



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y  
DE LA ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA.

STRASSEN'S THEOREM AND ERLANGIZATION: APPLICATIONS TO RISK  
THEORY

TESIS  
QUE PARA OPTAR POR EL GRADO DE:  
MAESTRO EN CIENCIAS

PRESENTA:  
OSCAR PERALTA GUTIÉRREZ

DIRECTOR DE LA TESIS:  
MOGENS BLADT PETERSEN  
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN  
SISTEMAS

MÉXICO, D. F. JUNIO DE 2015



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

# Agradecimientos.

Tesis realizada gracias al Programa UNAM-DGAPA-PAPIIT IA101014.

# Contents

|   |           |
|---|-----------|
| <b>1 Preliminaries from Analysis.</b>                                 | <b>11</b> |
| 1.1 Basic definitions and known results. . . . .                      | 12        |
| 1.1.1 Topological spaces. . . . .                                     | 12        |
| 1.1.2 Normal spaces. . . . .  | 14        |
| 1.1.3 Topological vector spaces. . . . .                              | 15        |
| 1.1.4 Metric spaces. . . . .  | 16        |
| 1.1.5 Normed spaces. . . . .  | 17        |
| 1.1.6 Nets. . . . .   | 17        |
| 1.1.7 Set functions. . . . .  | 18        |
| 1.1.8 Weak topology on spaces of measures. . . . .                    | 19        |
| 1.1.9 Weak and weak* topologies. . . . .                              | 20        |
| 1.2 Upper semicontinuous and convex functions. . . . .                | 21        |
| 1.3 Compactness of measures. . . . .                                  | 24        |
| <b>2 Strassen's Theorem and convex order.</b>                         | <b>32</b> |
| 2.1 The general Strassen's Theorem. . . . .                           | 33        |
| 2.2 Strassen's Theorem for convex order. . . . .                      | 40        |
| 2.3 A review of convex order in $\mathbb{R}$ . . . . .                | 45        |
| 2.3.1 An actuarial view. . . . .                                      | 45        |
| 2.3.2 A mathematical view. . . . .                                    | 51        |
| <b>3 Phase-type distributions theory.</b>                             | <b>58</b> |
| 3.1 Markov jump processes. . . . .                                    | 60        |
| 3.2 Phase-type distributions. . . . .                                 | 61        |
| 3.3 Renewal theory for phase-type distributions. . . . .              | 66        |
| 3.4 Uniformization. . . . .   | 69        |
| 3.5 The least variable phase-type distribution of order $n$ . . . . . | 71        |

|          |   |            |
|----------|---|------------|
| <b>4</b> | <b>Fluid-flow and risk models.</b>            | <b>76</b>  |
| 4.1      | The fluid-flow model. . . . .                 | 77         |
| 4.2      | Risk models. . . . .                          | 86         |
| 4.2.1    | <i>PH</i> -Cramér-Lundberg process. . . . .   | 88         |
| 4.2.2    | <i>PH</i> -Sparre-Andersen process. . . . .   | 90         |
| <b>5</b> | <b>Finite-horizon Ruin Probabilities.</b>     | <b>94</b>  |
| 5.1      | The <i>PH</i> -Sparre-Andersen case. . . . .  | 95         |
| 5.2      | The <i>PH</i> -Cramér-Lundberg case . . . . . | 99         |
| 5.2.1    | Erlang case. . . . .                          | 100        |
| <b>6</b> | <b>Parisian Ruin Probabilities.</b>           | <b>104</b> |
| 6.1      | The <i>PH</i> -Sparre-Andersen case. . . . .  | 106        |
| 6.1.1    | Recovery probability. . . . .                 | 106        |
| 6.1.2    | Main result . . . . .                         | 109        |
| 6.2      | The <i>PH</i> -Cramér-Lundberg case. . . . .  | 113        |
| 6.2.1    | Erlang case. . . . .                          | 115        |

Conclusions and future work.

Bibliography

# Introducción.

La variabilidad de variables aleatorias siempre ha sido un tema interesante en la Teoría de Probabilidad. La forma más simple y usual de medirla es comparando las varianzas de dos o más variables aleatorias, inclusive es posible inducir un orden<sup>1</sup> en el espacio de variables aleatorias con la misma esperanza fija: sin embargo, nada garantiza que esta manera de medir la variabilidad sea la correcta.<sup>2</sup> En este manuscrito estudiamos una manera alternativa de medir la variabilidad: mediante el orden convexo de medidas, también conocido como mayorización. En principio, decimos que una medida de probabilidad  $\mu$  es mayor que  $\nu$  en el sentido del orden convexo (o que  $\mu$  mayoriza a  $\nu$ ) sii

$$\int f d\mu \geq \int f d\nu \text{ para toda } f \text{ convexa.}$$

El orden convexo es más fuerte que el orden inducido por la varianza. Además, este orden ha tenido una gran variedad de aplicaciones en finanzas y teoría del riesgo (ver Kaas et. al. [9] para algunos ejemplos).

En 1865, Volker Strassen probó un celebrado resultado en [18] que exhibe una equivalencia entre orden convexo y la existencia de un espacio de probabilidad con cierta “propiedad de martingala”.<sup>3</sup> Específicamente, si  $(S, \|\cdot\|)$  es un espacio de Banach, y  $\mu$  y  $\nu$  son medidas de probabilidad sobre él, tenemos que  $\mu$  mayoriza a  $\nu$  sii existe un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$  en el cual podemos definir variables aleatorias  $X : \Omega \rightarrow S$  y  $Y : \Omega \rightarrow S$  tales que  $X \sim \mu$ ,  $Y \sim \nu$  y  $\mathbb{E}(X|Y) = Y$  (bajo  $\mathbb{P}$ ). Este teorema, conocido comúnmente como teorema de Strassen para orden convexo (o para may-

---

<sup>1</sup>De hecho un orden parcial.

<sup>2</sup>Es difícil proponer una manera “óptima” sin dar argumentos subjetivos.

<sup>3</sup>De hecho, en dicho artículo se estudia un problema mucho más general y éste fue sólo un ejemplo.

orización), ha sido una herramienta poderosa para entender mejor la teoría detrás de la mayorización de medidas. Un gran “inconveniente” es que su prueba es extremadamente técnica, así que existen muy pocas pruebas en la literatura.<sup>4</sup>

En 1987, David Aldous y Larry Shepp probaron en [1] que dentro de la clase de distribuciones tipo fase<sup>5</sup>, las distribuciones Erlang son las menos variables en el sentido de coeficiente de variación.<sup>6</sup> Las distribuciones tipo fase han sido una herramienta eficiente en varias ramas de Probabilidad Aplicada, y este problema de variabilidad había sido estudiado mucho antes del famoso artículo de Aldous y Shepp. Algunos años después, Colm Art O’Cinneide probó en [13] que no solamente las distribuciones Erlang son las de menor variabilidad en el sentido de coeficiente de variación, sino también en el sentido del orden convexo: su prueba contiene una elegante aplicación del teorema de Strassen para mayorización. Una gran implicación de este resultado es la validación del método de erlangización, que consiste en aproximar una distribución puntual en  $c > 0$ , digamos  $\delta_c$ , con una secuencia de distribuciones Erlang. Específicamente, los artículos de O’Cinneide, y Aldous y Shepp prueban que si deseamos aproximar  $\delta_c$  con una secuencia de distribuciones tipo fase, el método de erlangización es la mejor manera de hacerlo.

La aplicación del método de erlangización en la que estamos interesados tiene dos partes:

- Por un lado, podemos usarlo para dar un método de aproximación de la probabilidad de ruina a tiempo finito para dos procesos de riesgo clásicos: el proceso Cramér-Lundberg y el proceso Sparre-Andersen. Éste ha sido un problema difícil de resolver, pero con la ayuda de las distribuciones tipo fase y el método de erlangización será posible aproximar arbitrariamente bien esta probabilidad. El enfoque que utilizaremos está basado en el trabajo de Asmussen et. al. [4].
- Por otro lado, deseamos considerar una manera moderna de declararse

---

<sup>4</sup>Existen pruebas alternativas más cortas del teorema de Strassen en el caso  $S = \mathbb{R}$ , pero nos centramos en estudiar la versión más general posible.

<sup>5</sup>Recordemos que una distribución tipo fase está asociada al tiempo de absorción de una cadena de Markov de saltos.

<sup>6</sup>El coeficiente de variación de una variable aleatoria  $X$  está definido como  $\text{Var}(X)/\text{E}(X)^2$ .

en bancarrota, llamado ruina parisina: esta manera de arruinarse, introducida por Angelo Dassios y Shanle Wu en [6], dicta que la reserva de la compañía de seguros puede quedar por debajo del nivel 0 sin declararse en bancarrota, siempre y cuando ésta se recupere antes de algún tiempo (posiblemente aleatorio). Existe muy poca literatura acerca de este tipo de ruina, y la mayoría está inclinada hacia el estudio de la ruina parisina para procesos de Lévy, cuyos procedimientos son más analíticos que probabilísticos. Uno podría imaginar que un estudio análogo al de Asmussen [4] puede ser llevado a cabo para la ruina parisina en procesos Cramér-Lundberg y Sparre-Andersen.

La estructura de este manuscrito es la siguiente:

En el Capítulo 1 se da una breve revisión de algunos temas y resultados útiles del área de análisis. Este capítulo fue planeado para ser auto-contenido, aunque se hacen diversas referencias externas cuando nos topemos con resultados famosos de análisis funcional y topología. Además exponemos dos resultados no-tan-conocidos que necesitaron ser demostrados: que una función semicontinua superiormente y convexa es siempre continua, y una condición para la compacidad de una red de medidas de Radon sobre un espacio normal basado en Tøpsoe [19].<sup>7</sup> Ambos serán de vital importancia en el Capítulo 2.

En el Capítulo 2 estudiamos el teorema de Strassen para el orden convexo en el caso de espacios de Banach. Para esto seguimos el trabajo de Hoffmann-Jørgensen en [8]. Varios detalles (y una corrección mínima) necesitaban ser añadidos para que esta prueba fuera totalmente entendible.<sup>8</sup> Más adelante, estudiamos el orden convexo en el caso  $\mathbb{R}$ : tal estudio nos permitirá comprender a la mayorización como un método para estudiar la variabilidad.

En el Capítulo 3 se estudia el bloque fundamental de nuestro trabajo: las distribuciones tipo fase. Se presentan algunas propiedades y algunos resultados auxiliares que se usarán en los capítulos siguientes. Después, seguimos el método de O’Cinneide [13] para validar el método de erlangización.

---

<sup>7</sup>El caso del espacio normal es bastante más complicado que el caso de espacio métrico, que es el más famoso y puede ser encontrado en una variedad de libros.

<sup>8</sup>Para estándares personales.

En el Capítulo 4 algunos modelos estocásticos son estudiados. El primero es llamado modelo de flujo de fluidos, el cual es un proceso que describe el nivel de fluido en una reserva sujeto a periodos aleatorios de llenado y vaciado: la teoría detrás de este proceso será de gran ayuda en los siguientes capítulos, específicamente, la teoría encontrada en Asmussen [3]. Después de esto, introduciremos dos modelos de riesgo: los procesos Crámer-Lundberg y Sparre-Andersen. Su probabilidad de ruina a tiempo infinito será calculada en el caso en que sus componentes tengan una distribución tipo fase.

En el Capítulo 5 trabajamos con la probabilidad de ruina a tiempo finito para el proceso Sparre-Andersen y el proceso Cramér-Lundberg, basado en el trabajo de Asmussen et. al. [4], que requiere de el encaje del proceso de riesgo en un modelo de flujo de fluidos y de un argumento de erlangización.

Finalmente, en el Capítulo 6 una teoría análoga a la del Capítulo 5 es desarrollada para la ruina parisina: esta es la parte novedosa del actual manuscrito.

# Introduction.

Variability of random variables has always been an interesting topic in Probability Theory. The simplest and most used way to measure it is by comparing the variance between two or more random variables, even an order<sup>9</sup> can be induced in the space of square integrable random variables with some fixed mean: nevertheless, nothing guarantees that this way of measuring variability is the correct one.<sup>10</sup> In this manuscript we study instead a different way of measuring variability: through the convex ordering of measures, also known as majorization. Namely, we say that some probability measure  $\mu$  is greater than  $\nu$  in the convex order sense (or that  $\mu$  majorizes  $\nu$ ) iff

$$\int f d\mu \geq \int f d\nu \text{ for all convex } f.$$

Convex order is stronger than the ordering induced by the variance, and it has a wide variety of applications in finance and risk theory (see Kaas et. al. [9] for some examples).

In 1965, Volker Strassen proved a celebrated result in [18] which gives an equivalence between convex ordering and the existence of a probability space with certain “martingale property”.<sup>11</sup> More specifically, if  $(S, \|\cdot\|)$  is a Banach space and  $\mu$  and  $\nu$  are probability measures over it, we have that  $\mu$  majorizes  $\nu$  iff there exists some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  in which we can define some random variables  $X : \Omega \rightarrow S$  and  $Y : \Omega \rightarrow S$  such that  $X \sim \mu$ ,  $Y \sim \nu$  and  $\mathbb{E}(X|Y) = Y$  (under  $\mathbb{P}$ ). This theorem, commonly known as Strassen’s theorem for convex order (or majorization), has been proved

---

<sup>9</sup>Actually a partial order.

<sup>10</sup>It is difficult to propose an “optimal” way without being subjective.

