t,\mathbf{x})|H_t) = \lambda_1(t,\mathbf{x}|H_t)f_0(m)f_1(c_1(t,\mathbf{x}))\ldots f_k(c_k(t,\mathbf{x})) \tag{1.3}$$

where λ_1 is a nonnegative predictable process and $f_0,f_1,\ldots,f_k$ are fixed nonnegative functions.

From the ecological and practical point of view, choosing between a separable or a non separable model may be a key issue given the different implications that

both models pose. A separable model implies that given two different regions A and $B \in D$, the ratio

$$\frac{\int_A \lambda(x,t)dxdt}{\int_B \lambda(x,t)dxdt} \tag{1.4}$$

does not change with time. On the other hand, testing the significance of spatial covariates is of interest in applications where the goal of the study is to identify which factors are related to the incidence of events and how the effect of such factor varies across an area of interest (Díaz-Avalos et al, 2013).

Motivated by previous papers published in both topics, we present here the results of two separate studies: one is focused on the evaluation of nonparametric test statistics for separability and the other focuses on nonparametric tests for significance for covariates in multi-dimensional point processes. Both analyses are exemplified with application in the modeling of forest fires incidences in two separate localities. The rest of the paper is organized as follows: in section 2 we present some basic concepts and definitions for space-time point processes, section 3 is devoted to the results of the analysis of separability test statistics, with an application to modeling forest fires in Oregon. Section 4 presents the results of the analysis of non parametric tests for significance in multi-dimensional point processes. We finish the paper with a discussion of the results and conclusions regarding the benefits and drawbacks of non parametric tests for point process inference.

2. Spatial Point Processes

A point process is a random collection of points in some metric space. In general, a point process is formed by locations or spatial coordinates, a temporal coordinate, and a uni- or multi-dimensional mark. In such case we are in the context of spatio-temporal marked point processes. In addition, it is also common to have covariates defined in the region where this point process lies. Modeling and inference for spatial and spatio-temporal point processes is an issue that has been broadly investigated in the last years. In their most basic form, see Gelfand et al. (2010), spatio-temporal point process data consist of a time-ordered sequence of events $((x_i,t_i),: i,...,n)$, where $\mathbf{x} = (x_1,...,x_n)$ denotes the spatial location, t denotes time of occurrence of an event that falls within a spatial region D and a time-interval $[0,T]$.

Following Chang and Schoenberg (2011), for the case of spatio-temporal marked point processes, each point may be represented as a point $(t,\mathbf{x},m) \in \mathbb{R}^{n+2}$, where t is a one-dimensional temporal coordinate, $\mathbf{x} = (x_1,\ldots,x_n)$ is an n-dimensional spatial coordinate, and m is a one-dimensional mark. Note that if $\mathcal{M} = \{1,\ldots,k\}$, the marked spatial point process $\mathcal{X}$ is a multitype point process with k different types of points. Multitype point processes may be thought as a $k-$tuple of unmarked spatial point processes $\{\mathcal{X}_1,\ldots,\mathcal{X}_k\}$.

A characteristic basic to point process modeling is the Conditional Intensity Function (CIF) $\lambda(t,\mathbf{x},m)|H_t)$, which was defined in equation (1.1). Following Schoenberg (2004), we say the marked point process is separable with respect to the mark m if its CIF can be expressed as

$$\lambda^S(t,\mathbf{x},m|H_t) = \lambda_1(t,\mathbf{x}|H_t)f(m) \tag{2.1}$$

where λ_1 is a nonnegative predictable process and f is a fixed nonnegative function.

Complete separability is particularly convenient in modeling point processes with covariates, since in such cases one may readily inspect and model the influence of the covariate on the intensity of the point process simply by inspecting this covariate individually, and under quite general conditions, the parameters governing each component of the model may be consistently estimated individually by maximum likelihood as in Schoenberg (2006).

Let $c_1(t,\mathbf{x}),\ldots,c_k(t,\mathbf{x})$ be k covariates. The process N is completely separable, i.e., the spatial-temporal coordinates, the mark, and the k covariates are all separable from each other, if the CIF can be expressed as

$$\lambda^S(t,\mathbf{x},m;c_1(t,\mathbf{x}),\ldots,c_k(t,\mathbf{x})|H_t) = \lambda_1(t,\mathbf{x}|H_t)f_0(m)f_1(c_1(t,\mathbf{x}))\ldots f_k(c_k(t,\mathbf{x})) \tag{2.2}$$

where λ_1 is a nonnegative predictable process and $f_0,f_1,\ldots,f_k$ are fixed nonnegative functions.

Then, one may inspect the separability of a covariate with respect to a mark of the process. Note that we have done explicit the possible dependence of the covariates on time. In the case of separability of a covariate with respect to one or more spatio-temporal coordinates of the process, the CIF can be expressed as

$$\begin{aligned}\lambda^S(t,\mathbf{x},m;c_1,\ldots,c_k|H_t) &= \lambda_j(t,\mathbf{x};c_j|H_t)\lambda_{-j}(m,c_1,\ldots,c_{j-1},c_{j+1},\ldots,c_k|H_t)\\ &= \lambda_1(t,\mathbf{x}|H_t)f_j(c_j)h(m,c_1,\ldots,c_{j-1},c_{j+1},\ldots,c_k)\end{aligned}$$

(2.1) Estimation of CIF's. Following the strategy of previous works by Schoenberg (2004), Assunçao and Maia (2007), and Chang and Schoenberg (2011), we consider nonparametric kernel estimation of the conditional intensity functions. Assume that $\widehat{\lambda}^{NS}(t,\mathbf{x},m|H_t)$ is a nonparametric kernel estimate of the nonseparable CIF of a spatio-temporal-marked point process N, then

$$\widehat{\lambda}^{NS}(t,\mathbf{x},m|H_t) = \int_\chi K_{n+2}(t-t',\mathbf{x}-\mathbf{x}',m-m')dN(t',\mathbf{x}',m') \tag{2.3}$$

where K_d denotes a d-dimensional kernel density. Let $\widehat{\lambda}_1(t,\mathbf{x}|H_t)$ denote a nonparametric kernel estimate of the CIF of the marginal spatio-temporal point process consisting exclusively of the locations and times of points of N, and ignoring the marks, then

$$\widehat{\lambda}_1(t,\mathbf{x}|H_t) = \int_\chi K_{n+1}(t-t',\mathbf{x}-\mathbf{x}')dN(t',\mathbf{x}',m) \tag{2.4}$$

Similarly, consider a nonparametric kernel estimate $\widehat{f}(m)$ of the mark density f,

$$\widehat{f}(m|H_t) = \frac{1}{N(\chi)}\int_\chi K_1(m-m')dN(t,\mathbf{x},m') \tag{2.5}$$

where $N(\chi)$ is the total number of points in χ. Under the null hypothesis of separability, a separable estimate of $\lambda(t,\mathbf{x},m|H_t)$ is given by

$$\widehat{\lambda}^S(t,\mathbf{x},m|H_t) = \widehat{\lambda}_1(t,\mathbf{x}|H_t)\widehat{f}(m|H_t) \tag{2.6}$$

Under the null hypothesis of separability, the two CIF estimates, $\hat{\lambda}^{NS}(\cdot)$ and $\hat{\lambda}^{S}(\cdot)$, given in (2.3) and (2.6), respectively, should be similar. For details and guidelines for optimally selecting densities and bandwidths for kernel smoothing, and for correcting for boundary effects, see Silverman, (1986).

(2.2) CIF estimation in the presence of covariates. Suppose now that the CIF for a spatio-temporal marked point process depends not only on the time, spatial location, and mark but also on a set of covariates. While Schoenberg (2004) suggests ordinary kernel smoothing of the marginal densities of the process with respect to the coordinates of the process, it is clear that different approaches are necessary in examining the separability of a spatial-temporal marked point process model with covariates. In particular, if the distribution of a covariate is non-uniform, then a simple kernel regression estimate of $f_j[c_j(t,\mathbf{x})]$ will be substantially biased, and an adjustment must be made. Instead, an adjustment analogous to the Nadaraya-Watson estimator may be used (see Silverman, 1986), that is, one may compute a quantity of the form

$$\widehat{g}_j(c) = \frac{\int_\chi K_1(c-c')dN(t,\mathbf{x},m;c)}{\int_{\mathbb{R}} K_1(c-c')dN(c')} = \frac{\sum_{i=1}^{n_1} K_1(c-c'_i)}{\sum_{i=1}^{n_2} K_1(c-c'_i)} \tag{2.7}$$

where the summation in the numerator of equation (2.7) is taken over all n_1 observed points in which the covariate is observed at the same temporal instant that the phenomenon under study, and denoted by c'_i, and the summation in the denominator is taken over all n_2 temporal instants on which the covariate is observed.

The function $g_j(c)$ may be interpreted as a relative hazard, corresponding to those temporal instants when the jth covariate achieves a value of c. As with kernel CIF estimates, it is natural to desire a version of f_j scaled to integrate to n_1, so a natural estimate of the contribution $f(c)$ associated with a covariate of value c is given by

$$\widehat{f}_j(c) = \frac{\widehat{g}_j(c)\cdot n_1}{\int_{\mathbb{R}} \widehat{g}_j(v)dv} \tag{2.8}$$

Assume now that we have a complete point process given by $N=(t,\mathbf{x},m,c)\in\mathbb{R}^{n+3}$, i.e. we have spatial locations in $\mathbb{R}^n$, temporal instants $t\in\mathbb{R}$, a mark variable $m\in\mathbb{R}$, and a covariate $c\in\mathbb{R}$. We are now interested in analyzing the separability between the covariate and the rest of components. Let $\widehat{\lambda}^{NS}(t,\mathbf{x},m;c|H_t)$ denote the estimator of the nonseparable CIF, and $\widehat{\lambda}^{S}(t,\mathbf{x},m;c|H_t)=\widehat{\lambda}_1(t,\mathbf{x},m|H_t)\widehat{\lambda}_2(c)$ the corresponding one for the separable case.

Following now an adjustment analogous to the Nadaraya-Watson estimator, as commented before, we have that

$$\begin{aligned}\widehat{g}^{NS}(t,\mathbf{x},m;c) &= \frac{\int_\chi K_{n+3}(t-t',\mathbf{x}-\mathbf{x}',m-m',c-c')dN(t',\mathbf{x}',m',c')}{\int_{\mathbb{R}} K_1(c-c')dN(c')} \\ &= \frac{\sum_{i=1}^{n_1} K_{n+2}(t-t'_i,\mathbf{x}-\mathbf{x}'_i,m-m'_i)K_1(c-c'_i)}{\sum_{i=1}^{n_2} K_1(c-c'_i)}\end{aligned}$$

for the nonseparable case, and

$$\begin{aligned}\widehat{g}^S(t,\mathbf{x},m;c) &= \frac{\int_\chi K_{n+2}(t-t',\mathbf{x}-\mathbf{x}',m-m')dN(t',\mathbf{x}',m',c)\frac{1}{N(\chi)}}{n_1\cdot\int_{\mathbb{R}}K_1(c-')dN(')}\\ &\quad \frac{\int_\chi K_1(c-c')dN(t,\mathbf{x},m,c')}{}\\ &= \frac{\sum_{i=1}^{n_1}K_{n+2}(t-t'_i,\mathbf{x}-\mathbf{x}'_i,m-m'_i)\sum_{i=1}^{n_1}K_1(c-c'_i)}{n_1\cdot\sum_{i=1}^{n_2}K_1(c-c'_i)}\end{aligned}$$

for the separable one. Finally, the corresponding CIFs will be of the form

$$\widehat{\lambda}^{NS}(t,\mathbf{x},m;c|H_t)=\frac{\widehat{g}^{NS}(t,\mathbf{x},m;c)\cdot n_1}{\int_\chi \widehat{g}^{NS}(t',\mathbf{x}',m';c')dt'd\mathbf{x}'dm'dc'} \tag{2.9}$$

and

$$\widehat{\lambda}^{S}(t,\mathbf{x},m;c|H_t)=\frac{\widehat{g}^{S}(t,\mathbf{x},m;c)\cdot n_1}{\int_\chi \widehat{g}^{S}(t',\mathbf{x}',m';c')dt'd\mathbf{x}'dm'dc'} \tag{2.10}$$

(2.3) Non parametric test statistics . Schoenberg (2004) proposed several nonparametric test statistics, some based in absolute discrepancies (2.11), others in squared discrepancies (2.19) as well as likelihood ratios (2.20) of the following form

$$\begin{aligned}S_1 &= \sup\left\{\frac{|\widehat{\lambda}^{NS}(t,\mathbf{x},m|H_t)-\widehat{\lambda}^{S}(t,\mathbf{x},m|H_t)|}{\sqrt{\widehat{\lambda}^{S}(t,\mathbf{x},m|H_t)}};(t,\mathbf{x},m)\in N\right\}\\ S_3 &= \int_0^T\int_{\mathbb{R}^n}\int_{\mathbb{R}}[\widehat{\lambda}^{NS}(t,\mathbf{x},m|H_t)-\widehat{\lambda}^{S}(t,\mathbf{x},m|H_t)]^2dmd\mathbf{x}dt\\ S_4 &= \int_\chi[\log\{\widehat{\lambda}^{NS}(t,\mathbf{x},m|H_t)\}-\log\{\widehat{\lambda}^{S}(t,\mathbf{x},m|H_t)\}]dN\times\\ &\quad\times\int_0^T\int_{\mathbb{R}^n}\int_{\mathbb{R}}[\widehat{\lambda}^{NS}(t,\mathbf{x},m|H_t)-\widehat{\lambda}^{S}(t,\mathbf{x},m|H_t)]^2dmd\mathbf{x}dt\end{aligned}$$

Assunçao and Maia (2007) derived a score test statistic T and showed that there is a close relationship between the score test statistic T and the test statistic S_4 proposed by Schoenberg (2004). They assumed that, if the process is nonseparable, there exists a constant $\epsilon\geq 0$ and a certain predictable function $g(t,\mathbf{x},m)$ such that

$$\lambda(t,\mathbf{x},m|H_t)=\lambda_1(t,\mathbf{x}|H_t)f(m)[1+\epsilon g(t,\mathbf{x},m)] \tag{2.11}$$

where $g(t,\mathbf{x},m)$ is the relative difference between $\lambda(t,\mathbf{x},m|H_t)$ and $\lambda_1(t,\mathbf{x}|H_t)f(m)$. Note that if $\epsilon=0$, $\lambda(t,\mathbf{x},m|H_t)=\lambda_1(t,\mathbf{x}|H_t)f(m)$ and the CIF is separable with respect to the marks. Assunçao and Maia (2007)then derived the score test statistic T based on the log-likelihood of the model in (2.11):

$$T=\sum_i\frac{\widehat{\lambda}^{NS}(t_i,\mathbf{x}_i,m_i|H_t)}{\widehat{\lambda}^{S}(t_i,\mathbf{x}_i,m_i|H_t)}-\int_\chi\left[\widehat{\lambda}^{NS}(t,\mathbf{x},m|H_t)-\widehat{\lambda}^{S}(t,\mathbf{x},m|H_t)\right]dtd\mathbf{x}dm-n \tag{2.12}$$

and showed using first-order Taylor expansions that the test statistic S_4 is approximately equal to the score test statistic T under the null hypothesis of separability.

Díaz-Avalos et al. (2013a) considered test statistics based on the fact that, if

$$\int_{\chi} \frac{\lambda(t,\mathbf{x}|H_t)}{\int_{\chi}\lambda(u,\mathbf{y}|H_t)dud\mathbf{y}} = 1$$

then

$$f(t,\mathbf{x}) = \frac{\lambda(t,\mathbf{x}|H_t)}{\int_{\chi}\lambda(u,\mathbf{y}|H_t)dud\mathbf{y}}$$

is a density in $\mathbb{R}^d \times \mathbb{R}$.

A natural estimator for $f(t,\mathbf{x})$ is given in its general form by $\hat{p} = \frac{\hat{\lambda}(t,\mathbf{x}|H_t)}{\sum\hat{\lambda}(t,\mathbf{x}|H_t)}$, which is evaluated for each individual and using both the separable and nonseparable expressions for the CIFs. We thus have for each $i = 1,\ldots,n$

$$\widehat{p}_i^{NS} = \frac{\hat{\lambda}^{NS}(t_i,\mathbf{x}_i|H_t)}{\sum_j \hat{\lambda}^{NS}(t_j,\mathbf{x}_j|H_t)} \tag{2.13}$$

and

$$\widehat{p}_i^{S} = \frac{\hat{\lambda}^{S}(t_i,\mathbf{x}_i|H_t)}{\sum_j \hat{\lambda}^{S}(t_j,\mathbf{x}_j|H_t)} \tag{2.14}$$

Large values of any of these test statistics suggest departures from separability. As suggested in Schoenberg (2004), p-values for these test statistics may readily be obtained using simulations of separable marked point processes each with CIF equal to $\widehat{\lambda}^S(t,\mathbf{x},m|H_t)$.

The performance of the test statistics described in the previous subsection was analyzed using simulation under different scenarios:

- Inhomogeneous Poisson point patterns in the unit hypersquare in R^d, for $d = 3,4,5$ with intensity function

$$\lambda(t,x,y,m,c) = 100 + 100\epsilon\left(\frac{1}{4}\right)^{-k} I_{[.5,.75]^k}(\mathbf{v}) \tag{2.15}$$

  where ϵ is a non negative constant and $\mathbf{v}$ is a vector containing the k coordinates that are being tested for separability.

  to consider all combinations of dimensions for the spatio-temporal marked point patterns with covariates.
- A clustering structure given by Neymann-Scott point processes

Point patterns coming from inhomogeneous Poisson processes with intensity given in (2.15) and from Neymann-Scott processes become approximately separable for values of ϵ close to 0, and depart from separability as ϵ increases. We simulated point processes with intensity as in (2.15) and from the Neymann-Scott process, for a sequence of values of ϵ in the $[0,10]$ interval. For each resulting pattern Values of S_3, S_4, T, KL and H were computed to get a sample of 1000 values of each test statistic under the null and the alternative hypothesis. For each one of the test statistics, we computed their quantiles under the null hypothesis. Next, for $\epsilon \in (0,10]$ we counted the number of times that a particular test statistic value obtained under the alternative hypothesis was above the $i-th$ quantile under H_o,

and the result divided by the number of simulations. The quantiles under the null hypothesis can be considered as values for α when testing H_o vs H_a.

Figures 1 and 2 as well as Figures 3 and 4 show image plots (left panels) of the matrices corresponding to the $1-\beta$ values for the five test statistics. The matrices are arranged such that the $x-axis$ represents $1-\alpha$ whilst the $y-axis$ represents ϵ for the different simulation experiments. The right panels of these figures show the corresponding plots of the probability of Type II error for each test statistic and for $\epsilon = 3.0$ in (2.15). We can see that when the intensity function corresponds to the form in (2.15), the inhomogeneous Poisson case, the performance of the tests T, KL and H is similar (see Figures 1 and 2). In that case, the probability of Type II error for KL and H decreases faster than for any other test statistic, except at high values of α.

For the Neymann-Scott models, the statistics KL and H always showed a better performance in terms of β, except for high α when testing separability of time t and spatial coordinates (x,y) (see Figures 3 and 4). In that case, the test based on T has a lower value of β. We note and increased power associated with the KL and H statistics compared with S_3 and S_4 for the inhomogeneous cases. This is not generally true for the cluster cases. This is something expected as our statistics do not depend on the statistical distribution of the data, and on the probabilistic information coming out of it. Thus for Poisson cases where everything is random (and the statistical information is weak) our statistics do as well as when some spatial structure is present in the data, while providing strong power values compared with the other statistics. Of course the increase of power is directly related to an increase of the Type I error, the fact that the lines for the different tests crossed for some experiments indicates that under those conditions there is not a uniformly most powerful test among those studied here.

(2.4) Separability test for lightning ignitions in the US Pacific Nortwest. To illustrate the application of the separability test statistics to real data, we used the test statistics studied in this paper with a dataset of lightning caused forest fires in the US Pacific Northwest. The data are the ignition locations of 5847 fires in the Blue Mountains area, Oregon. The ignitions were recorded between April 1 1986 to November 30th 1993. This example was shown in Díaz-Avalos et al. (2013a). Figure 5 shows the geographic location of the Blue Mountains area as well as the spatial location of the ignitions and a non parametric kernel estimate of the spatial intensity function.

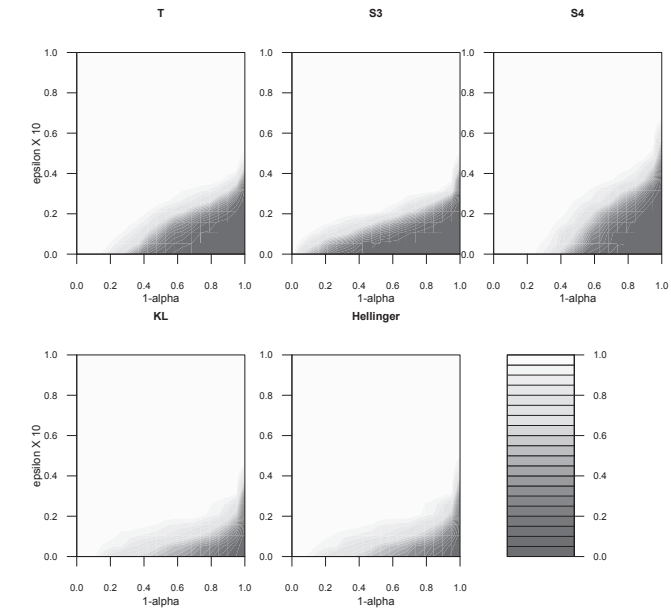
For each ignition, latitude, longitude, day of occurrence, elevation, vegetation type and slope were recorded. Also, the average daily temperature for the Blue mountains during the time span of the data was available, so we associated such temperature to the day of occurrence of each ignition. This dataset has been analysed by Díaz-Avalos et al. (2001) in the context of Markov random fields, and by Møller and Díaz-Avalos (2010) in the context of spatial point processes.