<sup>11</sup>His paper actually dealt with a much more general problem and this was only a quick example.

to be a powerful tool to get a better grasp of the theory of majorization of measures. A huge “drawback” is that its proof is extremely technical, so very few proofs of it are available on papers and books.<sup>12</sup>

In 1987, David Aldous and Larry Shepp proved in [1] that amongst the class of phase-type distributions<sup>13</sup>, Erlang distributions are the least variable ones in the coefficient of variation sense.<sup>14</sup> Phase-type distributions have quite an efficient tool in several branches of Applied Probability, and this problem of variability amongst them had been studied long before the famous paper of Aldous and Shepp. A few years later, Colm Art O’Cinneide proved in [13] that not only Erlang distributions are the least variable ones in the coefficient of variation sense, but also in the convex order sense: his proof uses an elegant application of Strassen’s theorem for majorization. A huge implication of this result is the validation of the method of erlangization, which consists in approximating a point mass at  $c > 0$ , say  $\delta_c$ , with a sequence of Erlang distributions. More specifically, O’Cinneide’s, and Aldous and Shepp’s papers prove that if we want to approximate  $\delta_c$  with a sequence of phase-type distributions, the erlangization method is the best way for doing it.

The application of the erlangization method we are interested in this manuscript is twofold:

- In one hand it can be used to give an approximation method for computing the finite-horizon probability of ruin for two classic risk processes: the Cramér-Lundberg and the Sparre-Andersen processes. This had been a difficult problem to solve, but with the aid of phase-type distributions and the erlangization method it is possible to approximate arbitrarily well this probability. The specific approach we are interested in is based in the work of Asmussen et. al. [4].
- On the other hand, we wish to consider a rather modern way of getting ruined, which is called parisian ruin: this way of ruin, introduced by Angelo Dassios and Shanle Wu [6], dictates that the reserve of a

---

<sup>12</sup>There exist shorter alternative proofs of Strassen’s theorem in the case  $S = \mathbb{R}$ , but we were interested in studying the most general one.

<sup>13</sup>Recall that a phase-type distribution is associated to the time of absorption of a Markov jump process.

<sup>14</sup>The coefficient of variation of a random variable  $X$  is defined as  $\text{Var}(X) / \mathbb{E}(X)^2$ .

company can get below level 0 without it being declared out of business, as long as it recovers before some (possibly random) time. There exists very little literature about this kind of ruin, and most of it is inclined towards the study of parisian ruin for Lévy processes, whose procedures are more analytical than probabilistic. One could imagine that an analogous study to that of Asmussen [4] could be carried on for parisian ruin for the Cramér-Lundberg and Sparre-Andersen process.

The structure of this manuscript is the following:

In Chapter 1 we give a brief review of useful topics and results borrowed from analysis. This chapter was made to be as self-contained as possible, although we do make external references whenever we encounter fairly famous results from functional analysis and topology. There are two not-so-known theorems which needed to be proved: that an upper semicontinuous and convex function is always continuous, and a condition for compactness of a net of Radon measures over a normal space based in Tøpsoe [19].<sup>15</sup> Both of them will be of huge importance in Chapter 2.

In Chapter 2 we review the Strassen's theorem for convex order in the Banach space case. To do this we follow the work of Hoffmann-Jørgensen in [8]. Several details (and one minor correction) needed to be added for this proof to be fully understandable.<sup>16</sup> Later, we make a study of convex order in the  $\mathbb{R}$  case: such study will give us more insight of why we use majorization as a way to measure variability.

In Chapter 3 the founding block of our work is studied: phase-type distributions. Some properties are presented, in addition to some auxiliary results which will be used in further chapters. Later, we follow the proof of O'Connide [13] to validate the erlangization method.

In Chapter 4 some useful stochastic models are reviewed. The first one is called a fluid-flow model, which is a process that describes fluid level in a reservoir subject to randomly determined periods of fillings and emptying: the theory behind this process will be of great help in further chapters, more

---

<sup>15</sup>The normal space case is far more difficult than the metric space case, which is the most famous one and can be found in a wide variety of books.

<sup>16</sup>By personal standards.

specifically the work of Asmussen [3]. After this, we introduce two classic risk models: the Crámer-Lundberg and Sparre-Andersen process, and their infinite-horizon probability of ruin is computed in the case their components are phase-type distributed.

In Chapter 5 we work with the finite-horizon probability of ruin for the Sparre-Andersen and the Cramér-Lundberg process, based in the work of Asmussen et. al. [4], which involves the embedding of the risk process in a fluid-flow model and an erlangization argument.

Finally, in Chapter 6 an analogous theory to that of Chapter 5 is developed for parisian ruin: this is the novel part of the manuscript.

# Chapter 1

## Preliminaries from Analysis.

In this chapter we introduce a wide variety of definitions and results in analysis which we will use throughout this text. Although most of them are classic, the reader might find some of them quite challenging and technical.

In Section 1.1 the most basic results are presented; even though the reader may be familiar with them, it is strongly recommended not to skip this section, since here we basically introduce all the notation used throughout the first part of this manuscript. The majority of definitions and results are based on Rudin [16], [17] and Willard [21].

In Section 1.2 we will prove that any upper semicontinuous and convex function over some topological vector space is actually continuous; this result will be of great help in Section 2.2. This section is based on Section 5.7 of Aliprantis et al. [2].

Finally, in Section 1.3 we will show a sufficient condition that guarantees the compactness of a net of measures over a fixed normal space; this result will be crucial in Section 2.1. This section is based on Tøpsoe [19].

## 1.1 Basic definitions and known results.

### 1.1.1 Topological spaces.

A **topological space** is a set  $X$  in which a collection  $\tau$  of subsets of  $X$  (called **open sets**) has been specified with the following properties:  $X$  is open,  $\emptyset$  is open, the intersection of any two open sets is open, and the union of every collection of open sets is open. Such a collection  $\tau$  is called a **topology on  $X$** . When we need to be clearer, we will write  $(X, \tau)$  to denote the topological space  $X$  equipped with the topology  $\tau$ .

A set  $E \subseteq X$  is **closed** iff  $X \setminus E$  is open. The **closure** of  $E$ , denoted by  $\overline{E}$ , is the intersection of all closed sets that contain  $E$ . The **interior of  $E$** , denoted by  $E^\circ$ , is the union of all open sets that are subsets of  $E$ . We say that  $\xi \subseteq \tau$  is a **base for  $\tau$**  if every element of  $\tau$  can be written as the union of elements in  $\xi$ . A **neighborhood of a point  $p \in X$**  is any open set that contains  $p$ . A collection  $\gamma$  of neighborhoods of a point  $p \in X$  is a **local base at  $p$**  if every neighborhood of  $p$  contains a member of  $\gamma$ . Every topology can be specified either by a base or by the collection of local bases at every point of  $X$ .  $(X, \tau)$  is a **Hausdorff space** and  $\tau$  is a **Hausdorff topology** iff for each two different points there exist two disjoint neighborhoods of said points. A set  $K \subseteq X$  is **compact** iff every open cover<sup>1</sup> of  $K$  has a finite subcover.

**Proposition 1.1.1.** *Compact sets of Hausdorff spaces are closed*

*Proof.* See Corollary to Theorem 2.5 (pp. 36) in Rudin [16]. □

**Theorem 1.1.1.** *Suppose  $K$  is compact and  $F$  is closed in a topological space  $X$ . Then  $F \cap K$  is compact.*

*Proof.* See Corollary to Theorem 2.5 (pp. 36) in Rudin [16]. □

**Proposition 1.1.2.** *The family  $\mathcal{K}$  of compact sets contains  $\emptyset$ , and is closed under finite unions and arbitrary intersections.*

*Proof.* The first two assertions are immediate from the definition of compactness. Proposition 1.1.1 yields that the arbitrary intersection of compact sets is closed; Theorem 1.1.1 shows that this intersection is also compact. □

---

<sup>1</sup>An open cover of a subset  $A \subseteq X$  is a collection of open subsets of  $X$  such that  $A$  is contained in the union of said collection.

**Theorem 1.1.2.** *If  $\{K_\alpha\}$  (countable or not) is a collection of compact subsets of a Hausdorff space and if  $\bigcap_\alpha K_\alpha = \emptyset$ , then some finite subcollection of  $\{K_\alpha\}$  also has empty intersection.*

*Proof.* See Theorem 2.6 (pp. 37) in Rudin [16]. □

**Definition 1.1.1.** Let  $X$  and  $Y$  be two topological spaces. The **product topology on  $X \times Y$**  is the one generated by the base

$$\xi = \{A \times B : A \text{ is open in } X, B \text{ is open in } Y\}.$$

From here on, when we encounter any Cartesian product of two topological spaces, we will assume that it is equipped with its product topology (unless otherwise stated).

**Theorem 1.1.3.**  *$K \times L$  is a compact subset of  $X \times Y$  iff  $K$  is compact in  $X$  and  $L$  is compact in  $Y$ .*

*Proof.* See Theorem 17.8 (pp. 120) in Willard [21]. □

**Lemma 1.1.1.** *Let  $K \times L$  be a compact subset of  $X \times Y$  and  $G$  an open subset of  $X \times Y$  such that  $K \times L \subset G$ . Then, there exist open subsets  $U$  and  $V$  of  $X$  and  $Y$  (respectively), such that*

$$K \times L \subset U \times V \subset G.$$

*Proof.* If either  $K = \emptyset$  or  $L = \emptyset$ , then it would suffice to take  $U = V = \emptyset$  for the statement to be true. Thus, suppose that  $K \neq \emptyset \neq L$ . Define

$$\Gamma_G = \{A \times B \subset G : A \text{ is open in } X, B \text{ is open in } Y\};$$

notice that the union of all the elements of  $\Gamma_G$  is in fact  $G$ . Fix  $x \in L$ . Clearly  $K \times \{x\}$  is a compact subset of  $G$ , so there exists a finite subcollection of  $\Gamma_G$ , say  $\gamma = \{A_i^x \times B_i^x\}_{i=1}^{n_x}$  (with  $n_x \in \mathbb{N}$ ), such that  $K \times \{x\} \subset \bigcup_{i=1}^{n_x} (A_i^x \times B_i^x)$ . If it happens that  $x \notin B_j^x$  for some  $j \in \{1, \dots, n_x\}$ , then  $(K \times \{x\}) \cap (A_j^x \times B_j^x) = \emptyset$ , so  $\gamma \setminus \{A_j^x \times B_j^x\}$  would still be a finite cover for  $K \times \{x\}$ ; this way, w.l.o.g. we can assume that  $x \in B_i^x$  for each  $i = 1, \dots, n_x$ . In other words, we have that

$$(K \times \{x\}) \subset W_x^1 \times W_x^2 \subset G$$

where  $W_x^1 := \bigcup_{i=1}^{n_x} A_i^x$  and  $W_x^2 := \bigcap_{i=1}^{n_x} B_i^x$  are both open subsets of  $X$  and  $Y$ , respectively.

Now define

$$\Delta = \{W_x^1 \times W_x^2 : x \in L\};$$

clearly  $\Delta$  covers  $K \times L$ , so there exists a finite subcollection of  $\Delta$ , say  $\delta = \{W_{x_k}^1 \times W_{x_k}^2\}_{k=1}^m$  (with  $m \in \mathbb{N}$ ), such that  $K \times L \subset \cup_{k=1}^m (W_{x_k}^1 \times W_{x_k}^2)$ . By construction we have that  $K \subset W_{x_i}^1$  for every  $k = 1, \dots, m$ . This means that

$$K \times L \subset U \times V \subset G$$

where  $U := \cap_{i=1}^m W_{x_i}^1$  and  $V := \cup_{i=1}^m W_{x_i}^2$  are both open subsets of  $X$  and  $Y$ , respectively.  $\square$

**Definition 1.1.2.** A mapping  $f$  from a topological space  $(X, \tau)$  into a topological space  $(Y, \varsigma)$  is **continuous** iff  $f^{-1}(U) \in \tau$  for every  $U \in \varsigma$ .

**Definition 1.1.3.** Let  $X$  be any topological space. Define  $C(X)$  to be the collection of continuous functions from  $X$  to  $\mathbb{R}$  and let

$$C_b(X) := \{f \in C(X) : f \text{ is bounded}\}.$$

Also, for any function  $\sigma : X \rightarrow \mathbb{R}$  which is continuous and bounded away from zero, we shall define the collection  $C_\sigma(X)^2$  by

$$C_\sigma(X) := \{f \in C(X) : f/\sigma \in C_b(X)\}.$$

## 1.1.2 Normal spaces.

A Hausdorff space  $X$  is **normal** iff whenever  $A$  and  $B$  are disjoint closed sets in  $X$ , there exist disjoint open sets  $U$  and  $V$  such that  $A \subseteq U$  and  $B \subseteq V$ .

**Lemma 1.1.2.** (*Urysohn's Lemma*) *A space  $X$  is normal iff whenever  $A \subseteq X$  is a closed set and  $B \subseteq X$  is an open set such that  $A \subseteq B$ , there exists a continuous mapping  $f : X \rightarrow [0, 1]$  with*

$$\mathbf{1}_A \leq f \leq \mathbf{1}_B.$$

*Proof.* See 15.6 (pp. 102) in Willard [21].  $\square$

---

<sup>2</sup> $b$  will never denote a function throughout this manuscript, avoiding any (possible) confusion concerning the definition of  $C_b(X)$ .

### 1.1.3 Topological vector spaces.

Let  $X$  be a vector space and  $\tau$  a topology on it such that

- every point of  $X$  is a closed set, and
- the vector space operations are continuous with respect to  $\tau$ .

Then  $\tau$  is said to be a **vector topology**, and  $X$  a **topological vector space**.

**Remark 1.1.1.** In this manuscript, we are only working with real vector spaces. Thus, from here on we shall omit the word “real” since any vector space we encounter will be a real vector space.

If  $X$  is a topological vector space, then the following notation will be used:

$$\begin{aligned}x + A &= \{x + a : a \in A\} \\x - A &= \{x - a : a \in A\} \\A + B &= \{a + b : a \in A, b \in B\} \\ \lambda A &= \{\lambda a : a \in A\},\end{aligned}$$

where  $\lambda \in \mathbb{R}$ ,  $x \in X$ , and  $A, B \subseteq X$ .