Møller and Díaz-Avalos (2010) proposed the following shot-noise point process model

$$\lambda(t,x,y,\mathbf{c},T) = \lambda_1(x,y,\mathbf{c})\lambda_2(t,T)S(t,x,y) \tag{2.16}$$

where

$$\lambda_1(x,y,\mathbf{c}) = \exp(\mathbf{c}\beta_1)$$



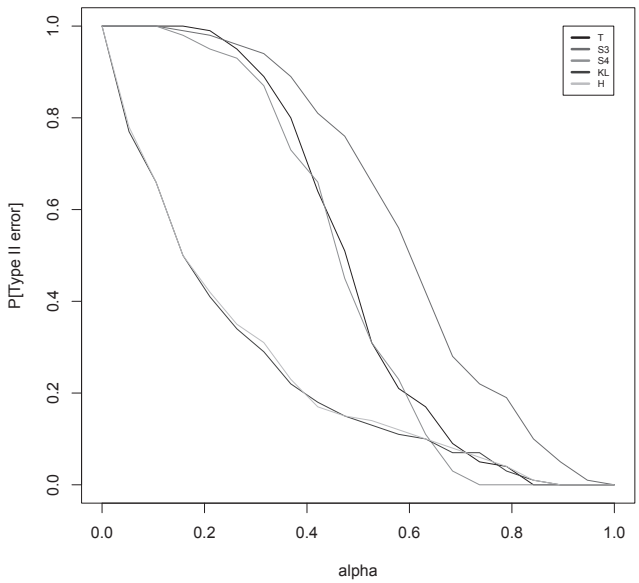
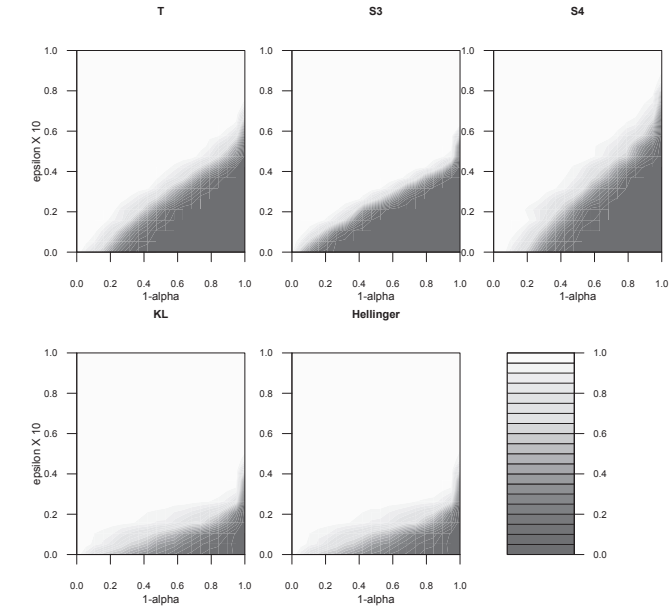


Figure 1. Inhomogeneous Poisson case with intensity function as in (2.15). Testing separability between (x, y) and t.



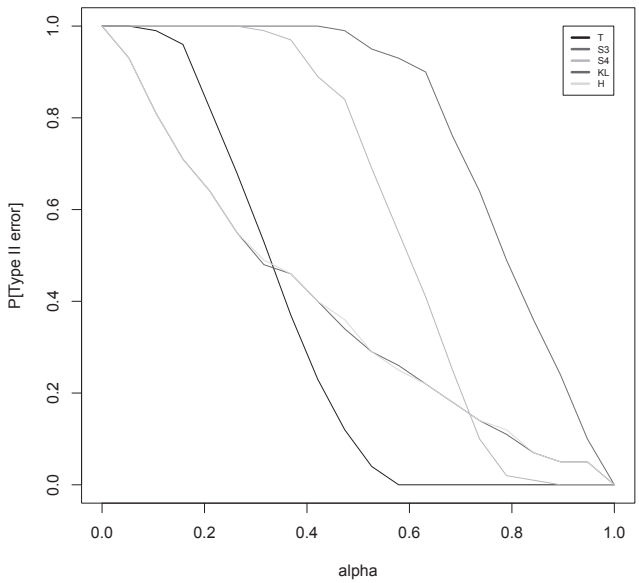
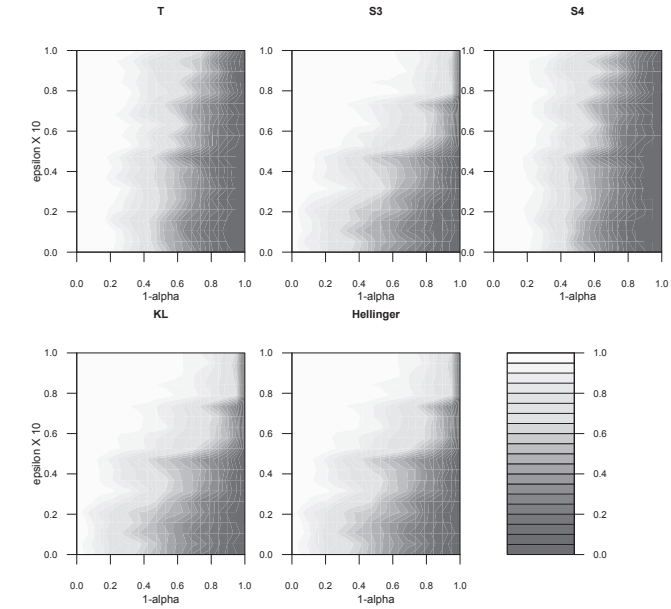


Figure 2. Inhomogeneous Poisson case with intensity function as in (2.15). Testing separability between (t,x,y) and m.



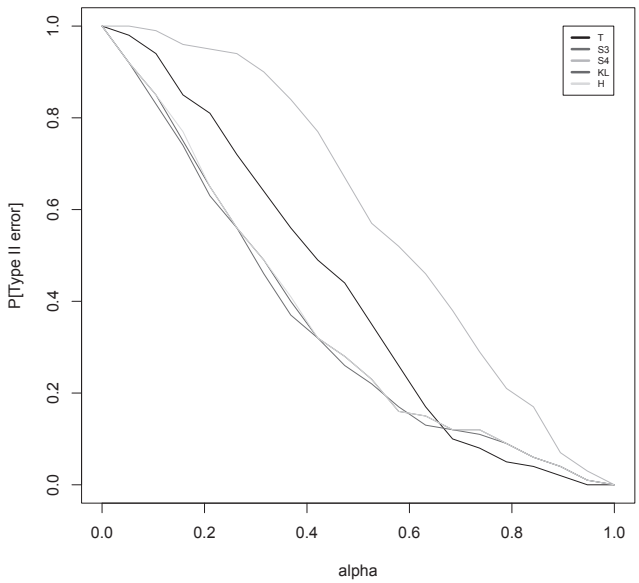
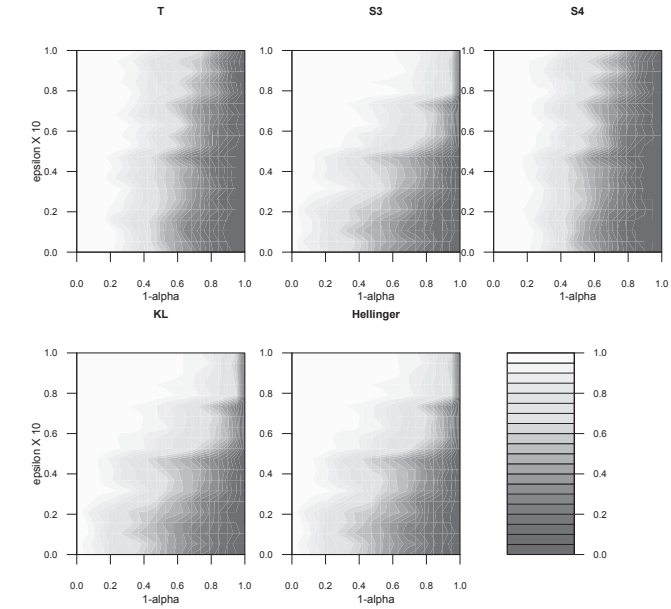


Figure 3. Neyman-Scott case. Testing separability between (x, y) and t.



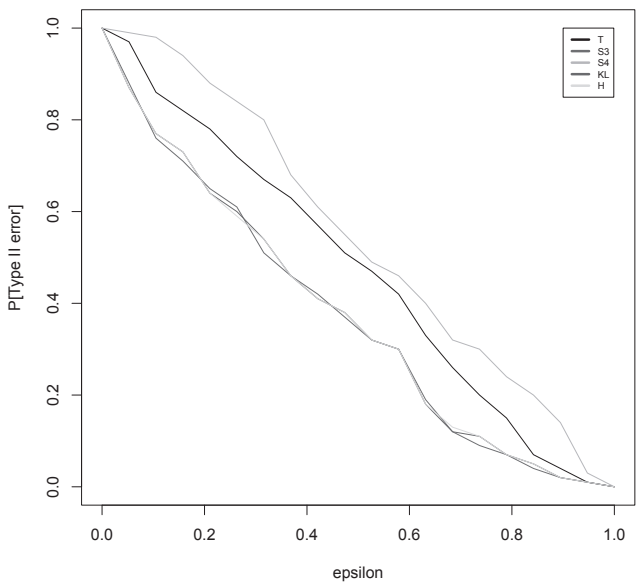


Figure 4. Neyman-Scott case. Testing separability between (t, x, y) and m.

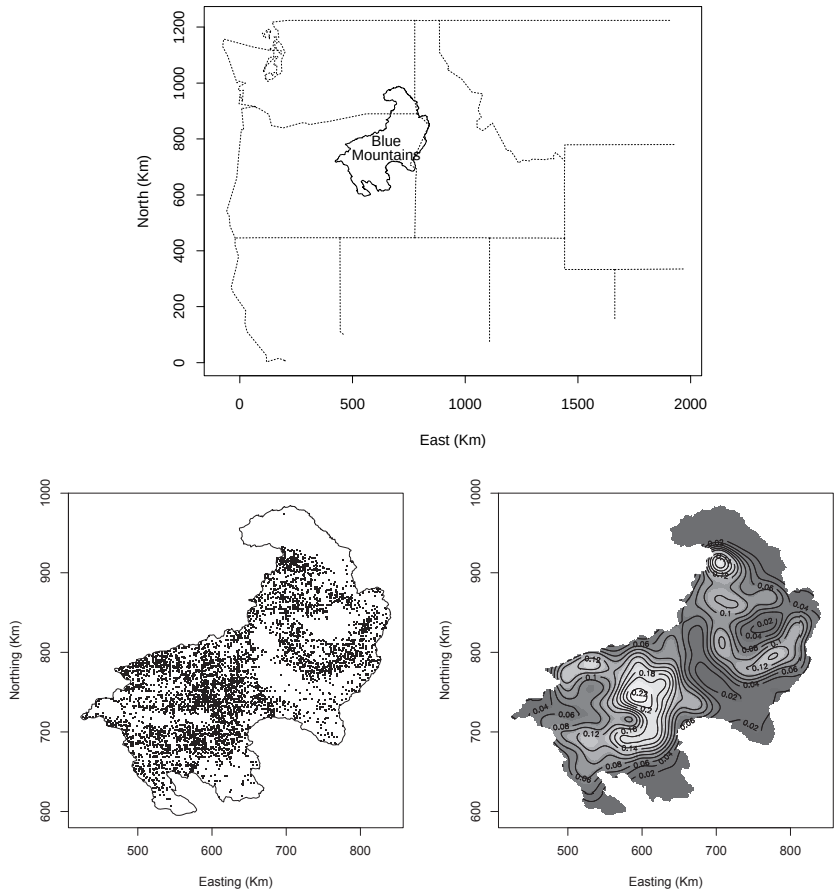


Figure 5. Location of the Blue Mountains area (upper figure) and the spatial location of 5847 lightning caused ignitions occurring between April 1, 1986 and November 25, 1993 (lower figure, left), together with a non-parametric kernel estimate (lower figure, right).

is the intensity function of the point process in space, which includes the effect of space-varying covariates (vegetation type, elevation, slope and exposure) in the linear predictor;

$$\lambda_2(t,T) = \exp[F(t,T)\beta_2]$$

is the intensity on time, which includes the effect of the time-dependent average temperature T and a polynomial time trend in the matrix F, and $S(t,x,y)$ is a separable shot noise point process.

Separability of space and time was assumed by Møller and Díaz-Avalos (2010), and although the fitted models did not show serious lack of fit problems, the hypothesis of separability was not tested either formally or non parametrically. Díaz-Avalos et al. (2013a) applied the testing procedures described in previous

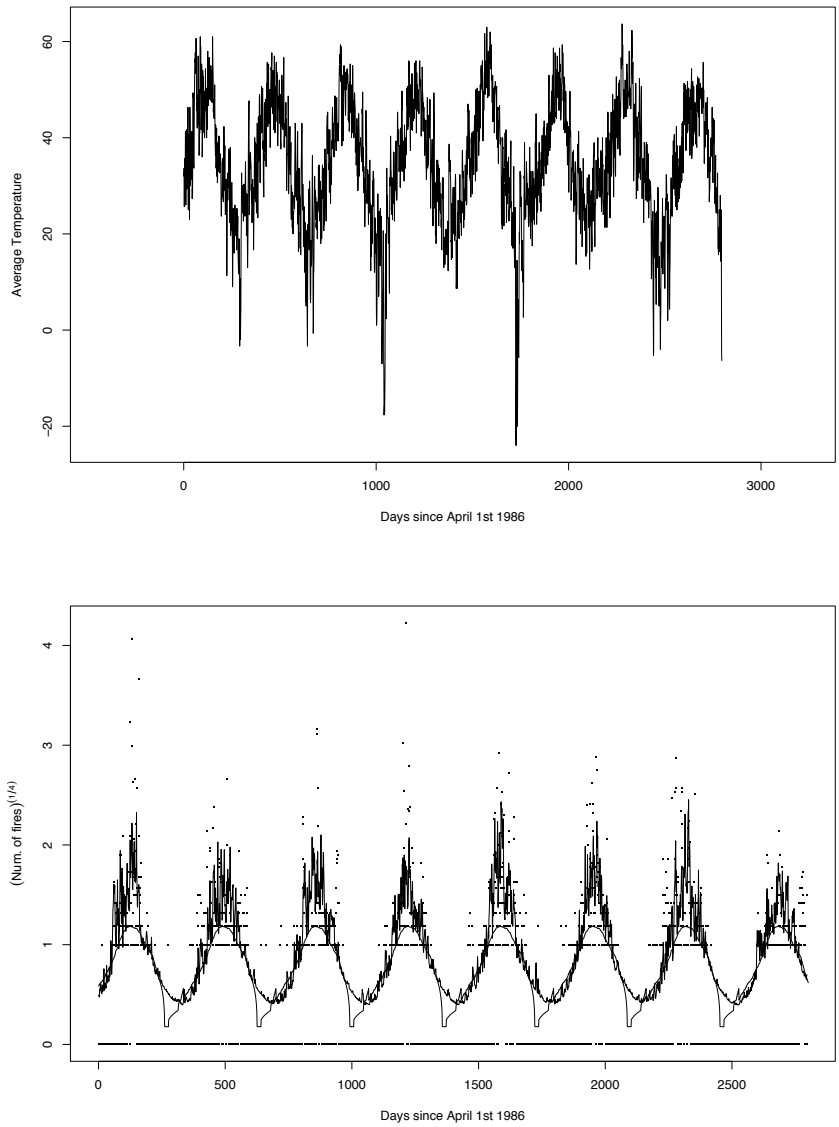


Figure 6. Average daily temperature in Fahrenheit (upper panel) and square root of the daily number of fire occurrences (lower panel, solid dots) in the Blue Mountains area, together with a non-parametric kernel estimate (lower panel, solid line) and a parametric estimate (lower panel, dashed line) of the temporal intensity function $\lambda_2(t)$. The x-axis in both panels shows days $t = 1,\ldots,2796$ corresponding to the period from April 1, 1986 to November 25, 1993.

sections to the lightning ignition data to check if the data set analysed by Møller and Díaz-Avalos (2010) shows evidence of spatio-temporal separability.

By following the procedure described by Møller and Díaz-Avalos (2010) to obtain simulated patterns from model (2.16), Díaz-Avalos et al. (2013a) obtained 100 such patterns and for each pattern they estimated the separable and nonseparable CIFs $\hat{\lambda}^S(t,x,y,c,T)$ and $\hat{\lambda}^{NS}(t,x,y,c,T)$. Using 100 simulated patterns from model (2.16), the nonseparable kernel estimator was computed as

$$\hat{\lambda}^{NS}(t,x,y,c,T) = \sum_{i=1}^{n_1} K_{n+2}(t-t_i', x-x_i', y-y_i', T-T_i') I_{\{\|c-c_i\|\leq\varepsilon\}}(c-c_i) \tag{2.17}$$

The factor $I_{\|c-c_i\|\leq\varepsilon}(c-c_i)$ was included to make the summation in the kernel only for those sites where the categorical covariates vegetation and slope-exposure had close values by choosing ε small enough. The separable kernel estimator is given by

$$\hat{\lambda}^S(t,x,y,c,T) = \sum_{i=1}^{n_1} K_n(x-x_i', y-y_i',) K_2(t-t_i', T-T_i') I_{\|c-c_i\|\leq\varepsilon}(c-c_i) \tag{2.18}$$

The hypothesis of separability of time and temperature was tested, that is, the validity of model (2.16) was under testing. With the resulting values for the test statistics computed from the kernel estimators, Díaz-Avalos et al. (2013a) obtained the empirical $p-values$ for each test statistic with the following results: $T = 0.22$; $S_3 = 0.38$; $S_4 = 0.53$; $KL = 0.37$ and $H = 0.32$.

None of the tests reject the hypothesis of separability of time and temperature, the two variables that describe the seasonal variation in the number of fire ignitions in the Blue Mountains. As a result, we may conclude that model (2.16) is adequate to describe the Blue Mountains ignition pattern. A consequence of the separability of space and time is that the shape of the intensity function λ_1 does not change in a significant way along time, and that such spatial pattern is only shifted upwards or downwards as time, and hence average temperature, vary along the different seasons. In other words, for two points $\mathbf{x}_i$ and $\mathbf{x}_j$, the ratio $\lambda(\mathbf{x}_i/\mathbf{x}_j)$ remains constant over time. A possible explanation for this is that fire ignition is associated to vegetation type through the litter shape, thickness and moisture (Omi, 2005). None of these factors changes along time, except for moisture, which is directly related to average temperature. The fact that separability is not rejected allows to give a local estimate of fire risk based only in covariate estimation and day of the year, something of great usefulness both for the public and for fire management agencies.

(2.5) Non parametric significance tests. For a wide variety of models, it is common to assume that the intensity function takes the form $\widehat{\lambda}(\cdot) = h(\mathbf{c}\beta) = h(\sum_{i=1}^{k} c_i\beta_i)$, where h is a link function. Assume now that we have a point process given by $N = (t,\mathbf{x},\mathbf{c}) \in \mathbb{R}^{1+n+k}$. Here we are interested in testing the significance of the j-th covariate, this is, we want to test the null hypothesis $H_o : \beta_j = 0$ against the alternative $H_a : \beta_j \neq 0$. Let $\widehat{\lambda}^F(t,\mathbf{x},\mathbf{c}|H_t)$ denote the CIF estimator under the alternative hypothesis (full model), and $\widehat{\lambda}^R(t,\mathbf{x},\mathbf{c}|H_t) = \widehat{\lambda}_1(t,\mathbf{x}|H_t)\widehat{\lambda}_2(\mathbf{c})$ the corresponding reduced model under the null hypothesis. Then, the discrepancies between $\widehat{\lambda}^F$ and $\widehat{\lambda}^R$ can be used to test the null hypothesis. We present in the next section some nonparametric tests of significance for spatial or spatio-temporal covariates.

Consider the case of interest of an unmarked spatio-temporal point process with covariates, $N = \{t, \mathbf{x}, \mathbf{c}\}$, and with corresponding CIF $\lambda(t, \mathbf{x}, \mathbf{c}|H_t)$ defining the limiting expected rate of occurrence of points per space-time-covariate volume conditional on the history of the process prior to time t. Schoenberg Schoenberg (2004) proposed several nonparametric test statistics, some based on squared discrepancies (2.19) and others based on likelihood ratios (2.20) of the following form

$$S_3 = \int_0^T \int_{\mathbb{R}^n} \int_{\mathbb{R}^k} [\hat{\lambda}^F(t, \mathbf{x}, \mathbf{c}|H_t) - \hat{\lambda}^R(t, \mathbf{x}, \mathbf{c}|H_t)]^2 d\mathbf{c} d\mathbf{x} dt \tag{2.19}$$

$$S_4 = \int_\chi [\log\{\hat{\lambda}^F(t, \mathbf{x}, \mathbf{c}|H_t)\} - \log\{\hat{\lambda}^R(t, \mathbf{x}, \mathbf{c}|H_t)\}] dN - \int_0^T \int_{\mathbb{R}^n} \int_{\mathbb{R}^k} [\hat{\lambda}^F(t, \mathbf{x}, \mathbf{c}|H_t) - \hat{\lambda}^R(t, \mathbf{x}, \mathbf{c}|H_t)] d\mathbf{c} d\mathbf{x} dt \tag{2.20}$$

A natural estimator for $f(t, \mathbf{x}, \mathbf{c})$ is given in its general form by

$$\hat{p} = \frac{\hat{\lambda}(t, \mathbf{x}, \mathbf{c}|H_t)}{\sum \hat{\lambda}(t, \mathbf{x}, \mathbf{c}|H_t)},$$

which is evaluated for each event and using both expressions for the CIFs, one under the null ($\hat{\lambda}^R$) and the other under the alternative hypothesis ($\hat{\lambda}^F$). We thus have for each $i = 1, \ldots, np$

$$\hat{p}_i^R = \frac{\hat{\lambda}^R(t_i, \mathbf{x}_i, \mathbf{c}_i|H_t)}{\sum_j \hat{\lambda}^R(t_j, \mathbf{x}_j, \mathbf{c}_j|H_t)} \tag{2.21}$$

and

$$\hat{p}_i^F = \frac{\hat{\lambda}^F(t_i, \mathbf{x}_i, \mathbf{c}_i|H_t)}{\sum_j \hat{\lambda}^F(t_j, \mathbf{x}_j, \mathbf{c}_j|H_t)} \tag{2.22}$$

Díaz-Avalos et al. (2013b) proposed the use of the Kullback-Leibler measure (KL) and the Hellinger (H) distance to measure the difference between the two estimated densities. Under the null hypothesis $H_o : \beta_j = 0$, $\hat{p}^R \equiv \hat{p}^F$, and the Kullback-Leibler measure can be defined as

$$KL = \sum_{i=1}^{np} \log\left(\frac{\hat{p}_i^F}{\hat{p}_i^R}\right) * \hat{p}_i^F \tag{2.23}$$

whilst the Hellinger distance is given by

$$H = \sum_{i=1}^{np} \left\{\left(\sqrt{\hat{p}_i^F} - \sqrt{\hat{p}_i^R}\right)^2\right\}. \tag{2.24}$$

The statistical properties of these two tests are analyzed through intensive simulation studies are shown in Díaz-Avalos et al. (2013b).

In all cases, we considered a trend given by intensity functions of the form

$$\log(\lambda(t, \mathbf{x}, \mathbf{c})) = \beta_0 + \beta_1 c_1(t, \mathbf{x}) + \cdots + \beta_k c_k(t, \mathbf{x}). \tag{2.25}$$

In particular, we restricted to planar spatial coordinates so that $\mathbf{x} = (x, y)$, and considered only three covariates (then $k = 3$) with coefficients β_0, β_1, and β_2. The first covariate, c_1, was built as the exponential of an isotropic Gaussian random

field in $[0,1]^2$ with zero mean and covariance with range 0.3, and C(0)=1.0. The second covariate, c_2, was obtained discretizing the exponential of the previous isotropic Gaussian random field in $[0,1]^2$, dividing it into four categories. Finally, the third covariate, c_3, was obtained as a non-linear transformation of $Cov1$. The first covariate was assumed to vary over time (in $[0,1]$) in a linear way, while keeping constant in time the second and third covariates.

Díaz-Avalos et al. (2013b) also fitted the corresponding intensity model and obtained the corresponding β values, obtaining information on their empirical distribution. They tested the hypothesis of normality using the Anderson-Darling test. To save space, we only show and compare the results coming from the new test based on KL with the previously proposed in the literature based on S_3. In their paper, Díaz-Avalos et al. (2013b) showed that the new test H provides similar results than KL, and S_4 is quite similar (although less powerful) than S_3. As an illustration, the empirical distributions of S_3 and KL for the Poisson and Gibbs structures, different number of events and β_1 values are shown in Figures 7 and 8. It is a general case, as expected, that the empirical distribution of the test statistics under the null hypothesis (in black) takes lower values, much closer to zero than those values coming from the test statistics under the alternative hypothesis (in red). In all cases we note that as β_1 increases, for a fixed sample size, the separation of the empirical distribution of the test statistics becomes more evident. This separation behaves clearly different for each test statistic and point structure. For Poisson and Gibbs models, the separation can be noted at very small values of β coefficient (as small as $\beta_1 = 0.5$), whereas in the case of Neymann-Scott processes, the separation is only detected at higher values of β coefficient.

Regarding the number of points in the pattern, increasing such number facilitates the gap or separation between both empirical distributions when considering inhomogeneous Poisson and Gibbs models. In the particular case of Neymann-Scott models, this separation as the number of points increases can be noticed only for high values of β_1. This reinforces the fact that in the presence of clustered structures testing for significance of a covariate has to be taken with more care. The performance of the test statistics S_3, S_4, KL, and H under the three distinct point structures and samples sizes is also summarized in terms of empirical $p-values$ (see Table 1). For inhomogeneous Poisson and Gibbs process, all the statistics perform in a similar fashion, except S_4, which is only able to reject the null model at big sample sizes. For cluster Neyman-Scott processes all the statistics can only reject the null model at high values of np and when the corresponding β_1 parameter is high enough.

In summary, in all cases we can see that as β_1 increases, for a fixed sample size, the separation of the empirical distributions of the test statistics under the null and alternative hypothesis becomes more evident. This separation is different for each statistic. In addition, for Poisson and Gibbs models, the separation can be noted at very small values of β parameter, while for clustered patterns the separation is detected at higher values of β. This gives evidence that in these cases, it is more difficult to detect significance of the covariates.

Table 2 shows the results of the Anderson-Darling Normality test applied to the estimated values of β_1 parameter. This test has been applied when considering the three point structures, three samples sizes and considering $\beta_1 = 0.0, 0.5, 2.0$.

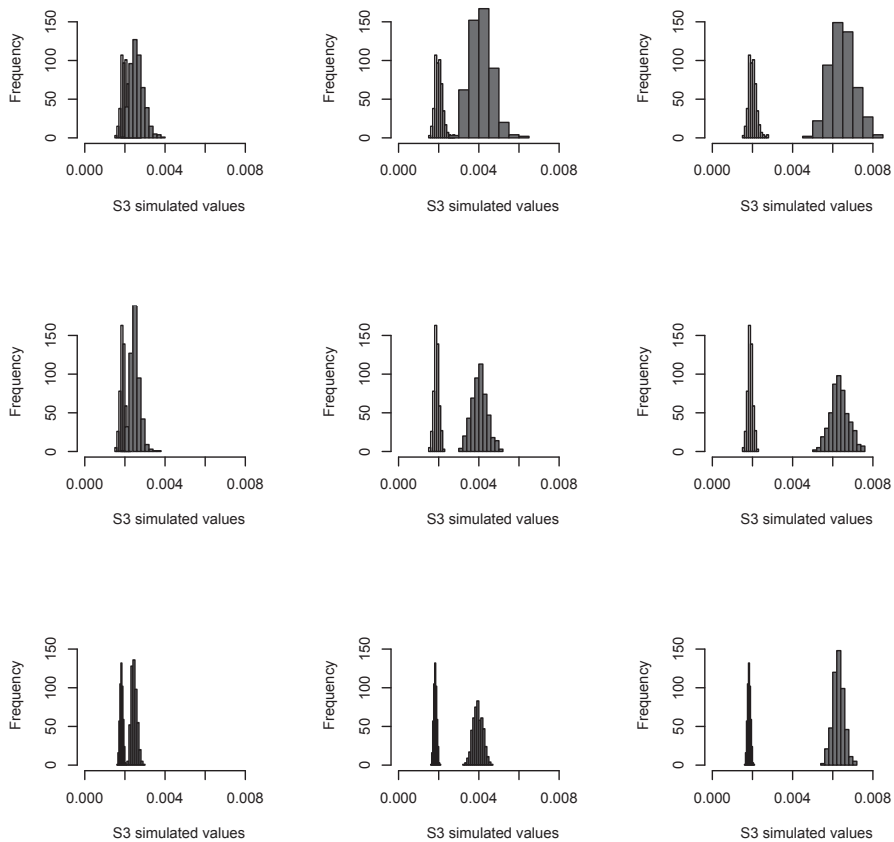


Figure 7. Empirical distribution of S_3 statistic under Inhomogeneous Poisson structures when testing the significance of the continuous covariate $Cov1$: *Left*: $\beta_1 = 0.5$, *Center*: $\beta_1 = 1.0$ and *Right*: $\beta_1 = 2.0$. First line: $n = 100$, Second line: $n = 200$, and Third line: $n = 500$. Empirical distribution of the test statistics under the null hypothesis (in black) and under the alternative hypothesis (in red).

The normality test is rejected in specific cases and it does not show a consistent behavior. For example, normality is rejected only for $\beta_1 = 2$ and $n = 100$ for the inhomogenous Poisson case, and for $\beta_1 = 0$ and all n in the case of the Gibbs model. For this last case, when $\beta_1 = 2.0$ and $n = 200$ we are close to rejection. Part of this inconsistency may be explained because as β_1 increases, the intensity function gives more weight to c_1, increasing the density of points in a smaller area. This suggests that the assumption of normality is not related to the number of points in D but to the density of points in areas where λ takes high values. The results in Table 2 show that the hypothesis of asymptotic normality has to be taken with caution. To reinforce these results, Figures 9-11 show the histograms of the empirical distributions of β_1 and the corresponding theoretical quantiles.

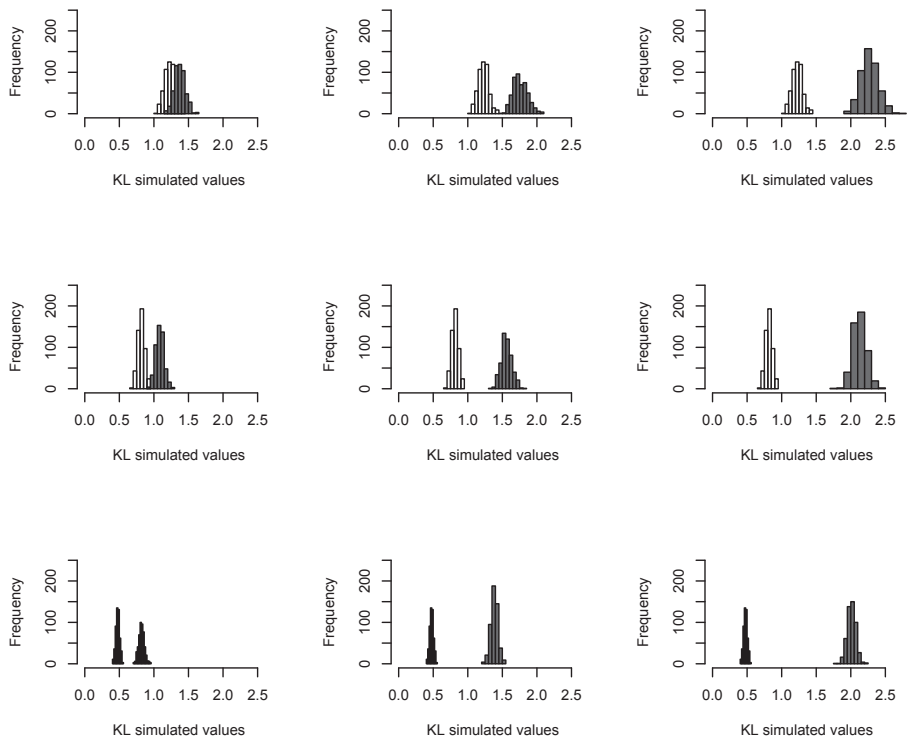


Figure 8. Empirical distribution of KL statistic under Inhomogeneous Gibbs structures when testing the significance of the continuous covariate $Cov1$: *Left*: $\beta_1 = 0.5$, *Center*: $\beta_1 = 1.0$ and *Right*: $\beta_1 = 2.0$. First line: $n = 100$, Second line: $n = 200$, and Third line: $n = 500$. Empirical distribution of the test statistics under the null hypothesis (in black) and under the alternative hypothesis (in red).

3. Application: Lightning ignitions in a Spanish province

To illustrate the application of the significance tests to real data, we use the test statistics presented in this paper to analyze the trend behavior of the spatial patterns produced by wildfire incidences in Castellón (Spain) (see Figure 12), during the years 2001 to 2006. Figure 12 also shows the image of the elevation in this zone, elevation that clearly decreases from West to East as we approach the sea. Testing the significance of this covariate in the spatio-temporal intensity function is shown in this paper.

We fitted and considered a space-time model exhibiting both a degree of spatial inhomogeneity that takes into account the trend in terms of covariates and an infinite order of interactions, as well as a seasonal time trend depending on a single covariate. These models are the so-called inhomogeneous Area-Interaction processes (Baddeley and van Lieshout, 1995). We considered three spatial covariates (Aspect, Elevation and Slope) affecting the spatial trend and daily average

	IP			NS			Gibbs		
S3	0.4713	0.4820	0.4825	0.4629	0.5519	0.8203	0.510	0.495	0.405
	0.0279	0.0035	0	0.6356	0.6461	0.7285	0.0074	0	0
	0	0	0	0.247	0.0966	0.0259	0	0	0
KL	0.4927	0.4857	0.4765	0.3737	0.4096	0.6604	0.514	0.490	0.501
	0.0071	0.0003	0	0.4934	0.4357	0.4244	0.0791	0.0001	0
	0	0	0	0	0.058	0.0208	0	0	0
H	0.4860	0.4827	0.5627	0.4426	0.4917	0.7529	0.511	0.493	0.493
	0.0188	0.0018	0	0.6049	0.5815	0.5933	0.051	0.0001	0
	0	0	0	0.2627	0.1256	0.0371	0	0	0
S4	0.4785	0.4984	0.4144	0.4962	0.5035	0.5014	0.526	0.585	0.560
	0.7592	0.3884	0.5464	0.6088	0.5529	0.3418	0.009	0	0
	0.2656	0	0	0.2627	0.1256	0.0371	0	0	0

Table 1. p-values when testing significance of the covariate $Cov1$ with the test statistics S_3, S_4, KL, and H under inhomogeneous Poisson (IP), Neymann-Scott (NS) and Gibbs models. For each statistic, rows correspond to values of β_1 $(0.0, 0.5, 2.0)$, and columns correspond to different sample sizes ($n = 100, 200, 500$).

Model	$n = 100$	$n = 200$	$n = 500$
IP	0.4308	0.0935	0.5979
	0.646	0.221	0.8481
	0.0397	0.3704	0.8678
NS	0.099	0.6119	0.2925
	0.8998	0.3805	0.7505
	0.1672	0.6046	0.05159
Gibbs	0.003965	0.03074	0.00191
	0.1243	0.512	0.4667
	0.1192	0.05599	0.1634

Table 2. Anderson-Darling Normality Test applied to the estimated values of β_1 parameter. Each row corresponds to $\beta_1 = 0.0, 0.5, 2.0$.

temperature as the covariate affecting the time trend. Slope is the steepness or degree of incline of a surface. Aspect is the orientation of the slope, measured clockwise in degrees from 0 to 360, where 0 is north-facing, 90 is east-facing, 180 is south-facing, and 270 is west-facing. Finally the altitude was considered as the elevation above sea level, and measured in meters.

We considered a separable model with intensity function of the form

$$\lambda(x,t) = \lambda_1(x)\lambda_2(t)$$

Both λ_1 and λ_2 are assumed to be densities in the spatial and temporal windows respectively. The multiplicative structure of the model assumed allows to estimate the covariate coefficients of λ_1 and λ_2 separately. We used the R library *Spatstat* to estimate the coefficients of the spatial component of the model, and the R *glm* function to estimate the coefficients of the temporal component. Given the

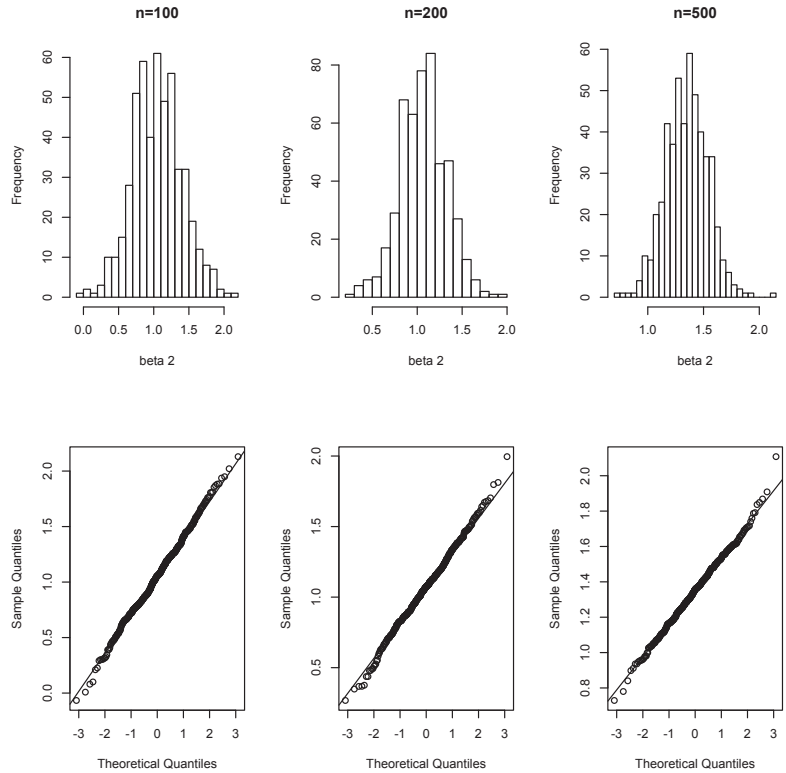


Figure 9. First row: Histograms of estimated values of β_1 under an inhomogeneous Poisson model obtained with $\beta_1 = 2.0$ in the trend term. Second row: Corresponding normal quantile plots for $n = 100$ (first column), $n = 200$ (second column) and $n = 500$ (third column).

seasonal pattern of fires in Castellón, we considered only models with a temporal component of the form

$$\lambda_2(t) = \alpha_0 + \alpha_1 T + \alpha_2 T^2 + \alpha_3 T^3 + \alpha_4 \sin(t) + \alpha_5 \cos(t) + \alpha_6 \sin^2(t) + \alpha_7 \cos^2(t) \tag{3.1}$$

with T denoting the average daily temperature in the province of Castellón. The estimates for the α coefficients are shown in Table 3. Given that most of the terms in the model assumed for $\lambda_2(t)$ lack a physical interpretation, for this temporal component of the intensity function we did not test for significance. Also, because we are more interested in testing the effect of the spatial covariates. The models we considered for $\lambda_1(\mathbf{x})$ are shown in Table 4. We chose a hierarchical structure for the spatial trend models in order to illustrate what a sequential modeling process would be.

The empirical distribution for the differences between the statistical test KL as well as the empirical quantiles are shown in Figure 13. The left column of plots shows the differences of $KL_o - KL_1$, i.e the differences after the inclusion of the elevation as a covariate in the spatial component of the intensity function.

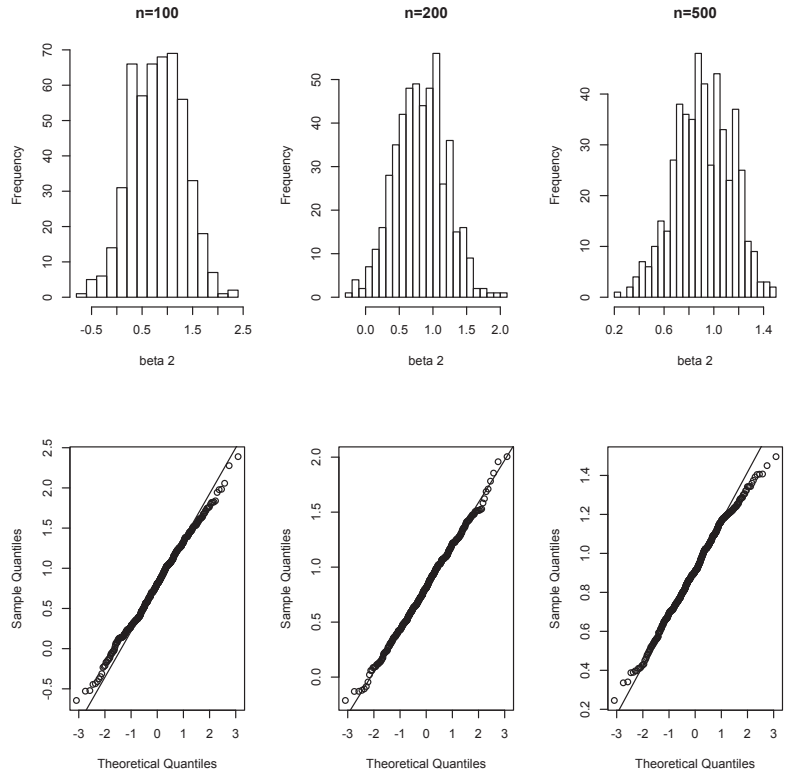


Figure 10. First row: Histograms of estimated values of β_1 under a Newmann-Scott model obtained with $\beta_1 = 2.0$ in the trend term. Second row: Corresponding normal quantile plots for $n = 100$ (first column), $n = 200$ (second column) and $n = 500$ (third column).

Coefficient	Estimate
α_0	-5.982
α_1	1.662e
α_2	-0.2385
α_3	0.01537
α_4	0.2830
α_5	-0.6244
α_6	0.3873
α_7	0.05248

Table 3. Coefficient estimates for the time component of the intensity function for the fire occurrences in the province of Castellón, Spain, for the years 2001-2006.

The center and the right columns of plots show the differences $KL_1 - KL_2$ and $KL_2 - KL_3$, respectively.