A set  $B \subseteq X$  is said to be **balanced** if  $\alpha B \subseteq B$  for every  $\alpha \in [-1, 1]$ .

**Theorem 1.1.4.** *In a topological vector space  $X$ , every neighborhood of 0 contains a balanced neighborhood of 0*

*Proof.* See Theorem 1.14 (pp. 12) in Rudin [17]. □

**Proposition 1.1.3.** *Every topological vector space is a Hausdorff space.*

*Proof.* See Theorem 1.10 (pp. 10) in Rudin [17]. □

A set  $C \subseteq X$  is said to be **convex** if

$$tC + (1 - t)C \subseteq C \quad \text{for all } t \in [0, 1].$$

**Theorem 1.1.5.** *Let  $X$  be a topological vector space. If  $C$  is a convex subset of  $X$ , so is  $\overline{C}$ .*

*Proof.* See Theorem 1.13 (pp. 11) in Rudin [17]. □

A topological vector space is said to be **locally convex** if for every  $x \in X$ , there exists a local base at  $x$  whose members are convex.

**Theorem 1.1.6.** *Let  $X$  be a locally convex topological vector space, let  $A \subseteq X$  be convex and closed, and let  $x_0 \in X$  be such that  $x_0 \notin A$ . Then there exists a continuous linear functional on  $X$ , say  $\Lambda_0$ , such that*

$$\Lambda_0(A) \leq 1 \text{ and } \Lambda_0 x_0 > 1.$$

*Proof.* See Corollary 5.80 (pp. 208) in Rudin [17]. □

### 1.1.4 Metric spaces.

Let  $X$  be a set.  $X$  is said to be a **metric space** if it is equipped with a function  $d : X \times X \rightarrow \mathbb{R}$  such that

1.  $0 \leq d(x, y) < \infty$  for all  $x, y \in X$ ,
2.  $d(x, y) = 0$  iff  $x = y$ ,
3.  $d(x, y) = d(y, x)$  for all  $x, y \in X$ ,
4.  $d(x, z) \leq d(x, y) + d(y, z)$  for all  $x, y, z \in X$ .

It is usual to call  $d(x, y)$  the **distance between  $x$  and  $y$** .

We can obtain a topology over any metric space  $X$  by declaring the local base at  $x$  to be the collection  $\{B_r(x)\}_{r>0}$ , where

$$B_r(x) = \{y \in X : d(x, y) < r\}.$$

From now on, when talking about metric spaces, we will imply that its topology is obtained this way.

**Theorem 1.1.7.** *Every metric space is normal.*

*Proof.* See 15.3 (pp. 100) in Willard [21]. □

A metric space  $X$  is said to be **complete** if every Cauchy sequence converges in  $X$ . In other words, if  $d(x_n, x_m) \rightarrow 0$  as  $n, m \rightarrow \infty$ , then there is some  $y \in X$  such that  $d(x_n, y) \rightarrow 0$  as  $n \rightarrow \infty$ .

### 1.1.5 Normed spaces.

Consider a real vector space  $X$ .  $X$  is said to be a **normed linear space** if to each  $x \in X$  there is an associated non-negative real number  $\|x\|$ , called the **norm of  $x$** , such that

1.  $\|x + y\| \leq \|x\| + \|y\|$  for all  $x, y \in X$ ,
2.  $\|\alpha x\| = |\alpha| \|x\|$  if  $x \in X$  and  $\alpha \in \mathbb{R}$ ,
3.  $\|x\| = 0$  implies  $x = 0$ .

Every normed linear space may be regarded as a metric space by defining the distance between  $x$  and  $y$  by  $\|x - y\|$ .

**Proposition 1.1.4.** *Every normed vector space is a normal topological vector space.*

*Proof.* That it is normal is a direct consequence of Theorem 1.1.7. It is easy to prove that the norm is continuous, so it is indeed a topological vector space.  $\square$

A **Banach space** is a normed linear space which is complete in the metric defined by the norm.

### 1.1.6 Nets.

Let  $X$  be a topological space. A **directed set** is a non-empty set  $D$  with a transitive and reflexive relation, denoted by  $\leq$ , such that any finite subset of  $D$  has an upper bound; i.e. if  $F$  is a finite subset of  $D$ , then there exists  $c \in D$  such that  $a \leq c$  for every  $a \in F$ . A **net** on  $X$  is a mapping of a directed set into  $X$ ; it is usually denoted by  $\{x_\alpha\}_{\alpha \in D}$  or  $\{x_\alpha\}$ . We say that a net  $\{x_\alpha\}_{\alpha \in D}$  is **eventually in**  $A$ , where  $A \subseteq X$ , if there exists  $a_0 \in D$  such that  $x_\alpha \in A$  for every  $\alpha \geq a_0$ .<sup>3</sup> A net  $\{x_\alpha\}_{\alpha \in D}$  is **universal** if for every subset  $A \subseteq X$  it is either true that  $\{x_\alpha\} \subset A$  eventually or else we have that  $\{x_\alpha\} \subset A^c$  eventually. Let  $\{x_\alpha\}_{\alpha \in D}$  and  $\{x_{\alpha_\beta}\}_{\beta \in E}$  be two nets on  $X$  such that the mapping  $\beta \rightarrow \alpha_\beta$  of  $E$  into  $D$  satisfies the requirement that for any  $a_0 \in D$ , we have  $\alpha_\beta \geq a_0$  eventually; then  $\{x_{\alpha_\beta}\}_{\beta \in E}$  is called a **subnet** of  $\{x_\alpha\}_{\alpha \in D}$ . The next theorem is a classic and easy result from General Topology.

---

<sup>3</sup>We are using the usual convention that the expression  $a \geq b$  is equivalent to  $b \leq a$ .

**Theorem 1.1.8.** *Every net has a universal subnet.*

Now suppose that  $X$  is a Hausdorff space. A net  $\{x_\alpha\}_{\alpha \in D}$  **converges to**  $x \in X$  if for every neighborhood  $U$  of  $x$ ,  $\{x_\alpha\} \subset U$  eventually; we denote this by  $\lim_\alpha x_\alpha = x$ .

**Definition 1.1.4.** A net is said to be **compact** if every universal subnet converges.

If  $\{x_\alpha\}_{\alpha \in D}$  happens to be a net on  $\mathbb{R}$ , then  $\limsup_\alpha x_\alpha$  and  $\liminf_\alpha x_\alpha$  are defined by

$$\limsup_\alpha x_\alpha = \inf_\alpha \left\{ \sup_{\beta \geq \alpha} x_\beta \right\}, \quad \liminf_\alpha x_\alpha = \sup_\alpha \left\{ \inf_{\beta \geq \alpha} x_\beta \right\}.$$

### 1.1.7 Set functions.

For now, let  $X$  be a topological space, and denote by  $\mathcal{G}$  the collection of open sets and  $\mathcal{K}$  the collection of compact sets; also let  $G, G_i$  be members of  $\mathcal{G}$ , and let  $K, K_i$  be members of  $\mathcal{K}$ . A **set function on  $\mathcal{F}$**  (where  $\mathcal{F}$  is a family of subsets of  $X$ ) is a non-negative, possibly infinite-valued function defined on the members of  $\mathcal{F}$ .

Let  $\beta$  be a set function on  $\mathcal{F}$  and let  $F, F_i$  be members of  $\mathcal{F}$ . Then<sup>4</sup>

- $\beta$  is **monotone** if  $F_1 \subseteq F_2$  implies that  $\beta(F_1) \leq \beta(F_2)$ .
- $\beta$  is **subadditive** if  $\beta(F_1 \cup F_2) \leq \beta(F_1) + \beta(F_2)$ .
- $\beta$  is **additive** if  $F_1 \cap F_2 = \emptyset$  implies that  $\beta(F_1 \cup F_2) = \beta(F_1) + \beta(F_2)$ .
- $\beta$  is called a **content** if  $\emptyset \in \mathcal{F}$ ,  $\mathcal{F}$  is closed under finite intersection and union, and  $\beta$  is finite, monotone, additive and subadditive.
- $\beta$  is **modular** if  $\emptyset \in \mathcal{F}$ , if  $\beta(\emptyset) = 0$ , and if  $\beta(F_1 \cup F_2) + \beta(F_1 \cap F_2) = \beta(F_1) + \beta(F_2)$ .
- A finite  $\beta$  is **tight** if  $F_1 \supseteq F_2$  implies that  $\sup\{\beta(F) : F \subseteq F_1 \setminus F_2\} = \beta(F_1) - \beta(F_2)$ .

---

<sup>4</sup>These statements are valid only when they make sense, that is, if  $\beta$  is correctly defined on the sets it is evaluated.

Now suppose that  $\mathcal{K} \subseteq \mathcal{F}$ ; then

- $\beta$  is  **$\sigma$ -smooth with respect to  $\mathcal{K}$**  if for any countable collection  $\mathcal{K}^*$  of elements of  $\mathcal{K}$  whose intersection is  $A_0$ , and  $A_0 \in \mathcal{F}$ , it is true that  $\beta(A_0) = \inf\{\beta(F) \mid \exists K \in \mathcal{K} : F \supseteq K\}$ , provided that the l.h.s. is finite. If  $\emptyset \in \mathcal{F}$  and we only require the last relation to hold when  $A_0 = \emptyset$ , then we say that  $\beta$  is  **$\sigma$ -smooth at  $\emptyset$  with respect to  $\mathcal{K}$** . If it happens that  $\mathcal{K} = \mathcal{F}$  in the previous settings, then we say that  $\beta$  is  **$\sigma$ -smooth** and  **$\sigma$ -smooth at  $\emptyset$** , respectively.
- $\beta$  is said to be **regular with respect to  $\mathcal{K}$**  if  $\beta(F) = \sup\{\beta(K) : K \subseteq F, K \in \mathcal{K}\}$ .

We need to be careful, because the previous statements were made for general set functions; some of those names, like tightness or regularity, have a different meaning when we are talking about measures.

**Definition 1.1.5.** Let  $\mathcal{B}(\mathcal{K})$  denote the smallest  $\sigma$ -field containing every set  $E \subseteq X$  for which  $K \cap E \in \mathcal{K}$  for every  $K \in \mathcal{K}$ . Define  $M_+(X)$  to be the set of finite, non-negative measures defined on  $\mathcal{B}(\mathcal{K})$ . Define  $M_+(X; \mathcal{K})$  to be the set of measures on  $M_+(X)$  which are regular with respect to  $\mathcal{K}$ ;  $M_+(X; \mathcal{K})$  is called the class of **finite Radon measures on  $X$** . Finally, for any function  $\sigma : X \rightarrow \mathbb{R}_+$  which is continuous and bounded away from 0, we shall define

$$Pr_\sigma(X) = \{\mu \in M_+(X; \mathcal{K}) : \mu(X) = 1 \text{ and } \sigma \text{ is } \mu\text{-integrable}\}.$$

**Remark 1.1.2.** We say that a topological space  $X$  is **compactly generated** if it satisfies the next condition: a set  $A$  is closed iff  $A \cap K$  is closed for every compact set  $K$ . This implies that if  $\mathcal{K}$  is the collection of compact sets in  $X$ , then  $\mathcal{B}(\mathcal{K})$  coincides with  $\mathcal{B}(X)$  (that is, the Borel  $\sigma$ -algebra of  $X$ ) when  $X$  is compactly generated. It can be shown that every metric space is compactly generated (see pp.282 in Munkres [12]), so in particular,  $\mathcal{B}(\mathcal{K}) = \mathcal{B}(X)$  whenever  $X$  is a metric space.

### 1.1.8 Weak topology on spaces of measures.

**Definition 1.1.6.** Let  $\mu \in M_+(X)$  and  $\{\mu_\alpha\} \subset M_+(X)$ . We say that  $\{\mu_\alpha\}$  converges in the *w-topology* iff

$$\int_X f d\mu_\alpha \rightarrow \int_X f d\mu \text{ for every } f \in C_b(X).$$

In other words, the  $w$ -topology on  $M_+(X)$  is the weakest topology that makes the mapping  $f \rightarrow \int_X f d\mu$  from  $C_b(X)$  to  $\mathbb{R}$  continuous.

**Theorem 1.1.9.** (*Portmanteau Theorem*) *If  $X$  is a normal space,  $\mu \in M_+(X; \mathcal{K})$  and  $\{\mu_\alpha\} \subset M_+(X; \mathcal{K})$ , then the following are equivalent:*

1.  $\mu_\alpha \xrightarrow{w} \mu$ .
2.  $\limsup \int_X f d\mu_\alpha \leq \int_X f d\mu$  for all  $f$  which are bounded from above and upper semicontinuous.
3.  $\liminf \int_X f d\mu_\alpha \geq \int_X f d\mu$  for all  $f$  which are bounded from below and lower semicontinuous.
4.  $\limsup \mu_\alpha(K) \leq \mu(K)$  for all  $K \in \mathcal{K}$  and  $\lim \mu_\alpha(X) = \mu(X)$ .
5.  $\liminf \mu_\alpha(G) \geq \mu(G)$  for all  $G \in \mathcal{G}$  and  $\lim \mu_\alpha(X) = \mu(X)$ .

*Proof.* See Theorem 8.1 (pp. 40) in Topsøe [20]. □

### 1.1.9 Weak and weak\* topologies.

Let  $X$  be a topological vector space. The **dual space of  $X$** , denoted by  $X^*$ , is the space whose elements are continuous linear functionals on  $X$ . It is regarded as a vector space when we define addition and scalar multiplication by

$$(\Lambda_1 + \Lambda_2)x = \Lambda_1x + \Lambda_2x, \quad (\alpha\Lambda)x = \alpha \cdot \Lambda x,$$

for all  $\Lambda_1, \Lambda_2, \Lambda \in X^*, \alpha \in \mathbb{R}$  and  $x \in X$ .

Let  $\mathcal{H}$  be a non-empty family of mappings  $f : X \rightarrow \mathbb{R}$ . Consider the weakest topology over  $X$  that makes every  $f \in \mathcal{H}$  continuous; this topology is called the **weak topology on  $X$  induced by  $\mathcal{H}$** . In the case  $\mathcal{H} = X^*$ , then this topology is called the **weak topology**.

Let  $\mathcal{H} \subseteq X^*$ ; we say that  $\mathcal{H}$  **separates the points of  $X$**  if for any pair  $x, y \in X$ , there exists  $\Lambda \in \mathcal{H}$  such that  $\Lambda x \neq \Lambda y$ .

**Theorem 1.1.10.** *Suppose  $X$  is a topological vector space and  $\mathcal{H}$  is a separating vector space of linear functionals. Then the weak topology on  $X$  induced by  $\mathcal{H}$  makes  $X$  into a locally convex space whose dual is  $\mathcal{H}$ .*

*Proof.* See Theorem 3.10 (pp. 64) in Rudin [17]. □

Now, instead of studying the topological properties of  $X$ , we shall study the structure of  $X^*$ . Note that every  $x \in X$  induces a linear functional  $f_x$  on  $X^*$  defined by

$$f_x \Lambda = \Lambda x, \quad \text{for all } \Lambda \in X^*.$$

Consider  $\Lambda, \Lambda' \in X^*$  such that  $f_x \Lambda = f_x \Lambda'$  for all  $x \in X$ ; then  $\Lambda x = \Lambda' x$  for all  $x$ , and so  $\Lambda = \Lambda'$ . This means that  $\{f_x : x \in X\}$  separates the points of  $X^*$ . Hence, we can topologize  $X^*$  with the weak topology induced by  $\{f_x : x \in X\}$ . This topology is called the **weak\* topology on  $X^*$** . Applying Theorem 1.1.10 we get immediately the following.

**Corollary 1.1.1.** *The weak\* topology makes  $X^*$  into a locally convex topological vector space, and every linear functional on  $X^*$  is of the form  $\Lambda \rightarrow \Lambda x$  for some  $x \in X$ .*

**Theorem 1.1.11.** *(Riesz-Markov Theorem) The dual space of  $C_b(X)$  (topologized by the weak\* topology) is homeomorphic to the the space of finite Radon measures on  $X$  (topologized by the  $w$ -topology).*

*Proof.* See Theorem 14.10 (pp. 495) in Aliprantis et al. [2]. □

Thus combining Corollary 1.1.1 and Theorem 1.1.11, we get the following:

**Corollary 1.1.2.** *The set of finite Radon measures equipped with the  $w$ -topology is a locally convex topological vector space, and every linear functional on it is of the form  $\mu \rightarrow \int f d\mu$  for some  $f \in C_b(X)$ .*

## 1.2 Upper semicontinuous and convex functions.

**Definition 1.2.1.** Let  $f$  be a real function on a topological space  $X$ . If

$$\{x : f(x) > \alpha\}$$

is open for every real  $\alpha$ , then  $f$  is said to be **lower semicontinuous**. We say that  $f$  is **lower semicontinuous at  $x_0$**  if for every  $\epsilon > 0$  there exists a neighborhood  $U$  of  $x_0$  such that  $f(x) \geq f(x_0) - \epsilon$  for all  $x \in U$ . If

$$\{x : f(x) < \alpha\}$$

is open for every real  $\alpha$ ,  $f$  is said to be **upper semicontinuous**. We say that  $f$  is **upper semicontinuous at**  $x_0$  if for every  $\epsilon > 0$  there exists a neighborhood  $U$  of  $x_0$  such that  $f(x) \leq f(x_0) + \epsilon$  for all  $x \in U$ . It is easy to prove that  $f$  is upper (lower) semicontinuous iff  $f$  is upper (lower) semicontinuous at  $x$  for every  $x \in X$ .

For this section, let  $X$  be a topological vector space.

**Lemma 1.2.1.** *If  $C$  is a convex subset of  $X$ , if  $f : C \rightarrow \mathbb{R}$  is convex, if  $x, x + z, x - z \in C$ , and if  $\delta \in [0, 1]$ , then*

$$|f(x + \delta z) - f(x)| \leq \delta \max\{f(x + z) - f(x), f(x - z) - f(x)\}.$$

*Proof.* Notice that  $x + \delta z = (1 - \delta)x + \delta(x + z)$ , so  $f(x + \delta z) \leq (1 - \delta)f(x) + \delta f(x + z)$ . Rearranging the previous inequality yields

$$f(x + \delta z) - f(x) \leq \delta[f(x + z) - f(x)]. \quad (1.2.1)$$

Replacing  $z$  with  $-z$  yields that

$$f(x - \delta z) - f(x) \leq \delta[f(x - z) - f(x)]. \quad (1.2.2)$$

On the other hand, note that  $x = (1/2)(x + \delta z) + (1/2)(x - \delta z)$ , so  $f(x) \leq (1/2)f(x + \delta z) + (1/2)f(x - \delta z)$ . This is equivalent to

$$f(x) - f(x + \delta z) \leq f(x - \delta z) - f(x); \quad (1.2.3)$$

thus, combining (1.2.2) and (1.2.3) we get that

$$f(x) - f(x + \delta z) \leq \delta[f(x - z) - f(x)]. \quad (1.2.4)$$

Then the theorem follows from (1.2.1) and (1.2.4).  $\square$

**Theorem 1.2.1.** *Let  $f : C \rightarrow \mathbb{R}$  be a convex function, where  $C$  is a convex subset of a topological vector space  $X$ . If there exists an interior point  $x$  and a neighborhood  $U$  of  $x$  such that  $f$  is bounded above on  $U$ , then  $f$  is continuous at  $x$ .*

*Proof.* By Theorem 1.1.4, there exists a balanced neighborhood of 0, say  $V$ , such that  $x + V \subseteq U$ . Fix  $y \in x + V$ . Let  $M > 0$  be such that

$$f(w) \leq f(x) + M \text{ for each } w \in x + V. \quad (1.2.5)$$

Fix  $\epsilon > 0$  and choose some  $\delta \in (0, 1)$  such that  $\delta M < \epsilon$ . Then, if we let  $y \in x + \delta V$  and  $z = (y - x)\delta$ , we have that

$$\begin{aligned} |f(y) - f(x)| &= |f(x + \delta z) - f(x)| \\ &= \delta \max\{f(x + z) - f(x), f(x - z) - f(x)\} \quad (\text{By Lemma 1.2.1}) \\ &\leq \delta M \quad (\text{Because } z, -z \in V \text{ and (1.2.5)}) \\ &< \epsilon; \end{aligned}$$

since this is true for every  $y \in x + \delta V$ , then  $f$  is continuous at  $x$ .  $\square$

**Theorem 1.2.2.** *If  $C$  is an open convex subset of  $X$ , and if  $f : C \rightarrow \mathbb{R}$  is convex, then the following are equivalent:*

1.  $f$  is continuous on  $C$ .
2.  $f$  is upper semicontinuous on  $C$ .
3. For each  $x \in C$ , there exists some neighborhood of  $x$  on which  $f$  is bounded above.
4. There is some  $x \in C$  and some neighborhood of  $x$  on which  $f$  is bounded above.
5. There is some  $x \in C$  at which  $f$  is continuous.

*Proof.* (1.  $\Rightarrow$  2.) Immediate.

(2.  $\Rightarrow$  3.) Fix  $x \in C$ . Since  $f$  is upper semicontinuous, then  $\{y \in C : f(y) < f(x) + 1\}$  is an open neighborhood of  $x$  on which  $f$  is bounded.

(3.  $\Rightarrow$  4.) Immediate.

(4.  $\Rightarrow$  5.) Follows from Theorem 1.2.1.

(5.  $\Rightarrow$  1.) Suppose  $f$  is continuous at  $x \in C$ . We need to prove that  $f$  is continuous at any  $y \in C$ ; fix  $y \in C$ . The mapping  $t \rightarrow x + t(y - x)$  of  $\mathbb{R}$  into  $X$  is continuous. Since it sends 1 to  $y$ , and  $C$  is open, then there is some  $t_0 > 1$  such that  $x + t_0(y - x) \in C$ . Set  $z = x + t_0(y - x) \in C$ ; then

$$y = \left(1 - \frac{1}{t_0}\right)x + \frac{1}{t_0}z,$$

or

$$y = \lambda_0 x + (1 - \lambda_0)z \text{ for some } \lambda_0 = 1 - 1/t_0 \in (0, 1).$$

Since  $f$  is continuous at  $x$ , then for every  $\epsilon > 0$  there exists a balanced neighborhood  $V$  of  $0$  such that if  $y \in x + V$ , then  $|f(y) - f(x)| < \epsilon$ ; since  $C$  is open,  $V$  can be chosen so that  $x + V \in C$ . Then there is some  $M > 0$  such that  $f(w) \leq M$  for  $w \in x + V$ . If  $v \in V$ , then

$$y + \lambda v = \lambda x + (1 - \lambda)z + \lambda v = \lambda(x + v) + (1 - \lambda)z; \quad (1.2.6)$$

since  $x + v, z \in C$ , then  $y + \lambda v \in C$ . Therefore  $y + \lambda V \subseteq C$ . Because  $f$  is convex,

$$\begin{aligned} f(y + \lambda v) &\leq \lambda f(x + v) + (1 - \lambda)f(z) \quad (\text{By (1.2.6)}) \\ &\leq \lambda M + (1 - \lambda)f(z). \end{aligned}$$

This holds for every  $v \in V$ , so in the neighborhood  $y + \lambda V$  of  $y$ , the function  $f$  is bounded by  $\lambda M + (1 - \lambda)f(z)$ . Hence applying Theorem 1.2.1, we get that  $f$  is continuous at  $y$ .  $\square$

The next corollary is immediate.

**Corollary 1.2.1.** *Let  $X$  be a topological vector space. If  $f : X \rightarrow \mathbb{R}$  is a convex upper semicontinuous function, then it is continuous.*

### 1.3 Compactness of measures.

Let  $X$  be a normal space and denote by  $\mathcal{G}$  and  $\mathcal{K}$  the families of open and compact sets in  $X$ , respectively. As before, we are going to use the letters  $G$  and  $K$  (with possible super-indices or sub-indices) to denote members of  $\mathcal{G}$  and  $\mathcal{K}$ , respectively.

**Lemma 1.3.1.** *Let  $\lambda$  be a finite set function on  $\mathcal{K}$  which is tight. Then it is also true that*

1.  $\lambda$  is monotone and modular,
2.  $\lambda$  is  $\sigma$ -smooth.

*Proof.* From the definition of tightness, we immediately get that  $\lambda$  is monotone. Also, note that

$$\lambda(\emptyset) = \sup\{\lambda(K) : K \subseteq \emptyset \setminus \emptyset = \emptyset\} = \lambda(\emptyset) - \lambda(\emptyset) = 0.$$

Since  $\lambda$  is defined only in  $\mathcal{K}$ , let us consider a new set function  $\mu$  over all the subsets of  $X$ , defined by

$$\mu(A) = \sup\{\lambda(K) : K \subset A\} \text{ for any } A \subseteq X. \quad (1.3.1)$$

The monotonicity of  $\lambda$  implies that  $\mu$  is an extension of  $\lambda$  which satisfies  $\mu(K_1 \setminus K_2) = \mu(K_1) - \mu(K_2)$  for all  $K_1 \supseteq K_2$  (since  $\lambda$  is tight). It follows from Proposition 1.1.2 that for any pair  $K'_1, K'_2$ , we have that

$$\mu(K'_1 \cup K'_2) - \mu(K'_2) = \mu(K'_1 \setminus K'_2) = \mu(K'_1) - \mu(K'_1 \cap K'_2), \quad (1.3.2)$$

and since  $\lambda$  and  $\mu$  coincides in  $\mathcal{K}$ , then

$$\lambda(K'_1 \cup K'_2) - \lambda(K'_2) = \mu(K'_1 \setminus K'_2) = \lambda(K'_1) - \lambda(K'_1 \cap K'_2); \quad (1.3.3)$$

hence  $\lambda$  is modular and then the statement 1 is proved.

To prove the statement 2, let us prove first that  $\lambda$  is  $\sigma$ -smooth at  $\emptyset$ . If  $\{K_n\}_{n \geq 1}$  is a decreasing sequence such that  $\bigcap_{n=1}^{\infty} K_n = \emptyset$ , then there exists  $N \in \mathbb{N}$  such that  $\bigcap_{n=1}^m K_n = \emptyset$  for all  $m \geq N$  (this follows from the Theorem 1.1.2). Then

$$0 = \lambda(\emptyset) = \lambda(\bigcap_{n=1}^N K_n) \geq \inf_{K \in \mathcal{K}} \{\lambda(K) : K \supseteq K_n \text{ for some } n \geq 1\} \geq 0,$$

so  $\lambda$  is  $\sigma$ -smooth at  $\emptyset$ .