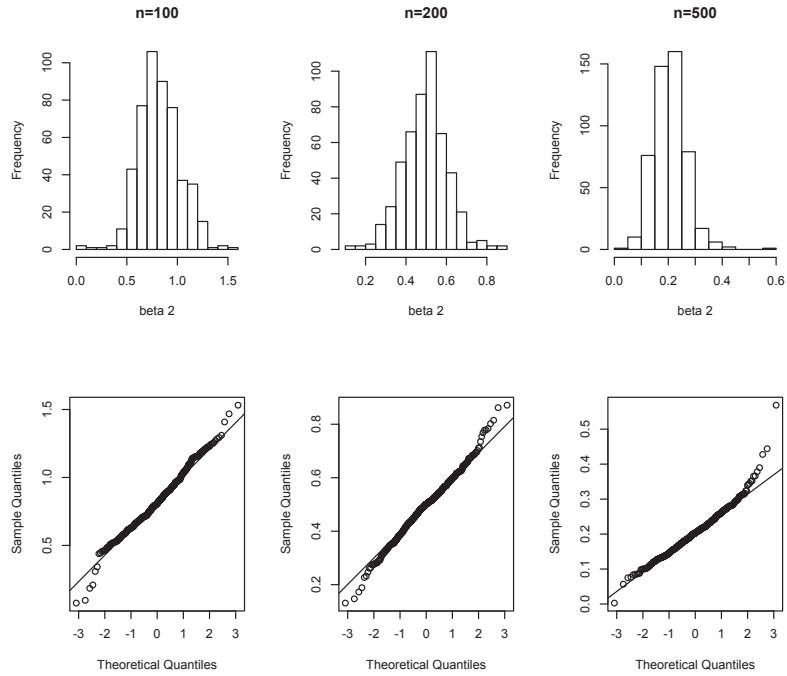


Figure 11. First row: Histograms of estimated values of β_1 under a Gibbs model obtained from the superposition of Strauss processes with $\beta_1 = 2.0$ in the trend term. Second row: Corresponding normal quantile plots for $n = 100$ (first column), $n = 200$ (second column) and $n = 500$ (third column).

Code	Model	β_0	β_1	β_2	β_3
0	Constant	7.268			
1	Elevation	7.722	-0.00068		
2	Elevation+Orientation	7.681	-0.00069	0.0003	
3	Elevation+Orientation+Slope	7.679	-0.00069	0.0003	0.0003

Table 4. Fitted models and parameter estimates for the trend under an inhomogeneous Area-Interaction model for Castellón fires.

From the histograms we note that there is a significant difference when Elevation is included in the model for $\lambda_1(x)$, but that inclusion of Elevation and Slope does not result in a significant improvement in terms of KL. The null model is rejected (empirical $p - value < 0.0116$), and we conclude that for the data set analyzed, Elevation explains a significant amount of the spatial variation of the occurrence of forest fires in Castellón. These results validate the field observations that most fires happen at low and intermediate elevations and that seasonal variation of average daily temperature is associated to the seasonal pattern of fire occurrences. A similar conclusion regarding significance of the covariates in the model for $\lambda_1(x)$ can be reached if the test statistic S4 is used ($p - value < 10^{-6}$).

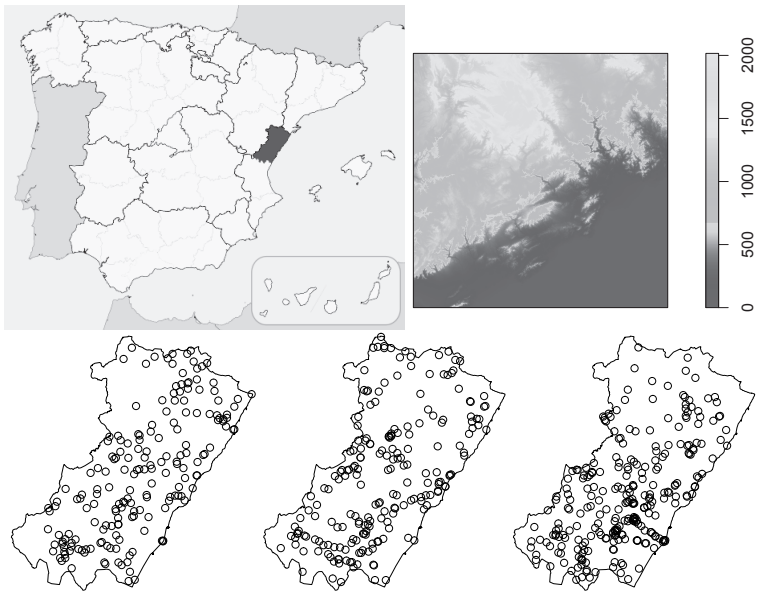


Figure 12. First row: Location of the provence of Castellón, Spain(left), and digital elevation map for the area that includes Castellón. Second row: wildfire occurrences in 2001-2002 (*left*), 2003-2004 (*center*) and 2005-2006 (*right*).

The empirical distribution for the differences $S3_i - S3_j$, $S4_i - S4_j$ and $H_i - H_j$ for $i = 0,1,2$ and $j = 1,2,3$ (not shown) as well as the corresponding normal quantile plots suggest that the differences have a normal-shape distribution, and in fact, the Anderson-Darling test for normality is not rejected in all cases. In all cases, we note that the addition of the Elevation to the model for $\lambda_1(x)$ is significant but the inclusion of Orientation and Slope does not provide a significant amount of information to explain the spatial variation of the observed fire occurrences in Castellón. The positive coefficient for Elevation in the model for the spatial intensity function indicates that the expected occurrence of fires will be higher with the Elevation.

A point of interest is the comparison between the different test statistics used to make inferences about the inclusion of spatial covariates to explain fire occurrences in Castellón. Table 5 shows the performance of the different test statistics in terms of the empirical $p-values$ for each added term in the model for $\log[\lambda_1(x)]$. As mentioned earlier, all the test statistics happen to detect that elevation is a significant term in the log intensity function model. However, the p-values in the first row of Table 5 show that the p-values of $S4$ and $S3$ are much smaller than those of KL and H distances, and therefore have a higher power for the added term tests.

4. Concluding Remarks

It is clear that the condition of separability in the modeling of multidimensional point processes cannot be overstated. It is typically extremely difficult to

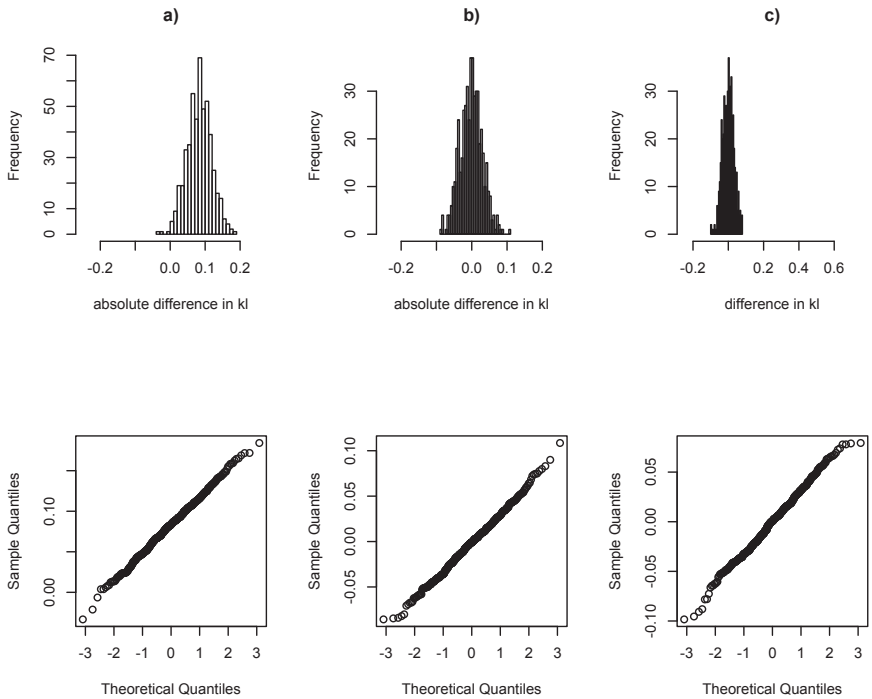


Figure 13. Histogram of estimated differences of KL values for different models for the spatial component $\lambda_1(x)$ of the intensity function , and corresponding normal quantile plots for Castellón fires.

Term	KL	$S4$	$S3$	Hellinger
β_1	0.0116	$< 10^{-6}$	$< 10^{-6}$	0.008
β_2	0.514	0.478	0.490	0.510
β_3	0.512	0.204	0.522	0.518

Table 5. Monte Carlo $p-values$ for the added terms in the log linear model for $\lambda_1(x)$, using the four different test statistics.

construct realistic models for multidimensional point processes in the presence of marks and with many covariates. Common practice looks for empirically-based models, and one method for constructing such a model would be to investigate individually the distribution of each coordinate, and the individual contribution to the conditional intensity of each, or perhaps small collections of, covariates. Under separability, these marginal distributions of the process could then be estimated separately, and the parametric forms for each could readily be inspected for goodness-of-fit.

As stated by Schoenberg (2006) and Schoenberg (2004), when separability may safely be assumed, not only is the derivation of a parametric form for the relationship between one covariate and the CIF greatly facilitated due to an effective

reduction of the dimensionality of the process, but in addition the parameters governing each component of the model may be estimated individually with consistency.

In this paper we are motivated by the fact that there was a need of new, simple and practical sepraability and significance tests under the case of marks and covariates in the context of spatio-temporal point processes, with indication and investigation of the power and other statistical properties of the tests. The inclusion of spatially and temporally varying covariates in the models for the intensity function is becoming of particular interest, and in this paper we have introduced different tests to study the significance of these covariates when playing a role in the intensity function of a point pattern. Testing the significance of covariates is important to explain which covariates have an effect in the spatial distribution of the point pattern observed. We opt here for the flexibility given by the conditional intensity function, and its nonparametric kernel-based estimation.

In this line, we have proposed new tests based on nonparametric estimation of the CIFs, we have developed an intensive simulation study to provide information on the statistical properties of our test, and have compared them with other existing tests. Our new tests are competitive with the existing ones, in particular under those cases of large dimensions. Note that since all the separability tests used here are nonparametric, such departures from nonseparability cannot be attributed to the lack of fit of a model used in the estimation of the contribution from a particular covariate.

Our approach calculates thinning probabilities under the conditions of presence and absence of a set of covariates and compares them through divergence measures. We approximate the statistical properties of our tests under a variety of practical scenarios.We also assess under which general conditions the asymptotic normality assumptions are adequate (Waagepetersen and Guan, 2009). An appealing characteristic of our approach is that it permits the use of covariates of any form, including both discrete and continuous, and that the model fitted is not restricted to have an intensity function of a lineal form. This allows the use of wider classes of trend functions in the process of modeling point patterns regardless of the application field. The conceptual simplicity of our approach makes it accessible to users with only a basic statistical background.

The differences between KL, S4, S3 and H distances for each added term may be thought of as an analogous of the extra sum of squares used to test the significance of an extra term in an ordinary linear model. Except for S3, such differences are not related to squared discrepancies between two models, but in neither case there is a warranty that the empirical distributions of the differences converge fast to a normal distribution. The use of Monte Carlo tests is widely justified in nonparametric situations (Besag and Diggle, 1977), thus justifying its use in the analysis of the fire patterns in Castellón. The use of Monte Carlo tests for statistical inference in spatial point patterns is not a novel idea. Recent use of the Monte Carlo approach for point processes inference are described in Diggle et al. (2005) and in Kelshall and Diggle (1998), but to the best of our knowledge, such tests have not been used in the context of significance tests for the inclusion of covariates.

The only restriction for the use of the method proposed here is the possibility to simulate point patterns from the parametric model proposed. Simulation of spatial point processes, however, has advanced considerably in the last decade, and today it is possible to simulate many complicated models (Møller and Waagepetersen (2003); Møller and Waagepetersen (2004); Møller and Rasmussen (2005). Although it is not a particular goal here, it would be wise to depict optimal bandwidth estimation, a long standing problem in nonparametric statistical estimation. Our study is also related to obtain optimal subsets of covariates when the number of present covariates is large. This is yet an open problem. Further research is also needed related to the analysis of the performance of our tests under large dimensions, when combining covariates, marks and very large sample sizes. In any case we should note that our approach is completely general and able to run under highly complicated cases combining times, locations, marks and covariates. The mathematical treatment and formulation is thus set for all such cases. The more practical problem we may face is the goodness-of-fit of the kernel estimate. Having complicated combinations of all those terms increases the number of dimensions and thus the dimension of the kernels involved in the tests. We in fact anticipate certain problems in dealing with the appropriate kernel function and with the election of the bandwidth parameter.

Received June 22, 2013

Final version received October 24, 2013

CARLOS DÍAZ-AVALOS
INSTITUTO DE INVESTIGACIONES EN
MATEMÁTICAS APLICADAS Y EN SISTEMAS,
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO,
APARTADO POSTAL 20-726
DEL. ALVARO OBREGÓN,
MÉXICO D.F., C.P. 0100.
carlos@sigma.imas.unam.mx

REFERENCES

[1] R. ASSUNÇAO AND A. MAIA, *A note on testing separability in spatial-temporal-marked point processes*, Biometrics **63**, (2007), 290–294.

[2] A. BADDELEY AND M.N.M. VAN LIESHOUT, *Area-interaction point processes*, Ann. Inst. Statist. Math. **47**, (1995), 601–619.

[3] C-H. CHANG AND F. P. SCHOENBERG, *Testing separability in marked multidimensional point processes with covariates*, Annals of the Institute of Statistical Mathematics **63**, (6), (2011), 1103–1122.

[4] C. COMAS, AND J. MATEU, *Statistical inference for Gibbs point processes based on field observations*, Stoch Environ Res Risk Assess **25**, (2011), 287–300.

[5] C. COMAS, M. PALAHI, T. PUKKALA, AND J. MATEU, *Characterising forest spatial structure through inhomogeneous second order characteristics*, Stoch Environ Res Risk Assess **23**, (2009), 387–397.

[6] N.A.C. CRESSIE, Statistics for Spatial Data (Revised ed.). New York, Wiley, 1993.

[7] C. DÍAZ-AVALOS, D. L. PETERSON, C. E. ALVARADO, S. A. FERGUSON, AND J. E. BESAG, *Space-time modeling of lightning-caused forest fires in the Blue Mountains, Oregon*, Canadian Journal of Forest Research **31**(9), (2001), 1579–1593.

[8] C. DIAZ-AVALOS, P. JUAN, AND J. MATEU, *Similarity measures of conditional intensity functions to test separability in multidimensional point processes*, Stochastic Environmental Research and Risk Assessment 27 (5), (2013a), 1193–1205.

[9] C. DIAZ-AVALOS, P. JUAN, AND J. MATEU, *Significance tests for covariate-dependent trends in inhomogeneous spatio-temporal point processes. Stochastic Environmental and Risk Assessment*, DOI 10.1007/s00477-013-0775-1, (2013b).

[10] D. DALEY AND D. VERE-JONES, An Introduction to the Theory of Point Processes, 2nd edition, Volume I. Springer, New York, 2003.

[11] P. J. DIGGLE, Statistical Analysis of Spatial Point Patterns, (2nd ed) Arnold, London UK, 2003.

[12] P. J. DIGGLE, P. ZHENG, AND P. DURR, *Nonparametric estimation of spatial segregation in a multivariate point process: bovine tuberculosis in Cornwall, UK*, Appl. Statist. **54** (3), (2005), 645–658.

[13] A. E. GELFAND, P. J. DIGGLE, P. GUTTORP, AND M. FUENTES, Handbook of Spatial Statistics, CRC Press, Boca Raton, 2010.

[14] P. GUTTORP, Stochastic modeling of scientific data, London: Chapman - Hall, 1995.

[15] P. JUAN, J. MATEU, AND M. SAEZ, *Pinpointing spatio-temporal interactions in wildfire patterns*, Stoch Environ Res Risk Assess, DOI: 10.1007/s00477-012-0568-y, (2012).

[16] J. E. KELSHALL AND P. J. DIGGLE, *Spatial variation in risk of disease: a non parametric binary regression approach*, Appl. Stat. **47**, (1998), 539–573.

[17] J. MØLLER AND C. DÍAZ-AVALOS *Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires*, Scandinavian Journal of Statistics **37**, (2010), 2–15.

[18] J. MØLLER AND J. G. RASMUSSEN, *Perfect simulation of Hawkes processes*, Advances in Applied Probability **37**, (2005), 629–646.

[19] J. MØLLER AND R. P. WAAGEPETERSEN, *An introduction to simulation-based inference for spatial point processes* In J. Møller, Spatial Statistics and Computational Methods, Lecture Notes in Statistics 173, Springer-Verlag, New York, (2003), 143–198.

[20] J. MØLLER AND R. P. WAAGEPETERSEN, Statistical Inference and Simulation for Spatial Point Processes, Chapman and Hall/CRC, Boca Raton, 2004.

[21] P. N. OMI, Forest Fires: A Reference Handbook, ABC-CLIO, Santa Barbara, CA, xviii, 2005.

[22] B. D. RIPLEY, *The second-order analysis of stationary point processes*, Journal of Applied Probability **13**, (1976), 256–266.

[23] F. P. SCHOENBERG, D. R. BRILLINGER, AND P. M. GUTTORP, *Point processes, spatialtemporal*, In Encyclopedia of Environmetrics, Volume 3, A. El-Schaarawi and W. Piegorsch (eds), 1573–1577. Wiley, New York, (2002).

[24] F. P. SCHOENBERG, *Testing separability in multi-dimensional point processes*, Biometrics **60**, (2004), 471–481.

[25] F. P. SCHOENBERG, A note on the separability of multidimensional point processes with covariates. UCLA Preprint Series, No.496.

[26] B. W. SILVERMAN, Density Estimation for Statistics and Data Analysis. Chapman and Hall, London, 1986.

[27] R. WAAGEPETERSEN, *An estimating function approach to inference for inhomogeneous Neyman-Scott processes*, Biometrics **63**, (2007), 252–258.

[28] R. WAAGEPETERSEN AND Y. GUAN, *Two-step estimation for inhomogeneous spatial point processes*, Journal of the Royal Statistical Society: Series B **71**, (2009), 685–702.

Bol. Soc. Mat. Mexicana (3) Vol. 19, 2013

ANALYSIS OF REPEATED EVENTS AND PANEL COUNT DATA

E. JUAREZ-COLUNGA AND C. B. DEAN

ABSTRACT. Recurrent event data arise when there is an event that may repeat over time. This type of data is common in many areas of the sciences, social sciences, medicine and elsewhere. In this article we review counting process models, in particular, Poisson and mixed Poisson models for recurrent event data, as well as models for aggregated data, where only the number of events occurring in intervals of time, called *panel data*, are available. Estimation is conducted through likelihood and estimating equations. We illustrate the methods through an example of recurrence of superficial bladder cancer tumors in a clinical trial. We also briefly discuss optimal recurrent event designs.

1. Introduction

In the last few decades there has been great interest in monitoring longitudinal processes. One type of longitudinal process arises when there is an event that may repeat over time; data arising from such a process are called *recurrent event data*, and are common in many fields. For example, in business, interest may be in monitoring insurance claims, in engineering, in monitoring lines of faulty code, and in medical research, in studying the occurrence of adenomas in the colon, or the occurrence of superficial bladder cancer tumors.

An example considered in greater detail later typifies the sort of scenario through which recurrent data arise in clinical trials. This study, conducted by the Veterans Administrative Co-operative Urological Research Group, investigated the effects of placebo pills, pyridoxine pills, and periodic instillation of thiotepa into the bladder on the frequency of recurrence of bladder cancer [2]. The data appear in [1]. All 116 patients had bladder cancer when they entered the study; the tumors were removed and the patients were randomly assigned to one of the three treatments. Two covariates were considered which may reflect cancer severity at baseline: the number of tumors and the size, in centimetres, of the largest tumor. Figure 1 displays the recurrences for the three treatment groups.

One of the main classifications of recurrent event analyses is based on whether the focus is on analyzing the timing between events, or the counts of events. Both are based on counting processes. While the first scenario focuses on modeling the hazard between events, the second develops models for the intensity, or the mean number of events. When the system is sharply different each time a new event occurs, either because of the effect of the event itself, or the effect of the interventions employed after events occur, other models than the ones considered here are utilized. In these cases, time to first event, then time between first and second

2010 *Mathematics Subject Classification*: 62M86, 62N86, 62P10.

Keywords and phrases: clinical trial, counting process, design, life-history data, Poisson regression.

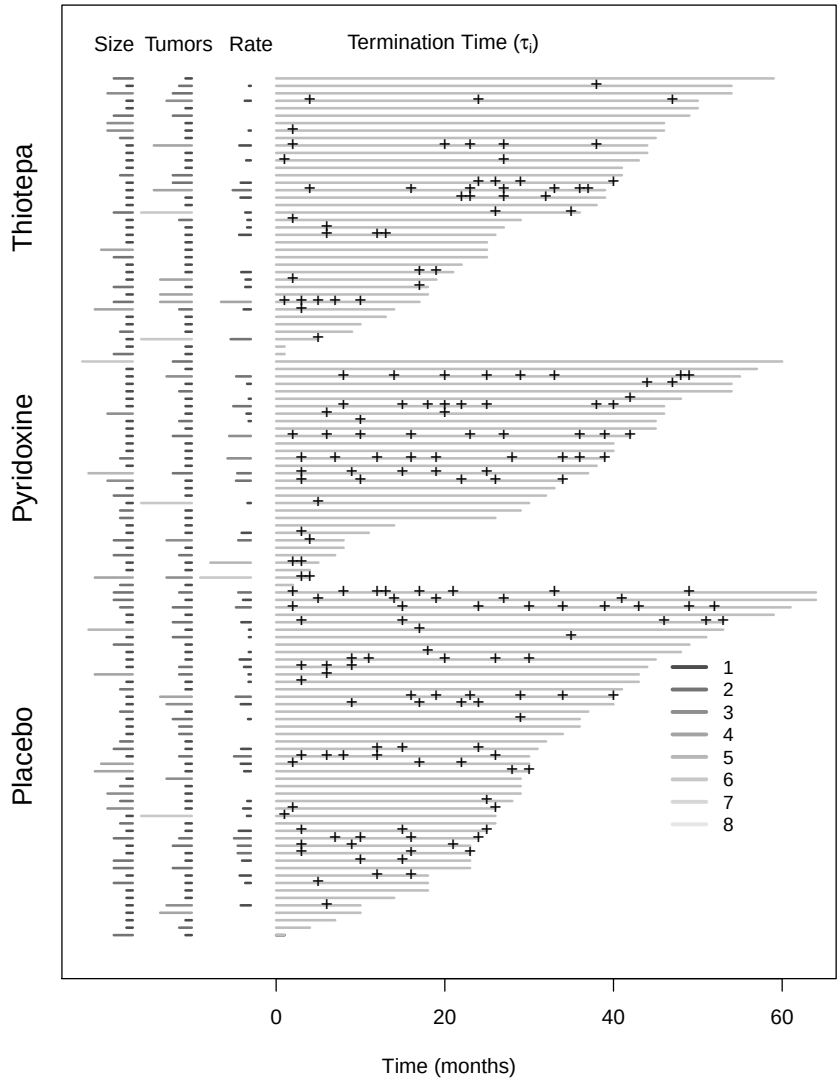


Figure 1. Bladder cancer data: the lines in the plot display the observation times ending at τ_i, termination times, for individuals, while the '+' indicates event times. For the display of the covariates (size and tumors) and individual-specific rates, the length of the lines are proportional to the size of the largest tumor, the number of tumors, and the rate of events, respectively; additionally, lighter colors are used to highlight large values of these variables.

event, etc., may be considered in separate models. Cook and Lawless [3] provides a comprehensive source on models and methods for recurrent event analyses.