Now, let  $\{K'_n\}_{n \geq 1}$  be a decreasing sequence of sets in  $\mathcal{K}$  such that  $K'_0 = \bigcap_{n=1}^{\infty} K'_n \in \mathcal{K}$ . Fix  $\epsilon > 0$  and choose  $K' \subset K'_1 \setminus K'_0$  such that  $\lambda(K') + \lambda(K'_0) \geq \lambda(K'_1) - \epsilon$ ; this choice can be made by using (1.3.3) and the tightness of  $\lambda$ . Choose  $n$  such that  $\lambda(K'_n \cap K') \leq \epsilon$ ; such an  $n$  exists because of the  $\sigma$ -smoothness at  $\emptyset$  of  $\lambda$ . Now,

$$\lambda(K'_n) = \lambda(K'_n \cup K') + \lambda(K'_n \cap K') - \lambda(K') \leq \lambda(K'_1) + \epsilon - \lambda(K') \leq \lambda(K'_0) - 2\epsilon,$$

so  $\lambda$  is actually  $\sigma$ -smooth.  $\square$

**Theorem 1.3.1.** *Let  $\lambda$  be a tight content on  $\mathcal{K}$ . Then  $\lambda$  can be extended to a measure  $\mu$  on  $\mathcal{B}(\mathcal{K})$  defined by  $\mu(A) = \sup\{\lambda(K) : K \subset A\}$  such that  $\mu$  is regular with respect to  $\mathcal{K}$ .*

*Proof.* Notice that  $\mu$  is exactly defined as in (1.3.1). Here we further assume that  $\lambda$  is content, so for fixed  $\epsilon > 0$ , and disjoint  $A, B \subseteq X$ , there exist  $K^1 \subseteq A$  and  $K^2 \subseteq B$  such that  $\lambda(K^1) \geq \mu(A) - \epsilon$  and  $\lambda(K^2) \geq \mu(B) - \epsilon$ , so  $\mu(A \cup B) \geq \mu(K^1 \cup K^2) = \lambda(K^1 \cup K^2) = \lambda(K^1) + \lambda(K^2) \geq \mu(A) + \mu(B) + 2\epsilon$ ; then it is true that

$$\mu(A \cup B) \geq \mu(A) + \mu(B) \quad (1.3.4)$$

for any disjoint subsets  $A$  and  $B$ . Next, consider the class

$$\mathcal{E} := \{E \subset X : \mu K = \mu(K \cap E) + \mu(K \setminus E) \quad \forall K \in \mathcal{K}\}.$$

Notice that  $\mathcal{E}$  is closed under complementation and that  $\mathcal{K} \subseteq \mathcal{E}$  (by (1.3.2)). Let us prove that  $\mathcal{E}$  is also closed under countable unions. Let  $\{E_n\}_{n \geq 1}$  be a sequence of sets in  $\mathcal{E}$ . Fix  $\epsilon > 0$  and  $K \in \mathcal{K}$ . Choose  $K'_n \subset K \cap E_n$  and  $K''_n \subset K \setminus E_n$  such that

$$\mu(K) \leq \mu(K'_n) + \mu(K''_n) + \epsilon 2^{-n}, \quad n \geq 1; \quad (1.3.5)$$

such sets can be found because of the very definitions of  $\mu$  and  $\mathcal{E}$ . The previous inequality implies that

$$\begin{aligned} & \lambda(K'_{n+1} \cap \{\cup_{i=1}^n K'_i\}) + \lambda(K''_{n+1} \cup \{\cap_{i=1}^n K''_i\}) \\ &= \lambda((K'_{n+1} \cap \{\cup_{i=1}^n K'_i\}) \cup (K''_{n+1} \cup \{\cap_{i=1}^n K''_i\})) \\ & \quad (\text{Because } \lambda \text{ is additive on } \mathcal{K} \text{ and Proposition 1.1.2}) \\ & \leq \lambda(K) \quad (\text{Because } \lambda \text{ is monotone and } K'_i, K''_i \subset K \quad \forall i \geq 1) \\ & \leq \lambda(K'_{n+1}) + \lambda(K''_{n+1}) + \epsilon 2^{-(n+1)} \quad (\text{Because of (1.3.5)}), \end{aligned}$$

which in turn, using that  $\lambda$  is modular on  $\mathcal{K}$ , is equivalent to

$$\lambda(\cup_{i=1}^{n+1} K'_i) + \lambda(\cap_{i=1}^{n+1} K''_i) \geq \lambda(\cup_{i=1}^n K'_i) + \lambda(\cap_{i=1}^n K''_i) - \epsilon 2^{-(n+1)},$$

and this finally implies that

$$\begin{aligned} \lambda(\cup_{i=1}^n K'_i) + \lambda(\cap_{i=1}^n K''_i) & \geq \lambda(K'_1) + \lambda(K''_1) - \epsilon \sum_{i=2}^n 2^{-i} \\ & \geq \lambda(K) - \epsilon \sum_{i=1}^n 2^{-i} \quad (\text{Because of (1.3.5)}) \\ & \geq \lambda(K) - \epsilon. \end{aligned}$$

Lemma 1.3.1 yields that  $\lambda$  is  $\sigma$ -smooth; then for sufficiently large  $n$  we have that

$$\lambda(\cup_{i=1}^n K'_i) + \lambda(\cap_{i=1}^\infty K''_i) \geq \lambda(K) - 2\epsilon,$$

so it follows from the definition of  $\mu$  that

$$\mu(K \cap \{\cup_{i=1}^\infty E_i\}) + \mu(K \setminus \{\cup_{i=1}^\infty E_i\}) \geq \lambda(K) - 2\epsilon;$$

this combined with (1.3.4) shows that  $\cup_{i=1}^\infty E_i \in \mathcal{E}$ . Using the subadditivity of  $\lambda$  we also get that

$$\sum_{i=1}^\infty \lambda(K'_i) + \lambda(\cap_{i=1}^\infty K''_i) \geq \lambda(K) - 2\epsilon.$$

It follows that

$$\sum_{i=1}^\infty \mu(E_i) + \mu(K \setminus \{\cup_{i=1}^\infty E_i\}) \geq \lambda(K) = \mu(K);$$

since  $\cup_{i=1}^\infty E_i \in \mathcal{E}$  and if we only consider  $K \subseteq \cup_{i=1}^\infty E_i$ , then

$$\sum_{i=1}^\infty \mu(E_i) \geq \mu(K \cap \{\cup_{i=1}^\infty E_i\}) = \mu(K) \quad \text{for all } K \subseteq \cup_{i=1}^\infty E_i.$$

This shows that

$$\mu(\cup_{i=1}^\infty E_i) = \sup \{ \mu(K) : K \subseteq \cup_{i=1}^\infty E_i \} \leq \sum_{i=1}^\infty \mu(E_i); \quad (1.3.6)$$

if the sequence  $\{E_n\}_{n \geq 1}$  were chosen to be disjoint, it is easy to see that equality in (1.3.6) must be attained, due to (1.3.4). This means that the restriction of  $\mu$  to  $\mathcal{E}$  has all the required properties.  $\square$

**Lemma 1.3.2.** *Let  $\nu$  be a monotone, additive and subadditive set function defined on  $\mathcal{G}$  such that to any  $K \in \mathcal{K}$  there exists  $G \in \mathcal{G}$  with  $K \subseteq G$  and  $\nu(G) < \infty$ . Define the set function  $\lambda$  on  $\mathcal{K}$  by  $\lambda(K) = \inf \{ \nu(G) : G \supseteq K \}$ . Then we have that  $\lambda$  is a tight content.*

*Proof.* It is immediate to see that  $\lambda$  is finite, monotone and subadditive.  $\mathcal{G}$  separates the sets in  $\mathcal{K}$  (since  $X$  is a normal space). Then, for disjoint sets

$K'_1, K'_2$  in  $\mathcal{K}$  choose disjoint sets  $G'_1, G'_2$  such that  $K'_1 \subseteq G'_1$  and  $K'_2 \subseteq G'_2$ . Using the properties of  $\nu$ , we get that

$$\begin{aligned}
\lambda(K'_1 \cup K'_2) &= \inf\{\nu(G) : G \supseteq K'_1 \cup K'_2\} \\
&\geq \inf\{\nu(\{G \cap G'_1\} \cup \{G \cap G'_2\}) : G \supseteq K'_1 \cup K'_2\} \\
&= \inf\{\nu(G \cap G'_1) + \nu(G \cap G'_2) : G \supseteq K'_1 \cup K'_2\} \\
&\geq \inf\{\nu(G \cap G'_1) : G \supseteq K'_1 \cup K'_2\} \\
&\quad + \inf\{\nu(G \cap G'_2) : G \supseteq K'_1 \cup K'_2\} \\
&\geq \inf\{\nu(G') : G' \supseteq K'_1\} + \inf\{\nu(G') : G' \supseteq K'_2\} \\
&= \lambda(K'_1) + \lambda(K'_2),
\end{aligned}$$

so  $\lambda$  is also additive. Next, assume that  $K_1 \supseteq K_2$ . Notice that for any  $K_0 \subseteq K_1 \setminus K_2$  with  $K_0 \in \mathcal{K}$ , we have that  $\lambda(K_0) + \lambda(K_2) = \lambda(K_0 \cup K_2) \leq \lambda(K_1)$  so

$$\sup\{\lambda(K) : K \subseteq K_1 \setminus K_2\} \leq \lambda(K_1) - \lambda(K_2). \quad (1.3.7)$$

Now fix  $\epsilon > 0$ . Choose  $G_2 \supseteq K_2$  such that  $\nu(G_2) \leq \lambda(K_2) + \epsilon$ . Put  $K = K_1 \setminus G_2$ ; then  $K \in \mathcal{K}$  because of Theorem 1.1.1. If  $G \supseteq K$ , then

$$\nu(G) \geq \nu(G \cup G_2) - \nu(G_2) \geq \lambda(K_1) - \nu(G_2) \geq \lambda(K_1) - \lambda(K_2) - \epsilon,$$

so it follows that  $\lambda(K) \geq \lambda(K_1) - \lambda(K_2) - \epsilon$ . In other words, we have that

$$\sup\{\lambda(K) : K \subseteq K_1 \setminus K_2\} \geq \lambda(K_1) - \lambda(K_2), \quad (1.3.8)$$

so the result follows from (1.3.7) and (1.3.8).  $\square$

**Theorem 1.3.2.** *Let  $\nu$  be a monotone, additive and subadditive set function defined on  $\mathcal{G}$  such that to any  $K \in \mathcal{K}$  there exists a  $G \in \mathcal{G}$  with  $G \supseteq K$  and  $\nu(G) < \infty$ . Then there exists a larger measure  $\mu$  regular w.r.t.  $\mathcal{K}$  such that  $\mu(G) \leq \nu(G)$  for all  $G \in \mathcal{G}$ . This measure is given by the formula*

$$\mu(A) = \sup_{K \subseteq A} \left\{ \inf_{G \supseteq K} \nu(G) \right\}, \quad A \in \mathcal{B}(\mathcal{K}).$$

*Proof.* Just apply Lemma 1.3.1 to construct a tight content  $\lambda$  on  $\mathcal{K}$  which is defined by  $\lambda(K) = \inf\{\nu(G) | G \supseteq K\}$ ; then apply Theorem 1.3.1 to construct a measure  $\mu$  on  $\mathcal{B}(\mathcal{K})$  defined by  $\mu(A) = \sup\{\lambda(K) | K \subseteq A\}$  such that  $\mu$  is

regular w.r.t.  $\mathcal{K}$ . Finally, fix  $G_0 \in G$  and  $K' \subseteq G_0$  such that  $K' \in \mathcal{K}$ . Then  $\inf_{G \supseteq K'} \nu(G) \leq \nu(G_0)$ , so it must happen that

$$\sup_{K' \subseteq G_0} \left\{ \inf_{G \supseteq K'} \nu(G) \right\} \leq \nu(G_0),$$

or in other words,  $\mu(G_0) \leq \nu(G_0)$ .  $\square$

**Lemma 1.3.3.** *If  $\{\mu_\alpha\}$  is a universal net of measures in  $M_+(X; \mathcal{K})$  and*

$$\limsup \mu_\alpha(X) < \infty, \tag{1.3.9}$$

*then  $\lim \mu_\alpha(A)$  exists and is finite for every  $A \in \mathcal{B}(\mathcal{K})$ . Moreover, the set function  $\nu$  on  $\mathcal{G}$  defined by  $\nu(G) = \lim \mu_\alpha(G)$  is monotone and modular.*

*Proof.* Fix  $A \in \mathcal{B}(\mathcal{K})$  and define

$$\beta_A = \inf\{a \geq 0 : \{\mu_\alpha(A)\} \subset [0, a] \text{ eventually}\}$$

and

$$\gamma_A = \sup\{a \geq 0 : \{\mu_\alpha(A)\} \subset [a, \infty) \text{ eventually}\}.$$

It is clear that  $\gamma_A \leq \beta_A$  and (1.3.9) implies that  $\beta_A < \infty$ . Assume that  $\gamma_A < \beta_A$  and fix  $\theta \in (\gamma_A, \beta_A)$ . Consider the subset  $\Delta \subseteq M_+(X; \mathcal{K})$  defined by

$$\Delta = \{\mu \in M_+(X; \mathcal{K}) : \mu(A) < \theta\};$$

then either  $\{\mu_\alpha\}$  is eventually in  $\Delta$  or  $\{\mu_\alpha\}$  is eventually in  $\Delta^c$ , since  $\{\mu_\alpha\}$  is a universal net. If we suppose the former were true, then  $\theta \geq \beta_A$ , which is a contradiction; if we suppose the latter were true, then  $\theta \leq \gamma_A$ , which is also a contradiction. Hence  $\gamma_A = \beta_A$ . This also implies that  $\lim \mu_\alpha(A)$  exists and is equal to  $\beta_A$ . That  $\nu$  is monotone and modular is straightforward to check.  $\square$

**Definition 1.3.1.** Let  $\mathcal{A}_i$  be a collection of subsets of  $X$ , for  $i = 1, 2$ . We shall say that  $\mathcal{A}_1$  dominates  $\mathcal{A}_2$  (and write  $\mathcal{A}_1 > \mathcal{A}_2$ ) if for all  $A_2 \in \mathcal{A}_2$  there exists  $A_1 \in \mathcal{A}_1$  such that  $A_1 \supseteq A_2$ .