Often, in order to monitor longitudinal processes, individuals are examined at specific followup times, and only the number of events that occurred between followup visits are recorded; these give rise to what is termed *panel data*. In this

article we focus on inference for both recurrent event data and panel data based on counting processes. We discuss estimation of recurrent event models through likelihood and estimating equations. Estimating equations provide a robust estimation approach. In Section 2 we introduce the counting process likelihood, as well as Poisson and mixed Poisson models, for both recurrent and panel count data. In Section 3 we discuss estimation of recurrent event models, while in Section 4 we illustrate these methods. We close with a discussion of other considerations in the analysis of recurrent event data in Section 5. In this section we outline a key result regarding optimal design of panel data studies, building on the robust approach to estimation developed in this paper using estimating equations.

2. Models Based on Counting Processes

(2.1) Counting Process Likelihood. Models for recurrent events are typically formulated through their intensity functions, which are the probability distribution of the events occurring in a small interval of time $[t, t+\delta t)$, given the history of events up to time t [3]. Let $\{N(t), t \geq 0\}$ be the right-continuous counting process that records the number of events for an individual over the interval $[0, t]$; we define $N(0) = 0$ for simplicity. If $H(t) = \{N(s) : 0 \leq s < t\}$ represents the history if the process up to time t, then the intensity process function is defined as

$$\lambda(t|H(t)) = \lim_{\Delta t \to 0} \frac{\Pr\{\Delta N(t) = 1 | H(t)\}}{\Delta t}. \tag{2.1}$$

It is assumed that the probability of two or more events occurring over the interval $[t, t+\Delta t)$ is $o(\Delta t)$, so $\Pr(\Delta N(t) = 1|H(t)) = \lambda(t|H(t))\Delta t + o(\Delta t)$, and $\Pr(\Delta N(t) = 0|H(t)) = 1 - \lambda(t|H(t))\Delta t + o(\Delta t)$. The likelihood, obtained via product integration [7], based on n event times $0 = t_0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq \tau$ occurring during the time $[0, \tau]$ for fixed τ is

$$L = \left[\prod_{j=1}^{n} \lambda(t_j|H(t)) \exp\left(-\int_{t_{j-1}}^{t_j} \lambda(u|H(u))du\right)\right] \exp\left(-\int_{t_n}^{\tau} \lambda(u)du\right) \tag{2.2}$$

$$= \left[\prod_{j=1}^{n} \lambda(t_j|H(t))\right] \exp\left(-\int_{0}^{\tau} \lambda(u|H(u))du\right). \tag{2.3}$$

Maximum likelihood methods can then be employed for inference by invoking usual asymptotic theory.

The definition of the intensity function in (2.1) can be extended to include covariate information up to and including time t through an extended definition of the history of the process, specifically $H(t) = \{N(s) : 0 \leq s < t; \boldsymbol{x}(s) : 0 \leq s \leq t\}$, where $\boldsymbol{x}(s)$ is the covariate vector at time s.

(2.2) Poisson and Mixed Poisson Models. The specification of the model through the intensity function in (2.1) is very general, but in practice assumptions are often made regarding the history of the process based on scientific input and these may aid estimation. Poisson processes are often utilized because of their memoryless property; they simplify the history to only that at time t, i.e. the instantaneous probability of the occurrence of a new event in a small window

of time depends on the history only through t. In this case, the intensity has the following form:

$$\lambda(t|H(t)) = h(t), \tag{2.4}$$

which implies that the number of events $N(s,t)$ in a time interval $(s,t]$ follows a Poisson distribution, i.e.

$$Pr\{N(t) - N(s) = n\} = \frac{\left\{\int_s^t \lambda(u)du\right\}^n \exp\left\{-\int_s^t \lambda(u)du\right\}}{n!}, \tag{2.5}$$

where the mean $E(N(s,t))$ is $\mu(s,t) = \int_s^t \lambda(u)du$. The expected cumulative number of events occurring from time 0 to time t is known as the cumulative intensity function or cumulative mean function, and has the form

$$\mu(t) = \int_0^t \lambda(u)du. \tag{2.6}$$

Another important property of the Poisson process is that the number of events occurring in disjoint intervals are independent random variables.

Since under a Poisson model the intensity function does not depend on the history of the process, we have $\lambda(t|H(t))dt = h(t)dt = E[dN(t)]$ (see Equation (2.4)), interpreted as both a conditional and a marginal probability, or an expectation. The function $h(t)$ is sometimes referred to as the rate function.

Covariate information is in practice commonly included in a multiplicative form as

$$\lambda(t|H(t)) = \rho(t)g(\boldsymbol{x}(t);\boldsymbol{\beta}), \tag{2.7}$$

where $\rho(t)$ is known as the baseline intensity function, and it may modeled nonparametrically or parametrically. Here we will focus on parametric forms through a parameter $\boldsymbol{\alpha}$ of dimension d_α, i.e. $\rho(t) = \rho(t;\boldsymbol{\alpha})$; forms such as the exponential ($\exp(\alpha t)$) and Weibull ($\alpha t^{\alpha-1}$) are often used. The function $g(\boldsymbol{x}(t);\boldsymbol{\beta})$ is often specified as $g(\boldsymbol{x}(t);\boldsymbol{\beta}) = \exp(\boldsymbol{x}'(t)\boldsymbol{\beta})$ to ensure $\lambda(t|H(t))$ is positive, and in this case, the coefficients $\boldsymbol{\beta}$ may be interpreted as log-relative risks. For simplicity in presentation, we will consider time-independent covariates, i.e. $\lambda(t;\boldsymbol{x}) = \rho(t;\boldsymbol{\alpha})\exp(\boldsymbol{x}'\boldsymbol{\beta})$.

Consider M individuals, and assume each individual is observed up to time τ_i referred to as the *termination time*, $i = 1,\ldots,M$. Let the observation process for the $i-$th individual $\{Y_i(t), t \geq 0\}$ be 1 if individual i is under study at t and 0 otherwise, and assume that $\{Y_i(t), t \geq 0\}$ is independent of the counting process $N_i(t)$. Thus the observed counting process may be defined as $\bar{N}_i(t) = \int_0^t Y_i(u)dN_i(u)$, $t \in [0,\tau_i]$ for individual i, $i = 1,\ldots,M$. The Poisson counting process has intensity $\lambda(t;\boldsymbol{x}_i) = \rho(t;\boldsymbol{\alpha})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})$, which implies that $\mathrm{E}(N_i(t)) = \mu_i(t) = R_i(t;\boldsymbol{\alpha})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})$, $i = 1,\ldots,M$, where $R_i(t;\boldsymbol{\alpha}) = \int_0^t Y_i(t)\rho(u;\boldsymbol{\alpha})du$. Let the total aggregated count of events for individual i be n_i, i.e. $N(\tau_i) = n_i$. For n_i events observed at times $\{t_{i1}, t_{i2},\ldots,t_{in_i}\}$, the likelihood for individual i has the following form [9]

$$L_i = \prod_{j=1}^{n_i} \lambda(t_{ij};\boldsymbol{x}_i)\exp\left\{-\int_0^{\tau_i}\lambda(u;\boldsymbol{x}_i)\right\} \tag{2.8}$$

$$= \left\{\prod_{j=1}^{n_i}\frac{\rho(t_{ij};\boldsymbol{\alpha})}{R_i(\tau_i;\boldsymbol{\alpha})}\right\} \times \left\{R_i(\tau_i;\boldsymbol{\alpha})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\right\}^{n_i}\exp\left\{-R_i(\tau_i;\boldsymbol{\alpha})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\right\}. \tag{2.9}$$

The first term in the likelihood (2.9) corresponds to the conditional distribution of the event times given the number of events, while the second is the likelihood kernel for the distribution of total number of events n_i. The Poisson model has the constraint that $\mathrm{Var}(N_i(\tau_i)) = \mathrm{E}(N_i(\tau_i)) = \mu_i(\tau_i)$, which is too limiting in many cases. Often, the variability observed exceeds what can be explained through covariates available. In these cases, is common to use a mixed Poisson model, in which the rate function for subject i is $\nu_i\lambda(t;\boldsymbol{x}_i)$ where ν_i's are independent positive random variables that follow a distribution $G(\nu;\phi)$ such that $\mathrm{Var}(\nu_i) = \phi$. The function $\lambda(t;\boldsymbol{x})$ is now interpreted as a population average rate function among subjects with covariate vector $\boldsymbol{x}$, since $\mathrm{E}(dN_i(t)|\boldsymbol{x}) = \lambda(t;\boldsymbol{x})dt$. As well, the random effect ν_i represents the effect of covariates which are unaccounted for in the regression model. Note that ν_i may also be a cluster effect, taking the same value for all individuals within the same cluster. This can be used to account for unknown hospital effects, for example, where patients are clustered within hospitals. When ν_i follows a Gamma distribution, the marginal distribution of n_i is negative binomial. In this case, the variance has the form $\mathrm{Var}(N_i(\tau_i) = \mu_i(\tau_i) + \mu_i^2(\tau_i)\phi$ accommodating extra-Poisson variation in the second term. In addition, if $s_1 < t_2$, $\mathrm{Cov}(N_i(t_1,s_1),N_i(t_2,s_2)) = \phi\mu_i(t_1,s_1)\mu_i(t_2,s_2)$. The parameter ϕ reflects both the degree of overdispersion and the degree of association between disjoint interval counts.

(2.3) Likelihood for Panel Count Data. Panel data arises when counts of events are recorded at specified followup times. Such data are quite common as continuous followup is expensive; when it is not ethically required, or seen to be too invasive given the clinical context, panel studies are employed. For example, studies investigating recurrence of superficial bladder cancer tumors commonly collect information in a panel followup framework; as an illustration, Vancouver Coastal Health in Canada is currently conducting a clinical trial on bladder cancer with followup every 3 months for the first two years, every 6 months for the next two years, and yearly thereafter [12].

Let the $e_i + 1$ individual-specific *panel followup times* be denoted by $T_{i,0} = 0 < T_{i,1} < T_{i,2} < \ldots < T_{i,e_i} = \tau_i$. Panel counts for individual i are denoted as $n_{ip} = \bar{N}_i(T_{i,p}) - \bar{N}_i(T_{i,p-1})$, $p = 1,2\ldots,e_i$, and the total aggregated count of events for individual i is $n_{i+} = \sum_{p=1}^{e_i} n_{ip}$.

The counting process $N_i(t)$ is again modeled here as a mixed Poisson process with intensity function $\lambda_i(t) = \nu_i\rho(t;\boldsymbol{\alpha})\exp(\boldsymbol{x}_i'\boldsymbol{\beta})$, given ν_i, an individual-specific random effect accounting for overdispersion. We set $\mathrm{E}(\nu_i) = 1$ without loss of generality, and let $\mathrm{var}(\nu_i) = \phi$. A particular scenario that may be of interest, as in the bladder cancer study, is where it is required to test treatment efficacy in reducing the number of recurrences. Let $\boldsymbol{x}_i$ be a $k\times 1$ treatment indicator vector for the i-th individual, such that $x_{i1} = 1$ represents an intercept term, and $x_{ij} = 1$ if individual i received treatment j, or 0 otherwise, $j = 2,\ldots,k$. Thus the corresponding β's are

parametrized such that the treatment effects are measured relative to treatment 1; β_1 reflects the overall mean, and $\boldsymbol{\alpha}$ describes the shape of the baseline intensity function $\rho(t,\boldsymbol{\alpha})$.

Writing the cumulative baseline intensity function for the entire followup time for individual i as $R_i = \int_0^{\tau_i} Y_i(t)\rho(t;\boldsymbol{\alpha})dt$, then $\mu_{i+} = \mu_i(\tau_i) = R_i \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$. Similarly, defining the cumulative baseline intensity function in panel period p as $R_{ip} = \int_{T_{i,p-1}}^{T_{i,p}} Y_i(t)\rho(t;\boldsymbol{\alpha})dt$, we have $\mu_{ip} = \mu_i(T_{i,p-1}, T_{i,p}) = \mathrm{E}(n_{ip}) = R_{ip}\exp(\boldsymbol{x}_i'\boldsymbol{\beta})$.

The likelihood based on panel count data may be written as the product of two terms 1) a conditional distribution of the number of events in each panel n_{ip} given the total number of events n_{i+}, and 2) the distribution of the total number of events n_{i+}. Specifically, the likelihood has the following form:

(2.10)
$$L_p(\boldsymbol{\theta}) = \prod_{i=1}^{M}\left[\binom{n_{i+}}{n_{i1},\ldots,n_{ie_i}}\prod_{p=1}^{e_i}\left(\frac{R_{ip}}{R_i}\right)^{n_{ip}}\right]\times\prod_{i=1}^{M}\int_0^{\infty}\frac{(\nu_i\mu_{i+})^{n_{i+}}e^{-\nu_i\mu_{i+}}}{n_{i+}!}G(\nu_i)d\nu_i.$$

Note that if there is a single panel, $L_p(\boldsymbol{\theta})$ (2.10) will reduce to the simple mixed Poisson kernel.

3. Use of Estimating Equations for Inference in Recurrent Event Models

Inference for Poisson and mixed Poisson models may be based on likelihood methods, and alternatively, on more robust approaches only assuming the form of the mean. These alternative approaches estimate the variance by using either a robust model-based variance or a robust empirical variance. In this section we provide details on how estimation can be carried out using estimating equations; we build the estimating equations for regression parameters from the negative binomial distribution for the counts of events.

Let the intensity function of the counting process of the events, given the subject-specific random effect ν_i, be

(3.1)
$$\lambda_i(t) = \nu_i\rho(t;\boldsymbol{\alpha})\exp\left\{\boldsymbol{x}_i'\boldsymbol{\beta}\right\}.$$

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}', \phi)'$, and let ω_{ipl} be the time of the l-th event, from the start of the study, for the i-th individual in panel period p, $i = 1,\ldots,M$, $p = 1,\ldots,e_i$, $l = 1,\ldots,n_{ip}$. The likelihood kernel based on either the full data (subscripted by $d = f$) or the panel data (subscripted by $d = p$) factorizes as:

(3.2)
$$L_d(\boldsymbol{\theta}) = L_{\alpha,d}(\boldsymbol{\alpha})\times\prod_{i=1}^{M}\int_0^{\infty}\frac{(\nu_i\mu_{i+})^{n_{i+}}e^{-\nu_i\mu_{i+}}}{n_{i+}!}G(\nu_i)d\nu_i, \qquad d\in\{f,p\}$$

where

(3.3)
$$L_{\alpha,f}(\boldsymbol{\alpha}) = \prod_{i=1}^{M}\prod_{p=1}^{e_i}\prod_{l=1}^{n_{ip}}\frac{\rho(\omega_{ipl};\boldsymbol{\alpha})}{R_i},$$

and

(3.4)
$$L_{\alpha,p}(\boldsymbol{\alpha}) = \prod_{i=1}^{M}\left[\binom{n_{i+}}{n_{i1},\ldots,n_{ie_i}}\prod_{p=1}^{e_i}\left(\frac{R_{ip}}{R_i}\right)^{n_{ip}}\right].$$

Lawless [8] provides the first and second derivatives of the negative binomial likelihood with respect the parameters, which provides the basis for likelihood estimation along with (3.2).

Let $\boldsymbol{g}_d = (\boldsymbol{g}'_\beta, \boldsymbol{g}'_{\alpha,d}, g_\phi) = \mathbf{0}'$ denote the full set of estimating equations for the panel ($d = p$) or the full ($d = f$) data. Since the likelihood is a function of $\boldsymbol{\beta}$ only through the second term in (3.2), the estimating equations for this parameter are developed here as the usual quasi-likelihood equations $(\partial\boldsymbol{\mu}/\partial\boldsymbol{\beta})' U_o^{-1}(\mathbf{n}-\boldsymbol{\mu}) = \mathbf{0}$, where $U_o = \text{diag}\{\mu_{i+}(1+\phi\mu_{i+}), i = 1,\ldots,M\}$, $\mathbf{n} = (n_{1+},\ldots,n_{M+})'$ is a vector of counts, and $\boldsymbol{\mu} = (\mu_{1+},\ldots,\mu_{M+})'$ is the vector of their expected values. Defining $U = \text{diag}\{\mu_{i+}, i = 1,\ldots,M\}$, this becomes

$$\boldsymbol{g}_\beta = X'UU_o^{-1}(\mathbf{n}-\boldsymbol{\mu}) = \mathbf{0}. \tag{3.5}$$

We obtain an estimating equation for $\boldsymbol{\alpha}$ by combining $\partial \log L_{\alpha,d}/\partial\boldsymbol{\alpha}$, $d = f, p$, with quasi-likelihood estimation as both first and second terms in (3.2) depend on $\boldsymbol{\alpha}$, yielding

$$\boldsymbol{g}_{\alpha,d} = \frac{\partial \log L_{\alpha,d}(\boldsymbol{\alpha})}{\partial\alpha} + W'UU_o^{-1}(\mathbf{n}-\boldsymbol{\mu}) = \mathbf{0}, \tag{3.6}$$

where W is a matrix with entries

$$w_{ia} = \frac{\partial \log R_i}{\partial\alpha_a}, \qquad i = 1,\ldots,M, \text{ and } a = 1,\ldots,d_\alpha. \tag{3.7}$$

Several choices may be considered for the estimating equation of the overdispersion parameter ϕ. In our examples, we use the pseudo-likelihood estimator, which has been popular since its introduction by [4]. It has performed well in simulation studies for simple overdispersed count analyses and has documented optimality properties [11]. The pseudo-likelihood estimating equation for ϕ is

$$g_\phi = \sum_{i=1}^{M} \frac{(n_{i+}-\mu_{i+})^2 - (1-h_i)\mu_{i+}(1+\phi\mu_{i+})}{(1+\phi\mu_{i+})^2} = 0, \tag{3.8}$$

where $h_i = \text{diag}(U^{1/2}V'(V'UV)^{-1}V'U^{1/2})$, $V = (\ X \quad W\ )$; h_i is the diagonal of the hat matrix and represents a correction to reduce small sample bias in this simple second moment equation.

Let the estimator of $\boldsymbol{\theta}$ from either full ($d = f$) or panel ($d = p$) data be denoted by $\hat{\boldsymbol{\theta}}_d$. Under standard conditions for the application of asymptotic results to estimating equations, $\sqrt{M}(\hat{\boldsymbol{\theta}}_d - \boldsymbol{\theta})$ is asymptotically normal with asymptotic covariance

$$E\left(-\lim_{M\to\infty}\frac{\partial \boldsymbol{g}_d}{\partial\boldsymbol{\theta}}\right)^{-1} E\{\lim_{M\to\infty}\boldsymbol{g}_d\boldsymbol{g}'_d\}\left\{E\left(-\lim_{M\to\infty}\frac{\partial \boldsymbol{g}_d}{\partial\boldsymbol{\theta}}\right)^{-1}\right\}'. \tag{3.9}$$

Finite sample variance estimates are obtained by substituting $\hat{\boldsymbol{\theta}}_d$ for $\boldsymbol{\theta}$ and omitting the expressions $\lim_{M\to\infty}$. In this case there are two options for approximating the expectation of the terms in (3.9). The first is a model-based approach, which in this case requires specification of 3rd and 4th moments of the counts. The second is an empirical approach, which substitutes $E\{\sum_{i=1}^M g_{id}g'_{id}\}$ by $\{\sum_{i=1}^M g_{id}g'_{id}\}$; where g_{id} denotes the contribution to the score equation from individual i.

	Full Data					2-Panel Data				
	Estimates		Standard Errors			Estimates		Standard Errors		
	NB	QL	NB	MB	EMP	NB	QL	NB	MB	EMP
β_1	-3.449	-3.428	0.348	0.275	0.333	-2.976	-2.947	0.503	0.477	0.486
β_2	0.133	0.118	0.316	0.302	0.287	0.118	0.108	0.313	0.301	0.286
β_3	-0.541	-0.540	0.317	0.264	0.291	-0.553	-0.551	0.315	0.263	0.290
β_4	0.245	0.240	0.072	0.057	0.065	0.240	0.236	0.072	0.056	0.065
α	1.019	1.015	0.069	0.055	0.068	0.886	0.880	0.124	0.117	0.121
ϕ	1.144	0.848	0.285	0.193	0.275	1.122	0.846	0.282	0.193	0.275

Table 1. Parameter estimates and their standard errors, resulting from the likelihood (NB) and quasi-likelihood (QL) analyses, using both a robust model-based (MB) and a robust empirical (EMP) approach, fit to the bladder cancer data. The regression parameters $\beta_1, \beta_2, \beta_3$ correspond to the three treatment groups, parametrized with respect to the placebo, and β_4 to the covariate, number of tumors at baseline.

4. Analysis of Bladder Cancer Data

In this section we estimate the treatment effects for the bladder cancer study [1] discussed in Section 1 under a design with continuous followup as well as a panel design, for illustrative purposes, with 2 equally spaced scheduled followup visits over 64 months; for the panel design, we record information on event recurrences at the scheduled followup times 32 and 64 months, and at termination times.

Figure 1 displays the recurrences of events. Note that the rate of occurrence of events seems to be slightly lower in the thiotepa group. Figure 1 shows that the three treatment groups are very similar in terms of the distributions of the termination times, the number of tumors, and the size of the largest tumor at baseline.

Table 1 reports parameter estimates for a model with Weibull baseline, and their standard errors: likelihood based, robust model-based and robust empirical, under a 2-panel design, as well as from an analysis of the full data with continuous followup. Since the estimated coefficient corresponding to the size of the largest tumor at baseline is not significant, this variable has been excluded in the subsequent discussion. The estimates under the negative binomial and quasi-likelihood are quite close, except for the overdispersion parameter, which is slightly larger based on the negative binomial analysis. The standard errors are also similar, with likelihood values larger than the robust model-based estimates. The robust empirical estimates are data-driven (not model-based) and are quite close to those from the negative binomial analysis. The difference is again more pronounced for estimates of the overdispersion parameter. The thiotepa treatment may have a significant protective effect on recurrences, while the pyridoxine treatment effect is non-significant. There is substantial overdispersion in the data, and the estimate of the Weibull shape parameter α is quite close to unity.

Importantly, note that the estimates of the treatment effects (relative to placebo) from the panel design are quite close to those from the full data analysis, indicating cost efficiency would be gained from well-designed panel designs with relatively little loss of information regarding treatment effects. There are larger

differences observed in panel versus full data analyses for estimates of other parameters. For example, the estimate of the standard error of $\hat{\alpha}$ is considerably higher in the 2-panel design than that of the corresponding estimate from the full data analysis.

5. Other Approaches and Topics

Dependent termination time of the recurrent event process is one of the topics that has drawn more attention in the last few years. This topic requires appropriate modeling when the end of the study, for example death, is related to the occurrence of the events. A popular approach to model this dependency is through shared individual frailties, where the recurrent event process and the survival process are linked [10].

This approach of shared individual frailties linking recurrent event and survival outcomes has also been used to model dependency when considering multivariate recurrent event processes [5, 13]. This may be useful when modeling recurrent event processes that may evolve differently over time, but are likely to be intrinsically associated.

In summary, in this article we have presented briefly some counting based models, and have discussed approaches where we have both full data available, through the timing of events, and where only panel counts are available; we also discussed inference based on likelihood and estimating equations in both situations. Juarez-Colunga et al [6] built from this inference framework an investigation of the loss of efficiency of a treatment estimator in panel designs as compared to designs using continuous followup. A main result of this investigation states that little is lost when the distributions of both the termination times and the covariates adjusted for in the analysis are close across treatment groups. A simple example of similarity in such distributions would be if the termination times over the covariates are identical across the treatment groups; for example, if the covariate is sex, and there are two treatments (A and B), this would mean that the same number of females and males are allocated to the treatment groups, and that the termination times for each gender under treatment A are the same as for those under treatment B. Juarez-Colunga et al [6] provides details on conditions for optimal panel designs, broadly as well as several illustrative examples.

Acknowledgments

We gratefully acknowledge support from the National Council of Science and Technology of Mexico (to EJC) and the Natural Sciences and Engineering Research Council of Canada (to CBD) for the conduct of this research.

Received June 23, 2013

Final version received October 24, 2013

E. Juarez-Colunga
Department of Biostatistics and Informatics,
University of Colorado Denver,
Aurora, CO 80045, USA.
Elizabeth.Juarez-Colunga@ucdenver.edu

C. B. DEAN
DEPARTMENT OF STATISTICAL AND ACTUARIAL SCIENCES,
WESTERN UNIVERSITY, WESTERN SCIENCE CENTRE,
LONDON, ON N6A 5B7, CANADA.
dean@stats.uwo.ca

REFERENCES

[1] D. F. ANDREWS AND A. M. HERZBERG, *Data: A collection of problems from many fields for the student and research worker*, Springer-Verlag, New York, 2000.

[2] D. BYAR, C. BLACKARD, AND THE VETERANS ADMINISTRATION CO-OPERATIVE UROLOGICAL RESEARCH GROUP, *Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage I bladder cancer*, Urology, **10**, (1977), 556–561.

[3] R. COOK AND J. LAWLESS, *The statistical analysis of recurrent events*, Springer, New York, 2007.

[4] M. DAVIDIAN AND R. J. CARROLL, *Variance function estimation*, Journal of the American Statistical Association, **82** (400), (1987), 1079–1091.

[5] X. HE, X. TONG, AND J. SUN, *Regression analysis of multivariate panel count data*, Biostatistics (Oxford, England), **9**, (2008), 234–248.

[6] E. JUAREZ-COLUNGA, R. BALSHAW, AND C. B. DEAN, *Efficient panel designs for longitudinal recurrent event studies recording panel counts*, Biostatistics. To appear.

[7] J. D. KALBFLEISCH AND R. L. PRENTICE, *The statistical analysis of failure time data*, John Wiley & Sons, 2nd ed., 2002.

[8] J. F. LAWLESS, *Negative binomial and mixed poisson regression*, The Canadian Journal of Statistics/La Revue Canadienne de Statistique, **15**, (1987), 209–225.

[9] ———, *Regression methods for poisson process data*, Journal of the American Statistical Association, **82**, (1987), 808–815.

[10] L. LIU, R. WOLFE, AND X. HUANG, *Shared frailty models for recurrent events and a terminal events*, Biometrics, **60** (3), (2004), 747–756.

[11] J. A. NELDER AND Y. LEE, *Likelihood, quasi-likelihood and pseudolikelihood: some comparisons*, Journal of the Royal Statistical Society. Series B (Statistical Methodology), **54**, (1992), 273–284.

[12] VANCOUVER COASTAL HEALTH, *A study for treatment of superficial bladder cancer using ogx-427*. Clinical Trials NIH Registry, 2012.

[13] X. X. ZHAO, L. L. LIU, Y. Y. LIU, AND W. W. XU, *Analysis of multivariate recurrent event data with time-dependent covariates and informative censoring.*, Biometrical Journal, **54** (5), (2012), 585–599.

Bol. Soc. Mat. Mexicana (3) Vol. 19, 2013

GROUP DECISION MAKING

M. D. L. DOLORES SÁNCHEZ CASTAÑEDA AND STEPHEN G. WALKER

ABSTRACT. In this paper, we describe a novel approach to group decision making. We develop theory specifically for a group rather than rely on standard maximizing expected utility rules based on treating group aggregations from individuals as a super–individual who adheres to axioms of rational behavior. The maximization of expected utility rule demands strong assumptions from decision makers. This problem is more complex for group decision making due to the conflict of interest within the group. To assume probabilistic preferences, and random utility models, seems a more realistic approach to group decision making. However, standard literature on probabilistic choice does not incorporate the uncertainty about certain events under which the decision making process takes place. Additionally, the aggregation of beliefs about the uncertain events of the world and the aggregation of utilities attached to them, become an important issue in the group decision making process. Our idea is to combine theory from individual rational choice and random utility models to propose a new rational choice for group decision making. We assume the decisions makers preferences are probabilistic due to uncertain states of the world. We present illustrations.

1. Introduction

Decision theory aims to give a decision rule on what action is the best choice assuming decision makers behave rationally. However, the concept of rationality restricts the individual to certain conditions on choice behavior that are normative. This is, normative approaches give a prescription of what action is the one that the decision maker should choose, they do not describe how people actually behave. Decision theory is closely related to utility theory, where the point of any agent facing any type of decision is to achieve an optimal outcome. In social choice, where decisions are made for the society, utility theory is more complex, since it involves both individual and the collective interests. The problem is how to reconcile these two aspects when people may have different views or considerations about what is the best for the society as a whole.

It is well known that decision theory is about taking actions with imperfect knowledge. The main idea is to construct a utility function on actions and outcomes with a probability specified on the uncertain future outcome. Hence, one is able to derive a fully specified expected utility, and [38] were among the first to formalize the axioms of rational behavior, leading to the well known maximization of expected utility rule. The axioms where restated within a subjective probabilistic framework by [36], whose work is also based on [32]. Whether it is considered with an objective or subjective probabilistic perspective, these set of axioms are well known as the coherence or rationality axioms. The normative

2010 *Mathematics Subject Classification*: 91B06, 91B10, 91B82.

Keywords and phrases: group decision making, social choice, expected utility, rationality, democracy, random utility, opinion pooling, weighted likelihood bootstrap.

approach to decision making consists on choosing the action which maximizes the expected utility, under the assumption that every decision maker adheres to the rationality axioms. However, given the uncertainty about external events that affect outcomes from an action taken and due to imperfect perception about the attributes of the set of alternatives, it has been found that such coherence axioms are not always satisfied [2, 37, 17]. As an alternative, other models of choice have been develop to describe individual choice behavior, i.e. how the individual actually behaves, instead of how she should behave. Such are the cases of random utility models, [3, 22]; and prospect theory, [16].

The aim of this work is to propose a coherent group decision rule which does not involve a voting process, and is based on a descriptive choice behavior, rather than a normative one. The utility is typically described as a cardinal number or an increasing monotonic function, but the approach to be used here is to represent the utility as a random function, taking ideas from random utility models; see [3], and [20]. We provide a formal framework and an axiom system for group decision making under probabilistic choice behavior.

To set the notation, we define the action space $\mathscr{A}$ as the set of all possible actions or alternatives $\{a_j \in \mathscr{A}\}$, $j = 1,\ldots,J$ that are available to the decision maker. $\mathscr{A}$ is typically finite, although considering the extension to the infinite case is also valid. The states of the world $\theta \in \Theta$ are the uncertain events. The decision maker has no control over them but they are relevant because they will have an effect on the outcome in conjunction with the set of alternatives. The states of the world form a partition of the algebra Θ, such that $\{\theta \in \Theta\}$ if the state space is a continuum or $\{\theta_l \in \Theta\}$ with $l = 1,\ldots,L$ if it is discrete. The description of the states of the world can include all the possible events, or it can be a subset including only the relevant events for the decision problem in question. The set of consequences $\{c_l \in \mathscr{C}\}$, $l = 1,\ldots,L$, are the outcomes derived from a realization of a particular state of the world and a particular action taken, i.e. $c = c(a,\theta)$. Although individuals show preferences over pure consequences, these consequences are always an associated result of the action taken. Hence, preferences over consequences induce a preference over actions. $\mathscr{C}$ is typically finite, although considering the extension to the infinite case is also valid. A binary or pairwise preference relation denoted by $\succeq$, is used to compare and order the elements in the action space $\mathscr{A}$, and in the set of consequences $\mathscr{C}$. The utility function over the set of alternatives and the uncertain events, $U(a,\theta)$ quantifies the preference relations, and it is derived from a valued-scale function v over the consequences, $v(c(a,\theta)) = U(a,\theta)$; see [14] for more details.

Describing the layout of the paper; in Section 2 we review the random utility model. (RUM). Section 3 introduces the representation of the random utility model which includes the random variable capturing the uncertainty of the external events of the world. Section 4 describes our approach to group decision theory. In Section 5 we present some examples and applications, and we have a brief discussion in Section 6.

2. Random utility models

Probabilistic theories of preference were developed to acknowledge inconsistencies and deviations from the normative expected utility rule. These theories

attempt to describe individual choice behavior in a more realistic way, assuming that for repeated choices under the same circumstances and, with the same available alternatives, individuals choose sometime one alternative and sometime they choose another. Therefore, one can assign a probability distribution on the space of alternatives to estimate and describe the proportion of times that the decision maker selects one alternative over the other. Probabilistic choice admits intransitivities and multiple comparisons of alternatives, [27, 33] a more realistic way to incorporate randomness from individual choice.

The random utility model in the probabilistic choice approach is derived from the problem in proving the existence of numerical scales of utility and subjective probability. However, it should also take into account external uncertainty in some way. [27] and [22] introduced the random utility model, in which the utility function is not known at the moment of choice, but is assumed to be randomly determined on each presentation to the subject. Once selected, the subject's decision or outcome is completely determined by the relevant utilities, just as in the algebraic models. Hence, the utility function is selected according to some postulated probability mechanism that is perceived as to be the largest value of the utility. Since non–random utility models are considered inadequate in descriptive theories of choice, these models were intended to relax the assumptions of normative theory in order to have weaker models that describe and forecast behavior in a better way.

As [3] remarks, a way to weaken the non-stochastic theory is to make the ordinal utility function as a random one. Hence, RUMs assume that choice probabilities reflect the underlying utility function of individual preferences and that individuals choose alternatives according to what they perceive as a maximization of that utility. Therefore, RUMs fill the gap between normative and descriptive behavior, while retaining the principle of optimality.

Probabilistic preferences are assumed to satisfy the set of axioms stated by [23] as the conditions for a simultaneous conjoint measurement for binary relations. These axioms where extended to the multiple comparison case by [19]. According to notation in Section 1, choice probabilities for binary preferences over alternatives in $\mathscr{A}$, are defined as follows.

Definition (2.1) (Choice probability). When the subject is forced to choose between two elements $\{a_i, a_j \in \mathscr{A}\}$, the probability of choosing a_i over a_j is $P(a_i, a_j)$, and $P(a_j, a_i)$ is the probability of choosing the reverse order, i.e. choosing a_j over a_i, where P is a real-valued function that satisfies

1. $P : \mathscr{A} \times \mathscr{A} \to [0,1]$
2. $P(a_i, a_j) + P(a_j, a_i) = 1$
3. $P(a_i, a_j) \in (0,1)\ \forall j \neq i$, follows from 1.

An individual will prefer alternative a_i over a_j if $P(a_i, a_j) > P(a_j, a_i)$. If $P(a_i, a_j) = 1$ then we say that there exists perfect discrimination between a_i and a_j. If there are only two elements in $\mathscr{A}$ and if $P(a_i, a_j) = 1/2 = P(a_j, a_i)$ we say that the individual is indifferent to both alternatives. Also, $P(a, a) = 1/2$, is a consequence of *2* in Definition (2.1). However, the contrary is not true. When two alternatives have probability 1/2 of being selected, it does not mean that both alternatives are necessarily the same. In economic terms, it can mean that both alternatives are the substitutes of each other.

In a similar way, we can define choice probabilities over subsets of alternatives. To say, for $S \subset T \subseteq \mathscr{A}$, $P_T(S)$ is the probability of choosing the set of alternatives S when the set T of alternatives is presented to the decision maker. In general, we assume that these probabilities also satisfy the following axiom stated by [20]:

AXIOM (2.2) (Probability axiom). 1. *For* $S \subset T$, $0 \leq P_T(S) \leq 1$.
2. $P_T(T) = 1$
3. *For all* $S_i \subset T$ *such that* $\bigcap_{i \in I} S_i = \emptyset$, *then*

$$P_T\left(\bigcup_{i \in I} S_i\right) = \sum_{i \in I} P_T(S_i).$$

A generalization of part 3. for elements of $S \subset T$ *implies*

$$P_T(S) = \sum_{a \in S} P_T(a)$$

Although, initially it was not specified, choice probabilities are subjective in the sense of [32] and [36]; since they represent individual views about personal preferences.

Another fundamental axiom in probabilistic choice by [20], is the choice axiom.

AXIOM (2.3) (The choice axiom). *Let* T *be a finite subset of* $\mathscr{A}$ *such that, for every* $R \subset S \subset T$, *where* $P_S(R)$ *is the probability of choosing the subset of alternatives* R *from the subset* S,

1. *If* $P(a_i, a_j) \neq 0$ *or* 1 *for all* $a_i, a_j \in T$, *then for* $R \subset S \subset T$
$$P_T(R) = P_S(R) P_T(S);$$
2. *If* $P(a_i, a_j) = 0$ *for some* $a_i, a_j \in T$, *then for every* $S \subset T$
$$P_T(S) = P_{T-\{a_i\}}(S - \{a_i\}).$$

In terms of conditional probability, this axiom is equivalent to say that if $P_T(S) > 0$, then the conditional probability of choosing the subset of alternatives R given that the individual has already chosen the set S from T is

$$P_T(R|S) = \frac{P_T(R \cap S)}{P_T(S)}.$$

The importance of this axiom is that with this assumption it is easy to deal with the problem involving both perfect and imperfect discrimination and it helps to define the properties of general choice probability. In words, the axiom says that if we include more alternatives in the relevant set, the choice probabilities are not affected. Hence, the extension of this axiom also applies when instead of considering just one irrelevant alternative a_i, we consider a set of irrelevant alternatives, say $E = \{a_i | a_i \in S\} \subset S$ such that $P(a_i, a_j) = 0$ for all $a_i \in E$.

A weak utility model allows to state any order-preserving function to be restated in terms of probabilities, given that the preference order is defined by a real-valued function w, this is equivalent to one alternative having probability 1/2 of being selected over the other one.

Definition (2.4) (Weak utility). A weak (binary) utility model is a set of preference probabilities for which there exists a real-valued function w over $\mathscr{A}$ such that

$$P(a_i, a_j) \geq \frac{1}{2} \text{ if and only if } w(a_i) \geq w(a_j), \ \forall\, a_i a_j \in \mathscr{A}. \tag{2.5}$$

Another important concept in probabilistic choice that reflects to certain extent the strength of preferences is the stochastic transitivity; which is a probabilistic extension of the algebraic transitivity property for probabilistic preferences, and it is defined for the binary relations as follows:

Definition (2.6) (Stochastic transitivity). Given three alternatives, a_i, a_j, $a_k \in \mathscr{A}$, and $\min[P(a_i,a_j), P(a_j,a_k)] \geq \frac{1}{2}$ the binary preferences satisfy:

1. Weak stochastic transitivity if and only if

$$P(a_i,a_k) \geq \frac{1}{2}$$

2. Moderate stochastic transitivity if and only if

$$P(a_i,a_k) \geq \min[P(a_i,a_j),P(a_j,a_k)]$$

3. Strong stochastic transitivity if and only if

$$P(a_i,a_k) \geq \max[P(a_i,a_j),P(a_j,a_k)]$$

The intuitive interpretation of this property is that an individual prefers a_i over a_j if $P(a_i,a_j) > 1/2$. This is equivalent to saying that when facing repeated decision situations, at least 50% of the times she chooses a_i over a_j. An equivalent condition to moderate stochastic transitivity is

$$P(a_i,a_j) \geq \frac{1}{2} \text{ if and only if } P(a_i,a_k) \geq P(a_j,a_k).$$

[4] link choice probabilities with utility functions. The implication is that weak stochastic transitivity is a necessary condition for the existence of a utility function. Hence, under some conditions the comparison of probabilities can be interpreted as a difference in cardinal utilities.

Definition (2.7). For a given subject, a real valued function w on $\mathscr{A}$ is called the *utility function* if and only if, for any a_1, a_2, a_3, $a_4 \in \mathscr{A}$,

$$P(a_1,a_2) \geq P(a_3,a_4) \text{ if and only if } w(a_1) - w(a_2) \geq w(a_3) - w(a_4).$$

provided neither $P(a_1,a_2)$ nor $P(a_3,a_4)$ are equal to 0 or 1.

Then one will feel compelled to look for a cardinal utility function w such that $P(a_1,a_2) > P(a_3,a_4)$ to be equivalent to $w(a_1)-w(a_2) > w(a_3)-w(a_4)$. At this point, we would be willing to make a link with normative theory through the cardinal utility. For instance, strong stochastic transitivity is a necessary and sufficient condition for the existence of a utility function on a set of alternatives.

Now we present the definition of a random utility model. Let $A \subseteq \mathscr{A}$ be the finite subset of alternatives, and $\mathbf{U}$ be a *random vector of utilities* on A, such that $\mathbf{U} = [U_1(a_1),\ldots,U_n(a_n)]$, with each $U_i(a_i)$ also denoted as $U(a_i)$, is called the random utility of $a_i \in A \subseteq \mathscr{A}$. No assumption is made about dependency relations between these random variables.

Definition (2.8) (Random utility model). A random utility model is a set of preference probabilities defined for all subsets of a finite $\mathscr{A}$ for which there is a random vector $\mathbf{U} = [U_1(a_1), \ldots, U_n(a_n)]$ on $\mathscr{A}$, such that for $a_i \in A \subseteq \mathscr{A}$,

$$P_A(a_i) = \mathbb{P}\left[U_i(a_i) \geq U_j(a_j)\, \forall j \in A\right] = \int_{-\infty}^{\infty} \mathbb{P}\Big[U_i(a_i) \in \mathrm{d}t, U_j(a_j) \leq t,\ \forall j \in A\Big]. \tag{2.9}$$

Here $\mathbb{P}$ is a probability measure over the Borel σ–algebra σ, the probability space is $(\mathscr{A}, \sigma, \mathbb{P})$ and $\{a_i \in \mathscr{A} : U(a_i) \geq U(a_j)$ for all $a_j \in \mathscr{A}, j \neq i\}$, defining a utility random variable U on $\mathscr{A}$. If the definition is only asserted for the binary preference probabilities, i.e.

$$P(a_i, a_j) = \mathbb{P}\Big[U_i(a_i) \geq U_j(a_j)\Big] \tag{2.10}$$

then the model is called a binary random utility model. If the random vector $\mathbf{U}$ consists of components that are independent random variables, then the model is said to be an independent random utility model.