**Theorem 1.3.3.** *Consider the space  $M_+(X; \mathcal{K})$  with the  $w$ -topology. Let  $\{\mu_\alpha\}_{\alpha \in D}$  be a net on  $M_+(X; \mathcal{K})$ . Then  $\{\mu_\alpha\}$  is compact (recall Definition 1.1.4) if the following conditions hold:*

1.  $\limsup_{\alpha} \mu_{\alpha}(X) < \infty$
2. For every subclass  $\mathcal{G}' \subset \mathcal{G}$  which dominates  $\mathcal{K}$ ,

$$\forall \epsilon > 0 \quad \exists G \in \mathcal{G}' : \limsup_{\alpha} \mu_{\alpha}(G^c) \leq \epsilon.$$

*Proof.* Suppose 1 and 2 hold and let  $\{\mu_{\alpha_{\beta}}\}$  be a universal subnet of  $\{\mu_{\alpha}\}$ . Applying Lemma 1.3.3, we get that the set function  $\nu : \mathcal{G} \rightarrow \mathbb{R}_+$  defined by  $\nu(G) = \lim \mu_{\alpha_{\beta}}(G)$  is monotone, additive and subadditive. Then, according to Theorem 1.3.2, there exists a measure  $\mu \in M_+(X, \mathcal{K})$  such that  $\mu(G) \leq \nu(G)$  for every  $G \in \mathcal{G}$ , which is actually defined by

$$\mu(A) = \sup_{K \subseteq A} \left\{ \inf_{G \supseteq K} \nu(G) \right\}, \quad A \in \mathcal{B}(\mathcal{K}). \quad (1.3.10)$$

Hence condition (5) from Theorem 1.1.9 is attained, so it is enough to prove that

$$\mu_{\alpha_{\beta}}(X) \rightarrow \mu(X) \quad (1.3.11)$$

to get that  $\mu_{\alpha_{\beta}} \xrightarrow{w} \mu$ . Assume that  $\mu(X) > \nu(X)$ . Hence, using (1.3.10) we note that there exists  $\delta > 0$  such that for all  $K \in \mathcal{K}$ ,

$$\inf_{G \supseteq K} \nu(G) + \delta > \nu(X),$$

which in turn is equivalent to

$$\delta > \nu(X) + \sup_{G \supseteq K} \{-\nu(G)\} = \sup_{G \supseteq K} \nu(G^c);$$

this means that for each  $K \in \mathcal{K}$  we can choose  $G_K \in \mathcal{G}$  such that  $G_K \supseteq K$  and  $\nu(G_K^c) > \delta/2$ , so  $\mu_{\alpha_{\beta}}(G_K^c) > \delta/4$  happens eventually for all  $K \in \mathcal{K}$ . Notice that  $\{G_K\}_{K \in \mathcal{K}}$  is a subclass of  $\mathcal{G}$  that dominates  $\mathcal{K}$ , but condition 2 is not attained. Hence we conclude that (1.3.11) must be true, thus  $\mu_{\alpha_{\beta}} \xrightarrow{w} \mu$ .  $\square$

**Remark 1.3.1.** It is possible to show that the conditions 1 and 2 in the statement of Theorem 1.3.3 are not only sufficient, but also necessary; however, in this manuscript necessity will not be crucial. See Tøpsoe [19] for more details.

**Lemma 1.3.4.** *Let  $\{\gamma_\alpha\}_\alpha$  be a net in  $M_+(S \times T)$  and let  $\mu_\alpha$  and  $\nu_\alpha$  represent the marginals of each  $\gamma_\alpha$ ; that is,  $\mu_\alpha(\cdot) = \gamma_\alpha(\cdot \times T)$  and  $\nu_\alpha(\cdot) = \gamma_\alpha(S \times \cdot)$ . If  $\mu_\alpha \xrightarrow{w} \mu$  and  $\nu_\alpha \xrightarrow{w} \nu$  for some  $\mu, \nu \in M_+(S)$ , then there exists a  $w$ -limit point  $\gamma \in M_+(S \times T)$  with marginals  $\mu$  and  $\nu$ .*

*Proof.* It is clear that any limit point of  $\{\gamma_\alpha\}$  must have marginals  $\mu$  and  $\nu$ ; if it were not the case, then either  $\mu_\alpha \not\xrightarrow{w} \mu$  or  $\nu_\alpha \not\xrightarrow{w} \nu$ . It follows that it is enough to prove the existence of a  $w$ -limit point of  $\{\gamma_\alpha\}$ . Let us verify the compactness criteria provided by Theorem 1.3.3; we may check that

1.  $\limsup_\alpha \gamma_\alpha(S \times T) < \infty$ , and that
2. for each  $\epsilon > 0$  there exists  $G \in \mathcal{G}'_{S \times T}$  such that  $\limsup_\alpha \gamma_\alpha(G^c) \leq \epsilon$ ,

whenever  $\mathcal{G}'_{S \times T}$  is a family of open subsets of  $S \times T$  that dominates the family  $\mathcal{K}_{S \times T}$  of compact sets of  $S \times T$ .

By Theorem 1.1.9, the statement 1 is trivially true since

$$\limsup_\alpha \gamma_\alpha(S \times Y) = \limsup_\alpha \mu_\alpha(S) < \infty.$$

Next, fix  $\epsilon > 0$ . Since  $\mu$  and  $\nu$  are Radon measures, there exist compact sets  $L_1 \subset T$  and  $L_2 \subset S$  such that  $\mu(L_1^c) \leq \epsilon/2$  and  $\nu(L_2^c) \leq \epsilon/2$ .  $L_1 \times L_2$  is a compact subset of  $S \times T$  (see Theorem 1.1.3). Let  $\mathcal{G}'_{S \times T}$  be such that  $\mathcal{G}'_{S \times T} > \mathcal{K}_{S \times T}$ ; then there exists  $G \in \mathcal{G}'_{S \times T}$  such that  $G \supseteq L_1 \times L_2$ . By Lemma 1.1.1, there exist open sets  $U \supseteq L_1$  and  $V \supseteq L_2$  such that  $U \times V \subset G$ .

By Theorem 1.1.9 again,

$$\limsup_\alpha \mu_\alpha(U^c) \leq \mu(U^c) \leq \epsilon/2 \text{ and } \limsup_\alpha \nu_\alpha(U^c) \leq \nu(U^c) \leq \epsilon/2,$$

and since  $G^c \subset (U \times V)^c \subset (U^c \times T) \cup (S \times V^c)$ , we have that

$$\limsup_\alpha \gamma_\alpha(G^c) \leq \epsilon,$$

so the statement 2 holds. This means that  $\{\gamma_\alpha\}$  is a compact net (with respect to the  $w$ -topology), so a  $w$ -limit point must exist.  $\square$

## Chapter 2

# Strassen's Theorem and convex order.

We devote this chapter to study a famous result by Volker Strassen concerning the existence of probability measures with given marginals, which originally appeared in Strassen [18].

In Section 2.1 we give a condition for said existence following the procedure of Hoffmann-Jørgensen [8] in the case each marginal measure is defined over some normal topological vector space.

In Section 2.2 we still follow the work of Hoffmann-Jørgensen [8], but now we work within the case that both marginal measures are defined over a Banach space. Later, we will find a sufficient and necessary condition to guarantee the existence of a probability space with measure  $\mathbb{P}$  on which some random variables  $X \sim \mu$  and  $Y \sim \nu$  are defined such that  $X = \mathbb{E}(Y|X)$  (under  $\mathbb{P}$ ); this condition is the so-called convex order (or majorization) for the measures  $\mu$  and  $\nu$ . This result is commonly known as the Strassen's Theorem for convex order (or for majorization).

In Section 2.3 we will be interested in working only with probability measures over  $\mathbb{R}$ . First, we will study majorization within the framework of reinsurance. Later we will move on to a more mathematical point of view and connect the majorization of measures with the Strassen's theorem explained in Section 2.2: this connection will allow us to make a much more detailed study of majorization.

## 2.1 The general Strassen's Theorem.

Let  $S$  and  $T$  be normal spaces. For any  $\sigma : S \rightarrow \mathbb{R}_+$  which is continuous and bounded away from 0, recall the definitions of  $C_\sigma(S)$ ,  $M_\sigma(S)$ ,  $M_+(S)$  and  $Pr_\sigma(S)$  introduced in Subsections 1.1.1 and 1.1.7.

**Definition 2.1.1.** ( $w_\sigma$ -topology) The  $w_\sigma$ -topology on  $M_\sigma(S)$  is the smallest topology which makes the mapping  $\mu \rightarrow \int_S f d\mu$  continuous for every  $f \in C_\sigma(S)$ . In other words, we have that if  $\{\mu_\alpha\} \subset M_\sigma(S)$ , then  $\mu_\alpha \xrightarrow{w_\sigma} \mu$  iff  $\int_S f d\mu_\alpha \rightarrow \int_S f d\mu$  for each  $f \in C_\sigma(S)$ .

**Remark 2.1.1.** We can also consider the  $w_\sigma$ -topology on  $Pr_\sigma(S)$  instead of  $M_\sigma(S)$ .

**Proposition 2.1.1.** *The  $w$ -topology is the same as the  $w_1$ -topology. Also, the mapping  $\mu(\cdot) \rightarrow \int_S \sigma \mathbb{1}_{(\cdot)} d\mu$  is a homeomorphism of  $(M_\sigma(S), w_\sigma)$  onto  $(M(S), w)$ .*

*Proof.* It follows from the Definitions 2.1 and 1.1.6. □

**Lemma 2.1.1.** 1. If  $\mu \in M_\sigma(S)$  and  $f \in C_\sigma(S)$  then  $f \in L^1(\mu)$ .

2.  $C_\sigma(S) = \sigma C_b(S) := \{\sigma h : h \in C_b(S)\}$ .

3. The dual space of  $C_\sigma(S)$  is (homeomorphic to)  $M_\sigma(S)$ .

*Proof.* 1. Just notice that since  $|f/\sigma| \leq K$  for some  $K \in \mathbb{R}_+$ , then

$$\int_S |f| d\mu = \int_S \left| \frac{f}{\sigma} \right| \sigma d\mu \leq K \int_S \sigma d\mu < \infty.$$

2. Immediate from the Definition 1.1.3.

3. Let  $\phi$  be a linear functional on  $C_\sigma(S)$ . Denote by  $\phi^*$  the linear functional on  $C_b(S)$  defined by

$$\phi^*(h) = \phi(\sigma h) \quad \forall h \in C_b(S).$$

By the Riesz-Markov theorem (Th. 1.1.11), there exists a finite Radon measure  $\mu$  such that

$$\phi^*(h) = \int_S h d\mu \quad \forall h \in C_b(S).$$

By the point 2 of the current Lemma, we have that for any fixed  $f_0 \in C_\sigma(S)$  there exists  $h_0 \in C_b(S)$  such that  $f_0 = \sigma h_0$ . Then

$$\phi(f_0) = \phi^*(h_0) = \int_S h_0 d\mu = \int_S \frac{f_0}{\sigma} d\mu = \int_S f_0 \frac{d\mu}{\sigma} = \int_S f_0 d\mu',$$

where  $d\mu' := d\mu/\sigma$ . Notice that  $\sigma \in L_1(\mu')$ . By the Radon-Nikodym theorem, any measure in  $M_\sigma(S)$  can be constructed this way, so the proof is complete.  $\square$

**Proposition 2.1.2.**  $\mu_\alpha \xrightarrow{w} \mu$  and  $\lim_\alpha \int_S \sigma d\mu_\alpha = \int_S \sigma d\mu$  iff  $\mu_\alpha \xrightarrow{w\sigma} \mu$

*Proof.* Suppose that  $\mu_\alpha \xrightarrow{w\sigma} \mu$  holds. Then, according to the point 2 of Lemma 2.1.1,

$$\int_S h \sigma d\mu_\alpha \rightarrow \int_S h \sigma d\mu \quad \forall h \in C_b(S).$$

Taking  $h = g/\sigma$  with  $g \in C_b(S)$  in the previous limit implies that  $\mu_\alpha \xrightarrow{w} \mu$  and taking  $h \equiv 1$  implies that  $\lim_\alpha \int_S \sigma d\mu_\alpha = \int_S \sigma d\mu$ .