In other words, there is a random vector $\mathbf{U} = [U_1(a_1), \ldots, U_n(a_n)]$, such that each $U_i(a_i)$ also denoted as $U(a_i)$, is called the random utility of a_i and it is unique up to an increasing monotone transformation such that,

$$P(a_i, a_j) = \mathbb{P}\Big[U(a_i) \geq U(a_j)\Big] \tag{2.11}$$

is the probability of choosing a_i over a_j. Definition (2.8) represents the decision maker's probability that a_i will maximize her random utility function. That is, if forced to choose, she will choose a_i over all the alternatives with probability $P(a_i)$. In general, the most common random utility models used to model and estimate choice probabilities are the Fechnerian and the strict random utility models [22]. The work of [5] is also relevant for random utility theory since he provides the necessary and sufficient conditions on choice probabilities for the existence of RUM models. The work of [28] is remarkable given that the random utility models he proposed are still used in econometrics and consumer choice, such as the conditional logit model. Also, [33] and [34] extend the representation and characterization problems of random utility models to the case of doing multiple comparisons or m-ary relations.

3. RUM with states of the world

In the preceding section, the ideas do not consider the states of the world approach to construct a random utility model. We now use the idea of random utility to model random choice within an individual, where actions are taken in the presence of an unknown state of the world θ. So let Θ denote the possible states of the world, and $\Pi(\theta)$ denote its probability distribution, that is, the decision makers belief distribution as to the possible outcome of θ: the true state of the world [32, 36]. Here, we are proposing a random utility model on a finite set $\mathscr{A}$ for a random vector of utilities $\mathbf{U} = (U_1, \ldots, U_J)$ that was defined in Section 2 as $U_j = U(a_j)$, but now it is defined as $U_j = U(a_j, \theta)$, $a_j \in \mathscr{A}$, $j = 1, \ldots, J$ and $\theta \sim \Pi(\theta)$. Hence, the random vector of utilities is now given by

$$\mathbf{U} = [U(a_1, \theta), \ldots, U(a_J, \theta)].$$

The probability that the decision maker selects a_i over all options in $\mathscr{A}$ is

$$P(a_i) = \mathbb{P}\left[U(a_i,\theta) \geq U(a_j,\theta)\ \forall a_j \in \mathscr{A}\right]. \tag{3.1}$$

Now, consider a partition $\{B_i\}$ of Θ such that it is known that if $\theta \in B_i$ then a_i maximizes $\{U(a_j,\theta);\ \forall j \neq i\}$. Hence, by (3.1), $P(a_i|\theta) = \Pi(B_i)$, and choosing the action a_i is equivalent to choosing the interval B_i. Moreover, if we define

$$B_{ij} = \{\theta : U(a_i,\theta) > U(a_j,\theta)\}, \tag{3.2}$$

so that $B_{ij} \cap B_{ji} = \emptyset$ and $B_{ij} \cup B_{ji} = \Theta$ then $P(a_i,a_j|\theta) = \Pi(B_{ij})$. In a similar way as above, it is known that if $\theta \in B_{ij}$, then a_i maximizes $\{U(a_j,\theta);\ \forall j \neq i\}$. The decision maker believes that action a_i is the best action with probability $\Pi(B_i)$, since she believes that θ lies in B_i with probability $\Pi(B_i)$. What the decision maker does with this piece of information and what we propose it can be used for with respect to decision making, in particular group decision making, is outlined in Section 4.3.

We briefly discuss the extension to an infinite $\mathscr{A}$. We define $a(\theta)$ to be the unique maximizer of $U(a,\theta)$ for $\theta \in \Theta$. That is $U(a(\theta),\theta) \geq U(a,\theta)$ for all $\{a \in \mathscr{A}\}$. Then define $B_A = \{\theta : a(\theta) \in A \subset \mathscr{A}\}$ so that $\theta \in B_A$ if and only if $a(\theta) \in A$. Therefore, Π_Θ induces a probability distribution on $\mathscr{A}$ via $P_{\mathscr{A}}(A) = \Pi_\Theta(B_A)$. In particular, we can easily sample from $P_{\mathscr{A}}$ by drawing a θ from Π_Θ and evaluating $a(\theta)$. Hence, instead of maximizing the expected utility, we maximize the induced distribution of a_i. Utility functions considered in this approach are all possible functions that can be expressed in terms of the alternatives a_i, and the states of the world θ, and those include typical utility or loss functions used for statistical inference. Examples of these are the quadratic utility function, the log–score utility function or the exponential utility function. This idea leads to a random function model in the next section.

(3.1) Random function model. With regard to probabilistic choice approaches, there is a strong relationship between the random utility approach, the random relations approach and the random function approach. In fact, [34] extend results in [33], [12] and [13], in which they investigate the characterization problem and prove that there is a general representation of random utilities in terms of a random preference relational structure or a random valued function. This is also true for the representation of a random preference relational structure in terms of a random utility representation or a random valued function; and finally of a random function in terms of a random relation or a random utility representation.

Based on the latter, here we describe an alternative probabilistic choice model based on the idea that the action a can be defined via a function $h(\cdot) \in \mathscr{H}$, $\mathscr{H} : \Theta \to \mathscr{A}$. Where h is a one-to-one function on the unknown state of the world θ, which, as previously, is assigned a subjective probability $\Pi(\theta)$. The action a depends on θ and h via the relation $a = h(\theta)$ which is known to maximise $\mathbf{U}(a,\theta)$. If $\mathscr{A}$ is finite then we may as well take Θ to be finite with elements the same number as in $\mathscr{A}$, and $a_i = h(\theta_i)$ for all $i = 1,\ldots,n$. Also, as discussed in Section 3, $P_{\mathscr{A}}(a_i) = \Pi(\theta_i)$. When $\mathscr{A}$ is infinite then $P(a \in A) = \Pi(\theta \in B_A)$ where $B_A = \{\theta : h(\theta) \in A \subset \mathscr{A}\}$. Hence, action a depends on the occurrence of a single θ.

One could attempt to put this in a utility framework via e.g.

$$U^*(a,\theta) = \mathbf{1}(a,h(\theta)),$$

where **1** is the indicator function for $a = h(\theta)$, but any choice of utility would appear to be arbitrary here. For instance, the choice of $U^*(a,\theta) = -(a - h(\theta))^2$ seems inappropriate and unnecessary to add such an arbitrary utility to the problem when one simply knows that $a = h(\theta)$ and it is only θ which is unknown. Hence, to known the explicit function on which a depends on θ, only needs maximization on the said random function, and attaching a utility to $a(\theta)$ and $h(\theta)$ is not necessary and appears to be redundant. Since any RUM can be expressed as a random function model; if θ is known, then the correct action is $a(\theta) = h(\theta)$, which is the unique maximizer of $U(a,\theta)$ for $a \in \mathscr{A}$. In addition, assuming h is strictly monotone,

$$\begin{aligned} \Pr(a \le y) &= \Pr\left(\theta \le h^{-1}(y)\right) \\ &= \Pi(h^{-1}(y)). \end{aligned} \tag{3.3}$$

Let $g = h^{-1}$, then the density function for a is

$$p(a) = g'(a)\Pi'(g(a)). \tag{3.4}$$

Hence, the action maximizing the random function model can be obtained via derivatives or sampling from $p(a)$. Attempts to incorporate descriptive with normative theories of choice have failed, but here, using the random function representation we have reached a stage where we have effectively used both normative and descriptive theory. However, we now require a means by which the decision maker confronted with an uncertain choice can formally make a decision.

4. Group decision making

This section represents our main contribution in the form of a novel approach to group decision making. A group decision problem is a joint decision for the benefit of the group or the society as a whole on whose behalf it is acting. Clearly conflict of interest is a milestone in this process. Different approaches to group decision theory aim to find a compromise solution to resolve the conflict of interests. [1] recognizes that there are two ways of arriving at a social decision; one of them is through a market mechanism, as in economic problems, and the other is through a voting process, as in politics and decisions made by committees. An important aspect to group decision is the notion of democracy understood as the ruling of the majority in which all members of a group or society have equal power and are equally represented, the term carries the sense of equality and the idea of solving conflicts, making decisions in a fair way, as well as the idea of justice, and the balance between the individual and the society's welfare. Social choice is based on individual preferences and its respective rationality assumptions. The main problem has been how to aggregate and characterize individual preferences and uncertainties into a group preference framework. Arrow's work was seminal in trying to find a method of aggregating preferences that satisfies desirable democratic, as well as the coherence conditions. He found that even if individuals behave or want to behave in a rational way, it does not imply that the group will be rational giving place to important contradictions. The first axiom that Arrow consider is that the preference of the society or group for the alternatives are a weak order, implying that these alternatives are reflexive, connected and transitive. The rest of the assumptions or axioms refer to the democratic requirements,

to mention the universal domain of the set of alternatives, the positive association of values, the independence of irrelevant alternatives, the sovereignty of the citizen's, and finally dictatorship (see [1]). Arrow proves that given the previous set of axioms a group of rational individuals in the sense described by [38] will not always satisfy the whole set of them. Arrow's Possibility theorem states that assuming a rational decision maker facing a group decision problem she might not be able to satisfy all axioms, and that at least one will be violated, mainly resulting in either a non-weak order of the alternatives, intransitivity of them or dictatorship and imposed welfare function. Democratic elections takes voting as a general method of preference aggregation, where the outcome is a winning alternative or a ranking of some or all of the alternatives. The outcome or final ranking may vary depending on the voting method or counting rule that is used. Arrow's results can be illustrated in voting methods in which results are typically to choose an alternative from an aggregated ordering that even though it comes from transitive individual orderings, it will not be transitive in the aggregated ordering. This means that the winner alternative, having the highest number of votes from pairwise comparisons, may not be transitive in a consistent way with the individual orderings and cycles can appear. This property is known as the intransitivity of the majority rule. Violation to non-dictatorship or sovereignty properties are also common.

Traditional procedures in social choice theory deal with completely deterministic preferences and are based on the entire or partial preference orderings of the alternatives by each individual [7]. On the other hand, [15] proposed a probabilistic model for social choice, in which the social probabilities are derived from the individual probabilities that will induce a social ordering of alternatives from which a winning alternative can be inferred. Statistical decision making has given main attention to both combining expert opinions, [29] and Nash equilibrium [21], which involves competing utilities, such as the multi-Bayesian approach, [40]. The latter develops the extension of Bayesian decision theory to a group. [40] consider a geometric pool of prior distributions for its Bayesian properties and also because the arithmetic pool leads to dictatorship and it is multimodal, as shown in [11]. However, their approach assumes that collective rationality of preferences is implied by individual rationality axioms of Von Neumann and Morgenstern, ignoring that maximum expected utility is not prescribed for the group context. Moreover, only a few theorists such as [30], [15] and [8] have considered the case when choices are probabilistic.

Despite the attempts to combine rational and probabilistic social choice, these theories always converge to Arrow's (1963) possibility theorem of a contradiction or inconsistency as [9] discuss, since they start from the assumptions that individual rationality implies group rationality. Before describing our approach to group decision making we address the aggregation issue for the individual utilities and probabilities.

(4.1) Aggregation of probabilities. With respect to the inclusion of uncertainty in the decision process, the aggregation of individual beliefs into a group's distribution of beliefs plays a significant role. However, it is also required to satisfy some desirable properties analogous to those stated by [1] (see [9]).

While there is a wide discussion on which aggregation methods is the most appropriate, we conclude that it depends on the specific problem. However, we propose that the logarithmic pool defined below, which consists on the geometric average of the prior distributions, and we consider best due to its properties.

Definition (4.1) (Logarithmic opinion pool). The logarithmic pool $T : [0,1]^n \to [0,1]$ is such that

$$T[\Pi_1,\ldots,\Pi_n](\theta) = \frac{\ell(\theta)\prod_{i=1}^n \Pi_i(\theta)^{w_i}}{\int \ell(\theta)\prod_{i=1}^n \Pi_i(\theta)^{w_i}\mathrm{d}\mu}, \qquad \mu-a.e., \tag{4.2}$$

where ℓ is a bounded function and in general it will be a likelihood function $\ell : \Theta \to [0,\infty)$ derived from some evidence on which the group agrees, the w_i's are constant non–negative weights such that $\sum_{i=1}^n w_i = 1$, μ is a dominating measure on Θ, and $0 < \int \ell(\theta)\Pi(\theta)\mathrm{d}\mu < \infty$.

The logarithmic pool consists of a weighted geometric mean of the individual beliefs; and it is a log–linear combination of the individual beliefs, hence its name. [10] proved that the logarithmic pool is the only pooling operator which satisfies external Bayesianity condition [24].

PROPERTY (4.3) (External Bayesianity). *Suppose that the group agrees on the likelihood function $\ell : \Theta \to [0,\infty)$ derived from some evidence. It is said that a pool operator T is Externally Bayesian if it satisfies*

$$T\left[\frac{\ell(\theta)\Pi_1}{\int \ell(\theta)\Pi_1\mathrm{d}\mu},\ldots,\frac{\ell(\theta)\Pi_n}{\int \ell(\theta)\Pi_n\mathrm{d}\mu}\right] = \frac{\ell(\theta)T[\Pi_1,\ldots,\Pi_n]}{\int \ell(\theta)T[\Pi_1,\ldots,\Pi_n]\mathrm{d}\mu} \qquad \mu-a.e. \tag{4.4}$$

This condition requires that the updating process commutes with the aggregation process in what it is known as a prior to posterior coherent way. This property ensures that in case of updating beliefs by means of new information, the process of combining the prior distributions will produce the same group distribution whether the updating process happens before or after the pooling. If this property is not satisfied, what the group learns from the data would depend on whether they learn it before or after forming their consensus. If the pooling operator is not externally Bayesian, the posterior distribution of beliefs is affected by the order in which the pooling and updating process are done. External Bayesianity relies on the fact that members of the group have agreed on the likelihood for the observed new evidence, which is equivalent to say that all are observing the same data.

Moreover, it is considered by [39] and others the best way to combine beliefs, given that it is typically uni-modal, scale invariant and less dispersed than for example, the linear pool. According to [41], this pool characterizes consensual values in a better way than the linear pool, and is Pareto optimal in the sense that there is no other alternative apart from the consensual value that makes one individual better off without making worse at least to one of the rest members of the group. Also, if the operator pool is multi-modal, then there is no characterization of consensual values. If the opinion pool is scale invariant, the group decision will depend on the normalization chosen, but decisions should not be affected should any order-preserving affine transformation is used. If the weights are the same for all members, the anonymity property is also satisfied [9].

The group distribution of beliefs about θ, $\Pi_G(\theta)$, using the logarithmic pool is:

$$\Pi_G(\theta) = \frac{\ell(\theta)\prod_{i=1}^n \Pi_i(\theta)^{w_i}}{\int \ell(\theta)\prod_{i=1}^n \Pi_i(\theta)^{w_i}\,\mathrm{d}\mu}, \qquad \mu - a.e., \tag{4.5}$$

where $\sum_{i=1}^n w_i = 1$.

(4.2) Aggregation of utilities. The additive aggregation may lack of certain desirable properties, but it is a practical way to synthesize the group preferences, [18]. We present a new idea to aggregate utility functions. Assume that individual utilities will differ not as functions but via parameters of functions, ϕ; i.e. now the utility function of individual i, is denoted as $U_i(a,\theta) = U(a,\theta;\phi_i)$. Then, additive utility aggregation seems wrong as the fundamental shape of the utility is disturbed. This means that the linear pool may produce a group utility that does not look anything like the individual utilities and these ones only differ in parameters, then the linear pool is misrepresenting the individual utilities. Hence, individual utilities have the same interpretation for all members but the difference is only about the values of the parameters of these functions. Therefore, more appropriate perhaps would be to aggregate on the parameters

$$U_G(a,\theta) = U\left(a,\theta;\sum_{i=1}^n \lambda_i\phi_i\right). \tag{4.6}$$

where $U_G(a,\theta)$ is the group utility, $\lambda_i > 0, \forall i\ \sum_{i=1}^n \lambda_i = 1$, are weight scaling constants and ϕ_i's are the parameters of the utility function. In this way we avoid misrepresentation of individual utilities. It is important to keep in mind that θ remains being the variable measuring uncertainty.

(4.3) Group decisions. In this section we present the main axioms and decision rule for group decision making under a random utility model with states of the world. The first assumption here is that the group has agreed in the aggregation of individual utilities and probability functions. Let $U_G(a,\theta)$ be the agreed group utility function. Since the uncertainty of θ induces uncertainty in the decision, $U(a,\theta)$ is also a random variable. The probability that the group chooses action a_k over all options is

$$P_G(a_k) = \mathbb{P}_G\left[U_G(a_k,\theta) \geq U_G(a_j,\theta)\ \forall j \in \mathscr{A}\right]. \tag{4.7}$$

with $\theta \sim \Pi_G(\mathrm{d}\theta)$. Given $\Pi_G(\mathrm{d}\theta)$, consider now the partition of Borel sets on Θ

$$\begin{aligned} \bigcup_{k=1}^{\infty} B_k &= \Theta, \text{ where} \\ B_k &= \left\{\theta : U(a_k,\theta) > U(a_j,\theta)\ \forall a_j,\ k \neq j\right\}. \end{aligned} \tag{4.8}$$

Then the partition $\{B_k\}$ exists on Θ since each θ is in a B_k for some k, and because the $\{U(a_k,\theta)\}$ are ordered for each value of θ, since additionally the non-coincidence condition holds: i.e.

$$\mathbb{P}_G\left[U_G(a_k,\theta) = U_G(a_j,\theta)\right] = 0. \tag{4.9}$$

Consequently, the group believes that θ lies in B_k with probability $\Pi_G(B_k)$. Hence, the probability that the group chooses action a_k is equal to the probability of

$\{\theta \in B_k\}$,

$$\Pi_G(B_k) = \mathbb{P}_G\left[U_G(a_k,\theta) \geq U_G(a_j,\theta)\ \forall j \in \mathscr{A}\right] = P_G(a_k). \tag{4.10}$$

Similarly, the group believes action a_k is the best with probability $\Pi_G(B_k)$. If the group is facing a voting process, one would expect that the votes would represent such beliefs and a proportion close to $\Pi_G(B_k)$ would vote for action a_k. But this can not be taken for granted since irregularities in individual voting strategies can occur. Hence, we look for a way to incorporate the beliefs without demanding a vote. We discuss the case when the choices of action are countable. To arrive at a decision rule we make an assumption about the group size:

AXIOM (4.11) (Group size). *Given U_G and Π_G, the size of the group is irrelevant to the decision making.*

In essence, the fact that the group have reached agreement about U_G and Π_G means that the work of the individual within the decision making process has now come to an end and some other mechanism not involving individuals should take over, to say the altruistic supra–decision maker. This is why now, the size of the group should not affect the decision made. Given this, we also have the following axiom:

AXIOM (4.12) (Equivalent groups). *Let G_1 and G_2 be two groups facing the same action space $\mathscr{A}$ and same states of the world Θ. If both groups have the same probability distribution of beliefs, $\Pi_{G_1} = \Pi_{G_2}$ and the same group utility function $U_{G_1} = U_{G_2}$, then both groups should reach the same decision.*

This assumption seems to be obvious, if two different groups have the same utility function and same distribution of beliefs, then the decision should be the same. This is also an implicit assumption of the expected utility rule, and it is important to give coherence to the decision making, except that for this rule there are additional assumptions to be made.

Let us consider the group G_2 of infinite size, or arbitrarily large size, for which the proportion of people who would vote for each alternative is known, i.e. the proportion of individuals who vote for action a_k is given by $\Pi_{G_2}(B_k)$, for $k = 1,2,\ldots$ We need to consider the infinite group so that we can obtain an arbitrary value for $\Pi_{G_2}(B_j)$ for each j. Actually, when agreement was sought for a group probability, each individual declared precisely their opinion about on which set from $\{B_k\}_{k=1}^{\infty}$ that θ lies. If the proportion of people providing B_k as the answer is $\Pi_{G_2}(B_k)$,then Π_{G_2} is the group probability for G_2. If this group G_2 now vote under the simple majority rule, they will do so according to these proportions, so that the proportion of people who would vote for action a_k is $\Pi_{G_2}(B_k)$ from equation (4.10). Moreover, if $\Pi_{G_2}(a_k) > \Pi_{G_2}(a_j)$ for all $j \in \mathscr{A}$ then a_k, having the greatest proportion of votes, is the action that wins. If we now set $\Pi_{G_1}(B_k) = \Pi_{G_2}(B_k)$ for all sets $\{B_k\}_{k=1}^{\infty}$ and let $U_{G_1} = U_{G_2}$ then, according to Axioms (4.11) and (4.12), and assuming a simple majority voting rule, the action a for G_1 to take is the mode of $P_{G_1}(a)$, since the mode is the action with the highest probability, or equivalently, the value with the greatest proportion of votes.

(4.4) Comments on Aggregation. The theory equally applies to an individual, a group of size 1. We have already declared that the size of a group should be irrelevant when a group utility and group probability have been established. However, this raises an issue of internal consistency and a way of pooling the probability beliefs. If individual i has probability $\Pi_i(B_j)$ of choosing action a_j, then according to our statement, individual i votes for action j if $\Pi_i(B_j) > \Pi_i(B_k)\,\forall k \neq j$. Therefore, overall, in a group of size n, action a_j attracts n_j number of votes, to say

$$n_j = \sum_{i=1}^{n} \mathbf{1}\Big(\Pi_i(B_j) > \Pi_i(B_k)\,\forall k \neq j\Big). \tag{4.13}$$

Where $\mathbf{1}$ is the indicator function. If $\Pi_G(B_j)$ represents the group allocation of probability to the set B_j, then to be consistent we would require that

$$n_j > n_k \Longleftrightarrow \Pi_G(B_j) > \Pi_G(B_k). \tag{4.14}$$

Therefore, we would define

$$\Pi_G(B_j) \propto \sum_{i=1}^{n} \mathbf{1}\Big(\Pi_i(B_j) > \Pi_i(B_k) \quad \forall k \neq j\Big). \tag{4.15}$$

This appears to be a new pooling strategy. An interesting outcome is that if $\Pi_i(B_j)$ is not maximal for any individual i, then the group probability assigned to B_j is precisely 0. This makes sense in that if an action is no one's first preference, then that action must not be adopted as the status quo would be preferred to take any other action. Also, only a partial specification of Π is determined but all the important probabilities $P(a_j)$ are defined.