Now assume that  $\mu_\alpha \xrightarrow{w} \mu$  and  $\lim_\alpha \int_S \sigma d\mu_\alpha = \int_S \sigma d\mu$ ; let  $f \in C_\sigma(S)$  and let  $K \in \mathbb{R}_+$  be such that  $|f| \leq K\sigma$ . Define  $f_1 = f - K\sigma \leq 0$  and  $f_2 = f + K\sigma \geq 0$ . Then

$$\begin{aligned} \int_S f d\mu - K \int_S \sigma d\mu &= \int_S f_1 d\mu \\ &\geq \limsup_\alpha \int_S f_1 d\mu_\alpha \quad (\text{By point 2 of Theorem 1.1.9}) \\ &\geq \limsup_\alpha \int_S f d\mu_\alpha - K \int_S \sigma d\mu, \end{aligned}$$

which implies that

$$\int_S f d\mu \geq \limsup_\alpha \int_S f d\mu_\alpha. \quad (2.1.1)$$

Similarly,

$$\begin{aligned}
\int_S f d\mu + K \int_S \sigma d\mu &= \int_S f_2 d\mu \\
&\leq \liminf_{\alpha} \int_S f_2 d\mu_{\alpha} \quad (\text{By point 3 of Theorem 1.1.9}) \\
&\leq \liminf_{\alpha} \int_S f d\mu_{\alpha} + K \int_S \sigma d\mu,
\end{aligned}$$

which implies that

$$\int_S f d\mu \leq \liminf_{\alpha} \int_S f d\mu_{\alpha}; \tag{2.1.2}$$

combining (2.1.1) and (2.1.2), we get that in fact

$$\int_S f d\mu = \lim_{\alpha} \int_S f d\mu_{\alpha},$$

so the proof is complete.  $\square$

If  $f : S \rightarrow \mathbb{R}$  and  $g : T \rightarrow \mathbb{R}$ , let us define  $f \oplus g : S \times T \rightarrow \mathbb{R}$  by

$$f \oplus g(s, t) = f(s) + g(t) \quad \text{for all } (s, t) \in S \times T.$$

**Theorem 2.1.1.** *Let  $\sigma : S \rightarrow \mathbb{R}_+$  and  $\tau : T \rightarrow \mathbb{R}_+$  be continuous and bounded from below away from 0, and define  $\rho = \sigma \oplus \tau$ . Suppose that  $\mu \in Pr_{\sigma}(S)$ ,  $\nu \in Pr_{\tau}(T)$ ,  $\Lambda \subset Pr_{\rho}(S \times T)$  and that*

$$\int_S f d\mu + \int_T g d\nu \leq \sup_{\lambda \in \Lambda} \int (f \oplus g) d\lambda \quad \forall f \in C_{\sigma}(S) \text{ and } \forall g \in C_{\tau}(T). \tag{2.1.3}$$

*Then there exists  $\gamma \in Pr_{\rho}(S \times T)$  with marginals  $\mu$  and  $\nu$ , which satisfies*

$$\int_{S \times T} \varphi d\gamma \leq \sup_{\lambda \in \Lambda} \int \varphi d\lambda \quad \forall \varphi \in C_{\rho}(S \times T). \tag{2.1.4}$$

*If  $\Lambda$  is chosen to be convex and  $w_{\rho}$ -closed, then any  $\gamma$  that satisfies (2.1.4) must be a member of  $\Lambda$ .*

*Proof.* Let us prove the last assertion first. Theorem 1.1.6 and Corollary 1.1.2 yields that if we suppose that  $\Lambda$  is convex,  $w_{\rho}$ -closed and  $\gamma \notin \Lambda$ , then there exists  $\varphi_0 \in C_{\rho}(S \times T)$  such that  $\int_S \varphi_0 d\lambda \leq 1$  for all  $\lambda \in \Lambda$  and  $\int_S \varphi_0 d\gamma > 1$ ,

which is a contradiction to (2.1.4). Hence, the last statement of the theorem is true; now, we need to prove the existence of  $\gamma$ .

Let us define

$$\Gamma = \left\{ \gamma \in Pr_\rho(S \times T) : \int_{S \times T} \varphi d\gamma \leq \sup_{\lambda \in \Lambda} \int \varphi d\lambda \quad \forall \varphi \in C_\rho(S \times T) \right\}.$$

It is fairly easy to prove that  $\Gamma$  is a convex and  $w_\rho$ -closed subset of  $Pr_\rho(S \times T)$ . Notice that we actually have that

$$\sup_{\gamma \in \Gamma} \int \varphi d\gamma = \sup_{\lambda \in \Lambda} \int \varphi d\lambda \quad \forall \varphi \in C_\rho(S \times T) :$$

one inequality ( $\leq$ ) is implied by the definition of  $\Gamma$  and the other ( $\geq$ ) by the fact that  $\lambda \subset \Gamma$ .

Now let  $\Delta$  be the set of all  $(\zeta, \eta) \in Pr_\sigma(S) \times Pr_\tau(T)$  for which there exists  $\delta \in \Gamma$  with marginals  $\zeta$  and  $\eta$ . Then  $\Delta$  is a convex subset of  $M_\sigma(S) \times M_\tau(T)$ , which is the dual of  $C_\sigma(S) \times C_\tau(T)$  under the pairing

$$\langle (f, g), (\zeta, \eta) \rangle = \int_S f d\zeta + \int_S g d\eta; \quad (2.1.5)$$

indeed, since any continuous functional over  $C_\sigma(S) \times C_\tau(T)$  is of the form  $\phi_\sigma(f) + \phi_\tau(g)$ , where  $\phi_\sigma(\cdot)$  is some continuous functional over  $C_\sigma(S)$  and  $\phi_\tau(\cdot)$  is some continuous functional over  $C_\tau(T)$ , (2.1.5) follows from the point 3 of Lemma 2.1.1. Next, (2.1.5) and (2.1.3) imply that

$$\langle (f, g), (\mu, \nu) \rangle \leq \sup_{(\zeta, \eta) \in \Delta} \langle (f, g), (\zeta, \eta) \rangle .$$

The closure of convex sets are also convex (see Theorem 1.1.5), so the same argument at the beginning of this proof shows that  $(\mu, \nu) \in \overline{\Delta}^{w_\sigma \times w_\tau}$ . Hence there exists  $\{\gamma_\alpha\} \subset \Gamma$ , where each  $\gamma_\alpha$  has marginals  $\mu_\alpha$  and  $\nu_\alpha$  such that  $\mu_\alpha \xrightarrow{w_\sigma} \mu$  and  $\nu_\alpha \xrightarrow{w_\tau} \nu$ ; thus, by Proposition 2.1.2 we have that  $\mu_\alpha \xrightarrow{w} \mu$  and  $\nu_\alpha \xrightarrow{w} \nu$ . By means of Lemma 1.3.4 we have that  $\{\gamma_\alpha\}$  has a  $w$ -limit point  $\gamma$  with marginals  $\mu$  and  $\nu$ . Note that

$$\begin{aligned} \int_{S \times T} \rho d\gamma_\alpha &= \int_S \sigma d\mu_\alpha + \int_T \tau d\nu \\ &\rightarrow \int_S \sigma d\mu + \int_T \tau d\nu \quad (\text{Because } \sigma \in C_\sigma(S), \tau \in C_\tau(T)) \\ &= \int_S \sigma d\mu_\alpha. \end{aligned}$$

Proposition 2.1.2 yields the desired result.  $\square$

**Corollary 2.1.1.** (*The general Strassen's Theorem*) Let  $\sigma, \tau$  and  $\rho = \sigma \oplus \tau$  be as in Theorem 2.1.1. Let  $\Lambda$  be a convex  $w_\rho$ -closed subset of  $Pr_\rho(S \times T)$ , and for all  $s \in S$  let  $\mathcal{M}(s)$  be a subset of  $Pr_\tau(T)$  such that

$$\delta_s \times \eta \in \Lambda, \quad \forall s \in S, \forall \eta \in \mathcal{M}(s). \quad (2.1.6)$$

Also, for each  $g \in C_\tau(T)$  define

$$g^*(s) = \sup_{\eta \in \mathcal{M}(s)} \int_T g d\eta, \quad g_*(s) = \inf_{\eta \in \mathcal{M}(s)} \int_T g d\eta. \quad (2.1.7)$$

If  $\mu \in Pr_\sigma(S)$  and  $\nu \in Pr_\tau(T)$ , then the inequalities

$$\int_S f d\mu + \int_T g d\nu \leq \sup_{s \in S} \{f(s) + g^*(s)\} \quad \forall f \in C_\sigma(S), \forall g \in C_\tau(T) \quad (2.1.8)$$

and

$$\int_S f d\mu + \int_T g d\nu \geq \inf_{s \in S} \{f(s) + g_*(s)\} \quad \forall f \in C_\sigma(S), \forall g \in C_\tau(T) \quad (2.1.9)$$

are equivalent. Furthermore, if any of them holds, then there exists  $\gamma \in \Lambda$  with marginals  $\mu$  and  $\nu$ .

*Proof.* To prove the equivalence between (2.1.8) and (2.1.9), recall that if  $A \subset \mathbb{R}$ , then  $\inf(-A) = -\sup(A)$ , so  $(-g)_* = -(g^*)$  and thus

$$\begin{aligned} \int_S f d\mu + \int_T g d\nu &\leq \sup_{s \in S} \{f(s) + g^*(s)\} \\ \Leftrightarrow -\int_S f d\mu - \int_T g d\nu &\geq \inf_{s \in S} \{-f(s) - (g^*(s))\} \\ \Leftrightarrow \int_S (-f) d\mu + \int_T (-g) d\nu &\geq \inf_{s \in S} \{(-f)(s) + (-g)^*(s)\}; \end{aligned}$$

since  $f \in C_\sigma(S)$  iff  $-f \in C_\sigma(S)$  and  $g \in C_\tau(T)$  iff  $-g \in C_\tau(T)$ , the equivalence follows.

Next, define

$$\widehat{\varphi}(s) := \sup_{\eta \in \mathcal{M}(s)} \int_T \varphi(s, t) \eta(dt) = \sup_{\eta \in \mathcal{M}(s)} \int_{S \times T} \varphi d(\delta_s \times \eta)$$

for each  $\varphi \in C\rho(S \times T)$ . Then

$$\widehat{\varphi}(s) \leq \sup_{\lambda \in \Lambda} \int_{S \times T} \varphi d\lambda \quad \forall s \in S$$

since  $\delta_s \times \eta \in \Lambda$  for all  $\eta \in \mathcal{M}(s)$ . Hence

$$\sup_{s \in S} \widehat{\varphi}(s) \leq \sup_{\lambda \in \Lambda} \int_{S \times T} \varphi d\lambda. \quad (2.1.10)$$

Thus,

$$\begin{aligned} \widehat{f \oplus g}(s) &= \sup_{\eta \in \mathcal{M}(s)} \int_T (f \oplus g)(s, t) \eta(dt) \\ &= \sup_{\eta \in \mathcal{M}(s)} \int_T (f(s) + g(t)) \eta(dt) \\ &= \sup_{\eta \in \mathcal{M}(s)} \left\{ \int_T f(s) \eta(dt) + \int_T g(t) \eta(dt) \right\} \\ &= f(s) + \sup_{\eta \in \mathcal{M}(s)} \int_T g(t) \eta(dt) \\ &\quad (f \text{ does not depend on } t \text{ and } \eta(T) = 1) \\ &= f(s) + g^*(s). \end{aligned}$$

If we suppose that (2.1.8) holds, then

$$\begin{aligned} \int_S f d\mu + \int_T g d\nu &\leq \sup_{s \in S} (f \oplus g)^*(s) \\ &\leq \sup_{\lambda \in \Lambda} \int_{S \times T} (f \oplus g) d\lambda \quad (\text{By (2.1.10)}) \end{aligned}$$

for all  $f \in C_\sigma(S)$  and for all  $g \in C_\tau(T)$ , so the result follows from Theorem 2.1.1.  $\square$

From here on we shall work with the particular case in which  $S$  is a normal space,  $\sigma : S \rightarrow \mathbb{R}_+$  is a continuous and bounded from below away from 0 function, and we take  $T = S$ ,  $\tau = \sigma$  and  $\rho = \sigma \oplus \sigma$ . Also, let  $\Lambda \subset Pr_\rho(S \times S)$  and for each  $s \in S$  let  $\mathcal{M}(s) \subset Pr_\sigma(S)$  be such that

$$\delta_s \in \mathcal{M}(s) \quad \forall s \in S, \quad (2.1.11)$$

$$\Lambda \text{ is convex and } w_\rho\text{-closed}, \quad (2.1.12)$$

$$\delta_s \times \alpha \in \Lambda \quad \forall \alpha \in \mathcal{M}(s), \forall s \in S. \quad (2.1.13)$$

As before, define

$$g^*(s) = \sup_{\eta \in \mathcal{M}(s)} \int g d\eta, \quad g_*(s) = \inf_{\eta \in \mathcal{M}(s)} \int g d\eta \quad (2.1.14)$$

for  $g \in C_\sigma(S)$ .

**Definition 2.1.2.** A function  $f : S \rightarrow \overline{\mathbb{R}}$  is called  $\mathcal{M}$ -convex if

$$f(s) \leq af(t) + (1-a)f(u) \text{ and}$$

$$a\mathcal{M}(t) + (1-a)\mathcal{M}(u) \subset \mathcal{M}(s) \text{ whenever } 0 \leq a \leq 1.$$

The set of  $\mathcal{M}$ -convex functions is denoted by  $\mathcal{M}$ .

**Lemma 2.1.2.** *Under the previous assumptions, we have that*

1.  $g_* \leq g \leq g^* \quad \forall g \in C_\sigma(S)$ ,
2.  $g_* \in \mathcal{M} \quad \forall g \in C_\sigma(S)$ ,
3. If  $f, g \in \mathcal{M}$  and  $b \geq 0$ , then  $f + g$ ,  $bf$  and  $\max\{f, g\}$  belong to  $\mathcal{M}$ , and,
4. Every constant function is  $\mathcal{M}$ -convex.

*Proof.* 1. It is a trivial consequence of (2.1.11) and the definition (2.1.14).

2. Let  $g \in C_\sigma(S)$ , fix  $a \in [0, 1]$  and  $s, t, u \in S$  such that

$$a\mathcal{M}(t) + (1-a)\mathcal{M}(u) \subset \mathcal{M}(s).$$

Then

$$\begin{aligned} ag_*(t) + (1-a)g_*(u) &= a \inf_{\eta \in \mathcal{M}(t)} \int g d\eta + (1-a) \inf_{\zeta \in \mathcal{M}(u)} \int g d\zeta \\ &= \inf_{\eta \in \mathcal{M}(t), \zeta \in \mathcal{M}(u)} \left\{ \int g d(a\eta + (1-a)\zeta) \right\} \\ &\geq \inf_{\alpha \in \mathcal{M}(s)} \int g d\alpha \\ &\quad (\text{Because } a\mathcal{M}(t) + (1-a)\mathcal{M}(u) \subset \mathcal{M}(s)) \\ &= g_*(s), \end{aligned}$$

so  $g_*$  is  $\mathcal{M}$ -convex.