In the case of the random function approach from Section 3.1 for a group, the idea that $a = h(\theta)$ is common knowledge, that all members of the group have previously accepted and agreed, then h would not be derived via aggregation from various individual h_i, since this makes no sense. Therefore, it is only the subjective probabilities of θ for each individual that need aggregation. However, dealing with another issue of internal consistency for aggregation of utilities needs to be mentioned. If all individuals believe θ is correct, then individual i chooses $h_i(\theta)$, which maximizes $U_i(a,\theta)$. Then for internal consistency, in order to guarantee that the group $h(\cdot)$ is the one providing the greatest number of votes, if $\mathscr{A} = \{1,2,\ldots\}$, then

$$h(\theta) = \arg\max_j \left\{ \sum_{i=1}^{n} \mathbf{1}\left(h_i(\theta) = j\right) \right\}. \tag{4.16}$$

This also appears a new strategy for aggregation of utilities.

5. Applications

This section presents some examples and applications of the decision rule presented in Section 4.3. The first example refers to a group decision theoretic approach to a parameter estimation problem, and the second one refers to a portfolio allocation problem.

(5.1) Statistical parameter estimation. [31] introduced the weighted likelihood bootstrap (WLB) to approximate a posterior distribution of a parameter without the need for a prior distribution. The WLB is an easy method to implement when the maximum likelihood estimator exists and can be computed. The WLB is described as an extension of the Bayesian bootstrap [35] and it consists in maximizing a weighted likelihood function. For example, let $x_1,\dots x_n$ be an i.i.d. sample with probability density function $f(x;\theta)$. Consequently, the likelihood function is given by

$$L(\theta)=\prod_{i=1}^{n} f(x_i;\theta). \tag{5.1}$$

Instead of maximizing $L(\theta)$ to obtain the maximum likelihood estimator, the WLB maximizes the random function $\widetilde{L}(\theta)$, given by

$$\widetilde{L}(\theta)=\prod_{i=1}^{n} f(x_i;\theta)^{w_{n,i}}, \tag{5.2}$$

where the vector $w_n=(w_{n,1},\dots,w_{n,n})$ is a random vector of weights. The natural distribution for the weight vector is a Dirichlet distribution with all the parameters set to 1. The simulation of an approximate posterior distribution works on repeated samples of weight vectors and the maximization of $\widetilde{L}$ as a function of θ. Therefore, the sample is not from any posterior distribution, but an approximation to a possible posterior. For this reason, [31] recommend adjusting the WLB by a kernel density estimate and then with SIR (sample importance re–sampling), and in this way the WLB sample will supposedly be a better approximation of a true posterior. It should be mentioned at this point that the paper of [31] was not universally accepted; the prime reason was that it lacked any formal justification. Let us now consider our approach to statistical parameter estimation. An often used utility function when using $f(x;\theta)$ as an approximate or substitute for the true distribution $P(x)$ is the logarithmic score, defined as

$$U(\theta,P)=\int \log f(x;\theta)\,\mathrm{d}P(x).$$

The negative of this is a commonly used loss function. Suppose, now that we use a Dirichlet process prior for P [6].

If we now have conditionally i.i.d. observations $x_1,\dots,x_n|P\sim P$ and P has a Dirichlet process prior, written as $P\sim Dir(c,P_0)$, then the conditional distribution of P given $\underline{\mathbf{x}}$ is also a Dirichlet process;

$$(P|x_1,\dots,x_n)\sim Dir\left(c+n,\frac{cP_0+nF_n}{c+n}\right), \tag{5.3}$$

where F_n is the empirical distribution, $F_n=n^{-1}\sum_{i=1}^{n}\delta_{x_i}$, and δ_x is the Dirac measure. In particular, the posterior mean is

$$P_n=\frac{cP_0+nF_n}{c+n}.$$

Therefore, returning to the decision problem for θ, the expected utility with respect to the Dirichlet posterior is

$$U_n(\theta)=\int U(\theta,P)Dir(\mathrm{d}P|\underline{x}), \tag{5.4}$$

and so with the log-score utility function,

$$U_n(\theta) = w_n \int \log f(x;\theta)\,\mathrm{d}P_0(x) + (1-w_n)\int \log f(x,\theta)\,\mathrm{d}F_n(x)$$

where $w_n = c/(c+n)$. An interesting observation here is that if we now choose to put $c = 0$, so the posterior Dirichlet process becomes

$$(P|x_1,\ldots,x_n) \sim Dir(n, F_n),$$

commonly known as the Bayesian bootstrap [35]. While the prior may not exist with $c = 0$, the posterior does. In particular then, $E(P|x_1,\ldots,x_n) = F_n$. So now

$$U_n(\theta) = \int \log f(x,\theta)\,\mathrm{d}F_n(x) = n^{-1}\sum_{i=1}^{n} \log f(x_i,\theta)$$

and the maximizer of the posterior expected utility is precisely the maximum likelihood estimator.

Now, P is distributed according to the Bayesian Bootstrap and we can sample such a P via

$$P = \sum_{i=1}^{n} w_{n,i}\delta_{x_i} \tag{5.5}$$

where the weights $(w_{n,1},\ldots,w_{n,n})$ are drawn from a Dirichlet $(1,\ldots,1)$ distribution. This follows quite easily from the representation of a Dirichlet process from its finite dimensional distributions; see Rubin (1981).

Consider now

$$U(\theta,P) = \sum_{i=1}^{n} w_{n,i}\log f(x_i;\theta), \tag{5.6}$$

which is maximized via the maximization of the weighted likelihood bootstrap

$$\prod_{i=1}^{n} f(x_i,\theta)^{w_{n,i}}.$$

The posterior distribution of θ obtained by the weighted likelihood bootstrap approach is equivalent and hence justified by the random utility model and the theory we have presented in Section 4.3. For us, the distribution created on Θ space is a valid distribution of beliefs about the parameter θ when the current state of beliefs about the true distribution function is represented by the Bayesian bootstrap. Such a distribution could be warranted in robust parameter estimation problems. Of course now we would recommend the mode of the density function induced on Θ space as the optimal parameter estimate.

(5.2) Portfolio analysis. In finance, investors face the problem of capital allocation in assets under uncertainty. Investors diversify, allocating a proportion of the capital into different assets. Moreover, according to maximum expected utility, the rational investor will try to optimize the proportion invested in the different assets in order to get the maximum expected total return. However, due to the volatility of assets, the combination yielding maximum expected return can lead to a higher risk, therefore another goal of the investors is to minimize the risk or the variability, which is another reason why diversification is recommended.

According to [25], the criteria to allocate portfolio diversification should not be only in terms of maximum expected return, but also looking to minimize the variance given that in many cases the assets with the highest returns tend to

have a higher risk associated. Hence, the idea of evaluating portfolio efficiency by the two elements is not in conflict, since decision makers tend to be risk averse in the domain of gains and to show risk preference in the domain of losses, and this is the basis for the mean–variance portfolio analysis. Therefore, there will be a set of portfolios that for a fixed level of expected return minimize the variance, or for a fixed variance maximize the expected return, these are called efficient portfolios, and the set constitutes what is called the efficient frontier. It is important to remark that the portfolio problem is considered in a single-time period situation and not over time. Let X be the total return of the portfolio, then the expected portfolio income or return and the variance of the expected return are defined in the equations below as a linear combination of the expected rate of returns of each asset:

$$E(X) = \sum_{i=1}^{n} w_i R_i \tag{5.7}$$

$$V(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j (R_i - E(R_i))(R_j - E(R_j)) \tag{5.8}$$

where w_i is the weight allocated in asset i and R_i is the corresponding annual expected return. The mean–variance theory optimizes the portfolio allocation by obtaining the efficient frontier, i.e. all the possible portfolios that have minimal variance given a certain level of returns or that have a maximal return for a given level of variance. Then the investor may choose any portfolio in the efficient frontier, in fact it is recommended to choose assets that move in opposite directions given that this will compensate the risk. The classic portfolio theory assumes that the vector of returns is normally distributed with mean $\mathbf{R}$ and covariance matrix $\mathbf{Q}$. Thus the portfolio return also has a normal distribution with mean $E(X) = \mathbf{R^T w}$ and variance V(X)=$\mathbf{w^T Q w}$. Assuming also that the utility function of the total return is exponential, i.e.

$$v(X) = 1 - \exp\{-kx\},$$

where $k > 0$ is a constant related to risk aversion. Hence, the expected value of the preference scaling function is given by

$$\begin{aligned} E(v(X)) &= \int_{-\infty}^{\infty} (1 - \exp\{-kx\}) P(\mathrm{d}x) \\ &= 1 - \exp\left\{-kE(X) + \frac{1}{2}k^2 V(X)\right\} \\ &= 1 - \exp\left\{-k\mathbf{R^T w} + \frac{1}{2}k^2 \mathbf{w^T Q w}\right\}, \end{aligned} \tag{5.9}$$

where $P(\mathrm{d}x)$ is the density function of the total returns. The exponential utility and the normal distribution of returns are assumed due to their convenience of giving exact results for the expected utility maximization. Although it is well known that the returns are not normally distributed. Hence, due to the monotonicity of the exponential function, maximizing (5.9) is equivalent to maximizing the following:

$$U\Big(\mathbf{w}, (\mathbf{R}, \mathbf{Q})\Big) = R^T \mathbf{w} - \frac{k}{2} \mathbf{w}^T Q \mathbf{w}. \tag{5.10}$$

[26] argues that the reason for performing a mean–variance analysis is more convenient in terms of costs and feasibility rather than doing the theoretical expected utility maximization. Also, he says that it is more expensive to find a utility maximizing portfolio than to trace out an entire mean–variance frontier. Hence, the need to maximize the part between the brackets of (5.9) or equivalently (5.10), with respect to $\mathbf{w}$, subject to the constraints $\sum_{i=1}^{n} w_i = 1$ and $w_i > 0$. This idea assumes $(\mathbf{R}, \mathbf{Q})$ are known. However, using our approach in section 4.3 and 5.1, we assign a possible prior distribution to $(\mathbf{R}, \mathbf{Q})$, say

$$\boldsymbol{R} \sim N(\boldsymbol{\mu}, \Sigma) \tag{5.11}$$

$$\boldsymbol{Q}^{-1} \sim W(\rho, (\rho\Omega)^{-1}). \tag{5.12}$$

From this we can simulate $(\mathbf{R}, \mathbf{Q})$, yielding a random maximizer $\mathbf{w}$ of

$$h(\mathbf{w}; \mathbf{R}, \mathbf{Q}) = \mathbf{R}^T\mathbf{w} - \frac{k}{2}\mathbf{w}^T\mathbf{Q}\mathbf{w}. \tag{5.13}$$

Hence, simulation of $(\mathbf{R}, \mathbf{Q})$ builds up a distribution on $\mathbf{w}$ space. For the illustration of this example we have a portfolio of three assets. For the simulation, the following values were chosen; $\boldsymbol{\mu} = (0.1, 0.2, 0.15)$, $\rho = 3$, $k = 3$

$$\Omega = \Sigma = \begin{pmatrix} 0.005 & -0.010 & 0.004 \\ -0.010 & 0.040 & -0.002 \\ 0.004 & -0.002 & 0.023 \end{pmatrix}.$$

Then we procede to do the maximization using the quadratic programming with 100,000 iterations. The three–dimensional histogram in Figure 1 for the frequencies of (w_2, w_3) shows a diagonal histogram as a consequence of the symmetric covariance. This confirms that it is always better to diversify in assets that are complements, because they will compensate the risk. Complement assets will give returns in the opposite direction, hence they will have negative correlation. A histogram with very thin bins will produce a good estimate of the mode. The mode corresponds to the vector of weights $\widehat{\mathbf{w}} = (0, 0.51, 0.49)$. With this we confirm that the asset 1 is not significant and the other two assets compensate, permitting almost the same allocation of capital.

6. Discussion

Normative theories of choice are not satisfied by all individuals, less so when aggregated into a group, and descriptive theories do not provide a decision rule. Moreover, group decision theory cannot give a decision rule because of the existence of conflicting interests and when decisions are taken by voting schemes, each strategy or counting rule produces a different winner. Whereas, probabilistic choice models have been for the benefit of an observer of individuals making choices.

Our perspective is of a decision maker modeling herself who having followed axioms of rational behavior is equipped with a utility function $U(a, \theta)$ which depends on an unknown state of the world θ and a subjective probability for it $\Pi(\theta)$. The random or probabilistic aspect is clear. The best action a^* is uncertain and has a probability $P(a^*)$ attached to it simply because the whereabouts of the unknown state of the world is uncertain and has a probability for it lying in any

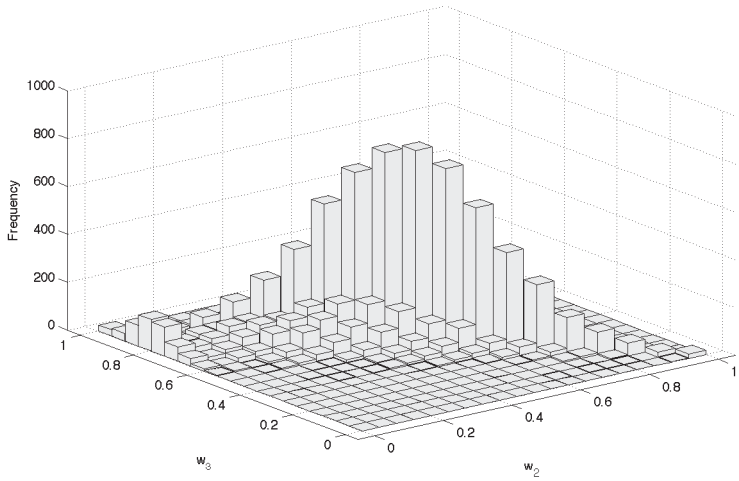


Figure 1. Histogram of w_2, w_3.

particular set or interval. So just as in probabilistic choice theory the decision maker can write down the probability of choices.

The inclusion of the states of the world as a latent variable in the RUM's representation is novel yet necessary since it captures the uncertainty about the states of the world as well as the randomness in the choice itself. On the other hand, to make a group decision when the simple majority rule applies, pooling the corresponding individual utilities and beliefs, and finding the mode of the group probability distribution is equivalent to a voting mechanism without actually asking the group to vote. The new proposal is to replace the decision rule of maximizing the expected utility $U(a)$ with the maximization of $P(a)$.

Acknowledgements

The authors would like to thank an anonymous referee for comments which have lead to an improvement of the presentation of the paper.

Received March 21, 2013

Final version received October 24, 2013

María de los Dolores Sánchez Castañeda
CIMAT (Mathematical Research Centre)
Fray Bartolomé de las Casas 314,
Aguascalientes, 20259, México
maria.sanchez@cimat.mx

Stephen G. Walker
Department of Mathematics and
Division of Statistics and Scientific Computation
University of Texas at Austin, USA
s.g.walker@math.utexas.edu

REFERENCES

[1] K. J. ARROW, *Social Choice and Individual Values*, Cowles Foundation for Research in Economics at Yale University, John Wiley & Sons, Inc., USA, 1963.

[2] G. M. BECKER, M. H. DEGROOT, AND J. MARSCHAK, *Stochastic models of choice behavior*, Behavioral Science, **8**, (1963), 41–55.

[3] H. BLOCK AND J. MARSCHAK, *Random orderings and stochastic theories of response*, Cowles Foundation Discussion Papers 66, Cowles Foundation, Yale University, 1959.

[4] D. DAVIDSON AND J. MARSCHAK, *Experimental tests of stochastic decision theory*, in Measurement Definitions and Theories, C. West Churchman and Ratoos, eds., John Wiley and Sons, 1959.

[5] J. FALMAGNE, *A representation theorem for finite random scale systems*, J. Math. Psych., **18**, (1978), 52–72.

[6] T. S. FERGUSON, *A Bayesian analysis of some nonparametric problems*, Ann. Statist., **1** (2), (1973), 209–230.

[7] P. C. FISHBURN, *A probabilistic model of social choice: Comment*, Rev. Econom. Stud., **42** (2), (1975), 297–301.

[8] ———, *Probabilistic social choice based on simple voting comparisons*, Rev. Econom. Stud., **51** (4), (1984), 683–692.

[9] S. FRENCH AND D. RÍOS INSUA, *Statistical Decision Theory*, Kendall's Library of Statistics, Arnold, London, England, 2000.

[10] C. GENEST, *A characterization theorem for externally Bayesian groups*, Ann. Statist., **12** (3), (1984), 1100–1105.

[11] C. GENEST AND J. V. ZIDEK, *Combining probability distributions: A critique and an annotated bibliography*, Statist. Sci., **1** (1), (1986), 114–135.

[12] D. HEYER AND R. NIEDERÉE, *Elements of a model-theoretic framework for probabilistic measurement.*, in Mathematical Psychology in progress, E. Roskam, ed., Berlin: Springer, 1989, 99–112.

[13] ———, *Generalizing the concept of binary choice systems induced by rankings: One way of probabilizing detrministic measurement structures.*, Mathematical Social Sciences, **23**, (1992), 31–44.

[14] J. HIRSHLEIFER AND J. G. RILEY, *The Analytics of Uncertainty and Information*, Cambridge University Press, Great Britain, 1992.

[15] M. D. INTRILIGATOR, *A probabilistic model of social choice*, Rev. Econom. Stud., **40** (4), (1973), 553–560.

[16] D. KAHNEMAN AND A. TVERSKY, *Prospect theory: An analysis of decision under risk*, Econometrica, **47** (2), (1979), 263–292.

[17] ———, *Choices, values and frames*, American Psychologist, **39** (4), (1984), 341–350.

[18] R. L. KEENEY AND H. RAIFFA, *Decisions with Multiple Objectives, Preferences and Value Tradeoffs*, Cambridge University Press, 1976.

[19] D. KRANTZ, *A survey of measurement theory*, in Mathematics of the Decision Sciences, Part 2, G. B. Dantzig and J. Arthur F. Veinott, eds., American Mathematical Association, 1968, 314–350.

[20] D. R. LUCE, *Individual Choice Behavior*, John Wiley & Sons, Inc., USA, 1959.

[21] D. R. LUCE AND H. RAIFFA, *Games and Decisions*, John Wiley & Sons, Inc., USA, 1957.

[22] D. R. LUCE AND P. SUPPES, *Preference, utility, and subjective probability*, in Handbook of Mathematical Psychology, Vol. III, R. Luce, R. R. Bush, and E. Galanter, eds., John Wiley and Sons, Inc., USA, 1965.

[23] R. LUCE AND J. TUKEY, *Simultaneous conjoint measurement: a new scale type of fundamental measurement*, J. Math. Psych., **1**, (1964), 1–27.

[24] A. MADANSKY, *Externally Bayesian groups*. The RAND Corporation, RM-4141-PR, November 1964.

[25] H. M. MARKOWITZ, *Portfolio Selection: Efficient Diversification of Investment*, Yale University Press, London, 1959.

[26] ———, *Foundations of portfolio theory*, The Journal of Finance, **46** (2), (1991), 469–477.

[27] J. MARSCHAK, *Binary choice constraints on random utility indicators*, Cowles Foundation Discussion Papers 74, Cowles Foundation, Yale University, 1959.

[28] D. MCFADDEN, *Conditional logit analysis of quanlitative choice behavior*, in Frontiers in Econometrics, P. Zarembka, ed., New York: Academic Press, 1973, 105–142.

[29] P. A. MORRIS, *Combining expert judgments: A Bayesian approach*, Management Science, **23** (7), (1977), 679–693.

[30] H. MOSKOWITZ, *Some observation on theories of collective decisions*, in Utility, Probability and Human Decision Making, D. Wendt and C. Vlek, eds., vol. II, Dordrecht, Holland, 3-6 September 1973, Selected Proceedings of an Interdiciplinary Research Conference Rome, 381–395.

[31] M. A. NEWTON AND A. E. RAFTERY, *Approximate Bayesian inference with the Weighted Likelihood Bootstrap*, J. Roy. Statist. Soc. Ser. B, **56** (1), (1994), 3–48.

[32] F. RAMSEY, *Truth and probability.*, in The Foundations of Mathematics and other Logial Essays„ R. Braithwaite, ed., (1931) 1926, 156–198.

[33] M. REGENWETTER, *Random utility representations of finite m-ary relations*, J. Math. Psych., **40**, (1996), 219–234.

[34] M. REGENWETTER AND A. MARLEY, *Random relations, random utilities, and random functions*, J. Math. Psych., **45**, (2001), 864–912.

[35] D. RUBIN, *The Bayesian bootstrap*, Ann. Statist., **9**, (1981), 130–134.

[36] L. J. SAVAGE, *The Foundations of Statistics*, Dover Publications, Inc., New York, second revised edition, 1974 ed., 1954.

[37] P. SLOVIC AND A. TVERSKY, *Who accepts Savage's axiom?*, Behavioral Science, **19**, (1974), 368–373.

[38] J. VON NEUMANN AND O. MORGENSTERN, *Theory of Games and Economic Behavior*, Science Editions, John Wiley & Sons, Inc., USA, 1944.

[39] S. WEERAHANDI AND J. ZIDEK, *Elements of multi-Bayesian decision theory*, Ann. Statist., **11** (4), (1983), 1032–1046.

[40] S. WEERAHANDI AND J. V. ZIDEK, *Multi-Bayesian statistical decision theory*, J. Roy. Statist. Soc. Ser. A, **144** (1), (1981), 85–93.

[41] R. L. WINKLER, *The consensus of subjective probability distributions*, Management Science, **15** (2), (1968), B61–B75.