3. The proof is similar to the previous one.
4. It is obvious from the definition of  $\mathcal{M}$ -convexity. □

## 2.2 Strassen's Theorem for convex order.

Let  $S_1$  be a Banach space with norm  $\|\cdot\|_1$ , and let  $S_2$  be an ordered Banach space with norm  $\|\cdot\|_2$  and order  $\preceq$ . Also, let  $\sigma_i : S_i \rightarrow \mathbb{R}_+$  be defined by  $\sigma_i = \|\cdot\|_i + 1$ , and  $\rho_i = \sigma_i \oplus \sigma_i$ , for  $i = 1, 2$ . For this subsection, we shall define the triplet  $(S_1, \{\mathcal{M}_1(s)\}_{s \in S_1}, \Lambda_1)$  where

$$\mathcal{M}_1(s) = \left\{ \eta \in Pr_{\sigma_1}(S_1) : \int_{S_1} x \eta(dx) = s \right\}$$

and

$$\Lambda_1 = \left\{ \gamma \in Pr_{\rho_1}(S_1 \times S_1) : \begin{array}{l} \int y f(x) \gamma(dx, dy) = \int x f(x) \gamma(dx, dy) \\ \forall f \in C_b(S_1) \end{array} \right\},$$

and the triplet  $(S_2, \{\mathcal{M}_2(s)\}_{s \in S_2}, \Lambda_2)$  where

$$\mathcal{M}_2(s) = \left\{ \eta \in Pr_{\sigma_2}(S_2) : \int_{S_2} x \eta(dx) \succeq s \right\}$$

and

$$\Lambda_2 = \left\{ \gamma \in Pr_{\rho_2}(S_2 \times S_2) : \begin{array}{l} \int y f(x) \gamma(dx, dy) \succeq \int x f(x) \gamma(dx, dy) \\ \forall f \in C_b(S_1), f \geq 0 \end{array} \right\}.$$

**Theorem 2.2.1.** *For  $i = 1, 2$ , the triplet  $(S_i, \{\mathcal{M}_i(s)\}_{s \in S_i}, \Lambda_i)$  satisfies the conditions (2.1.11), (2.1.12) and (2.1.13). Also,  $\mathcal{M}_i$ -convexity implies ordinary convexity.*

*Proof.* The condition (2.1.11) is trivial for  $i = 1, 2$ . To check that the condition (2.1.12) is attained, just notice that

$$\left\| \frac{y f(x)}{\rho_i(x, y)} \right\|_i = \frac{\|y\|_i \|f(x)\|_i}{2 + \|x\|_i + \|y\|_i} \leq \|f(x)\|_i,$$

so  $y f(x) \in C_{\rho_i}(S_i \times S_i)$  whenever  $f \in C_b(S_i)$ . Similarly, we can check that  $x f(x) \in C_{\rho_i}(S_i \times S_i)$  if  $f \in C_b(S_i)$ . Hence, from the very definition of  $w_{\rho_i}$ -convergence, we get that  $\Lambda_i$  is a  $w_{\rho_i}$ -closed set for  $i = 1, 2$ . It is easy to check that  $\Lambda_i$  is convex for  $i = 1, 2$ . To check that the condition (2.1.13) is

attained, fix  $s_1 \in S_1$  and  $\eta_1 \in \mathcal{M}_1(s_1)$ . Then

$$\begin{aligned} \int_{S_1 \times S_1} yf(x)(\delta_{s_1} \times \eta_1(dx, dy)) &= \left( \int_{S_1} f(x)\delta_{s_1}(dx) \right) \left( \int_{S_1} y\eta_1(dy) \right) \\ &= f(s_1)s_1 \\ &\quad (\text{Because } \eta_1 \in \mathcal{M}_1(s_1) \text{ and } f \geq 0) \\ &= \int_{S_1 \times S_1} xf(x)(\delta_{s_1} \times \eta_1(dx, dy)), \end{aligned}$$

so  $\delta_{s_1} \times \eta_1 \in \Lambda_1$ . Now, fix  $s_2 \in S_2$  and  $\eta_2 \in \mathcal{M}_2(s_2)$ . Then

$$\begin{aligned} \int_{S_2 \times S_2} yf(x)(\delta_{s_2} \times \eta_2(dx, dy)) &= \left( \int_{S_2} f(x)\delta_{s_2}(dx) \right) \left( \int_{S_2} y\eta_2(dy) \right) \\ &\succeq f(s_2)s_2 \\ &\quad (\text{Because } \eta_2 \in \mathcal{M}_2(s_2)) \\ &= \int_{S_2 \times S_2} xf(x)(\delta_{s_2} \times \eta_2(dx, dy)), \end{aligned}$$

so  $\delta_{s_2} \times \eta_2 \in \Lambda_2$ . Hence, the condition (2.1.13) is valid for  $i = 1, 2$ . Finally, let  $a \in [0, 1]$  and notice that  $a\mathcal{M}_1(u) + (1-a)\mathcal{M}_1(v) \subset \mathcal{M}_1(s)$  iff  $au + (1-a)v = s$ . It is also true that  $au + (1-a)v = s$  implies  $a\mathcal{M}_2(u) + (1-a)\mathcal{M}_2(v) \subset \mathcal{M}_2(s)$  (although the implication in the opposite direction may not hold); in either case,  $\mathcal{M}_i$ -convexity implies ordinary convexity, for  $i = 1, 2$ .  $\square$

For each triplet  $(S_i, \{\mathcal{M}_i(s)\}_{s \in S}, \Lambda_i)$ ,  $g_*^i : S \rightarrow \mathbb{R}$  is defined by

$$g_*^i(s) := \inf_{\eta \in \mathcal{M}_i(s)} \int g d\eta, \quad \text{for } i = 1, 2, \quad (2.2.1)$$

just as in (2.1.14).

**Theorem 2.2.2.** *For  $i = 1, 2$ , if  $g \in C_{\sigma_i}(S_i)$  and  $0 \leq g \leq \sigma_i$ , then  $g_*^i$  is a continuous convex function such that  $0 \leq g_*^i \leq \sigma_i$ . Hence,  $g_*^i$  is measurable and  $\mu$ -integrable for every  $\mu \in Pr_{\sigma_i}(S_i)$ . Moreover,  $g_*^2$  is increasing (with respect to the order  $\preceq$  in  $S_2$ ).*

*Proof.* Fix  $i = 1, 2$ . Let  $g$  be continuous with  $0 \leq g \leq \sigma_i$ ; then by 1 and 2 of Lemma 2.1.2,  $0 \leq g_*^i \leq \sigma_i$  and  $g_*^i$  is  $\mathcal{M}_i$ -convex. By Theorem 2.2.1, this implies that  $g_*^i$  is a convex function. Now suppose  $g_*^i(s_0) < a$  for some

$s_0 \in S_i$  and  $a \in \mathbb{R}$ ; then there exists  $\eta \in \mathcal{M}_i(s_0)$  so that  $\int_{S_i} g d\eta < a$ . Define  $\eta_x(A) = \eta(A - x)$ ; then the function  $p : S_i \rightarrow \mathbb{R}$  defined by

$$p(x) = \int_{S_i} g d\eta_x = \int_{S_i} g(y + x)\eta(dy)$$

is effectively finite for each  $x$ ,<sup>1</sup> continuous<sup>2</sup> and  $p(0) = \int g d\eta < a$ . Hence, there exists  $\delta > 0$  so that  $p(x) < a$  for  $\|x\|_i < \delta$ . But  $\eta_x \in \mathcal{M}_i(s_0 + x)$  and so  $g_*^i(s_0 + x) < a$  for  $\|x\|_i < \delta$ . This coincides with the definition of upper semicontinuity in the Definition 1.2.1; just replace  $a$  with  $g_*^i(s_0) + \epsilon$ . Then  $g_*^i$  us an upper semicontinuous convex function on  $S_i$ , and applying Corollary 1.2.1 we conclude that it is actually a continuous convex function on  $S_i$ .

To see that  $g_*^2$  is increasing, just note that  $\{\mathcal{M}_2(s)\}_{s \in S_2}$  is a decreasing collection of sets (with respect the order  $\preceq$  in  $S_2$ ), so by the definition in (2.2.1),  $g_*^2$  must be increasing.  $\square$

**Lemma 2.2.1.** *Let  $\mu, \nu \in Pr_{\sigma_1}(S_1)$  ( $\mu, \nu \in Pr_{\sigma_2}(S_3)$ ) and assume that*

$$\int h d\mu \leq \int h d\nu \tag{2.2.2}$$

*for all  $h$  continuous and convex with  $0 \leq h \leq \sigma_1$  (for all  $h$  continuous, convex and increasing with  $0 \leq h \leq \sigma_2$ ); then for all  $g \in C_{\sigma_1}(S_1)$  ( $g \in C_{\sigma_2}(S_2)$ ) we have that*

$$\int g_*^1 d\mu \leq \int g_*^1 d\nu \quad \left( \int g_*^2 d\mu \leq \int g_*^2 d\nu \right). \tag{2.2.3}$$

*Proof.* (We are giving the full proof when  $g \in C_{\sigma_1}(S_1)$ ; for  $g \in C_{\sigma_2}(S_2)$  the reasoning is completely analogous). If  $g \in C_{\sigma_1}(S_1)$  is such that  $0 \leq g \leq \sigma_1$ , then  $g_*^1$  is continuous, convex and  $0 \leq g_*^1 \leq \sigma_1$  (by Theorem 2.2.2), so (2.2.3) is valid due to the hypothesis (2.2.2).

Now, take  $g \in C_{\sigma_1}(S_1)$  which is bounded from below; then there exist  $a, b > 0$  such that  $0 \leq ag + b \leq \sigma_1$  and since  $ag + b$  is convex and continuous, then the preceding case gives us that

$$\int_{S_1} (ag + b)_*^1 d\mu \leq \int_{S_1} (ag + b)_*^1 d\nu,$$

<sup>1</sup>Since  $|p(x)| \leq \int_{S_i} |g(y + x)|\eta(dy) \leq \int_{S_i} \sigma_i(y + x)\eta(dy) \leq \int_{S_i} \sigma_i(y)\eta(dy) + \|x\| < \infty$ .

<sup>2</sup>By the Dominated Convergence Theorem for Bochner Integrals; see Theorem 11.46 (pp. 427) of Aliprantis et al. [2].

and since  $(ag + b)_*^1 = ag_*^1 + b$ , then

$$\int_{S_1} g_*^1 d\mu \leq \int_{S_1} g_*^1 d\nu$$

so (2.2.3) is also valid in this case.

Finally, take any  $g \in C_{\sigma_1}(S_1)$ . Define

$$g_n = \max(g, -n), n \in \mathbb{N}.$$

Clearly each  $g_n$  is bounded from below and  $g_n \in C_{\sigma_1}(S_1)$ , so the previous case yields that

$$\int_{S_1} (g_n)_*^1 d\mu \leq \int_{S_1} (g_n)_*^1 d\nu.$$

Furthermore  $g_n \downarrow g$  for all  $n \in \mathbb{N}$ , so

$$(g_n)_*^1 \downarrow h_0 \geq g_*^1, \quad (2.2.4)$$

where  $h_0$  is some measurable function on  $S_1$ . Fix  $s \in S_1$ . If  $a > g_*^1(s)$ , then by definition of  $g_*^1$  there exists  $\eta_0 \in \mathcal{M}_1(s)$  such that  $\int_{S_1} g d\eta_0 < a$ . Since  $g_n \downarrow g$ , then by the MCT there is  $n_0 \in \mathbb{N}$  such that

$$\int_{S_1} g_n d\eta_0 < a \text{ for all } n \geq n_0;$$

hence

$$(g_n)_*^1(s) = \inf_{\eta \in \mathcal{M}_1(s)} \int_{S_1} g_n d\eta < a \text{ for all } n \geq n_0.$$

This means that  $h_0 = \lim_{n \rightarrow \infty} (g_n)_*^1 \leq g_*^1$ : combining this with (2.2.4), we get that  $h_0 = g_*^1$ . Notice that  $g_0 \in L_1(\mu) \cap L_1(\nu)$ , so by the MCT

$$\int_{S_1} g d\mu = \lim_{n \rightarrow \infty} \int_{S_1} (g_n)_*^1 d\mu \leq \lim_{n \rightarrow \infty} \int_{S_1} (g_n)_*^1 d\nu = \int_{S_1} g d\nu,$$

so (2.2.3) is true for all  $g \in C_{\sigma_1}(S_1)$ . □

**Lemma 2.2.2.** *Let  $\mu, \nu \in Pr_{\sigma_1}(S_1)$ ,  $(\mu, \nu \in Pr_{\sigma_2}(S_2))$  and assume that*

$$\int h d\mu \leq \int h d\nu$$

*for all  $h$  continuous and convex with  $0 \leq h \leq \sigma_1$  (for all  $h$  continuous, convex and increasing with  $0 \leq h \leq \sigma_2$ ). Then there exists a measure  $\gamma \in \Lambda_1$  ( $\gamma \in \Lambda_2$ ) with marginals  $\mu$  and  $\nu$ .*

*Proof.* For  $i = 1, 2$ , Let  $f, g \in C_{\sigma_i}(S_i)$ . Notice that

$$\begin{aligned}
\inf_{s \in S_i} \{f(s) + g_*^1(s)\} &\leq \int_{S_i} f(s) + g_*^1(s) \mu(ds) \\
&\quad \text{(Since } \mu \text{ and } \nu \text{ are probability measures)} \\
&= \int_{S_i} f d\mu + \int_{S_i} g_*^1 d\mu \\
&\leq \int_{S_i} f d\mu + \int_{S_i} g_*^1 d\nu \quad \text{(By Lemma 2.2.1)} \\
&\leq \int_{S_i} f d\mu + \int_{S_i} g d\nu \\
&\quad \text{(Since } g_*^i \leq g; \text{ see point 2 of Lemma 2.1.2);}
\end{aligned}$$

since this is true for all  $f, g \in C_{\sigma_i}(S_i)$ , the result follows from Corollary 2.1.1.  $\square$

**Theorem 2.2.3.** (*Strassen's Theorem for convex order*) Let  $\mu$  and  $\nu$  be Radon probability measures on the Banach space  $(S_1, \|\cdot\|_1)$  (ordered Banach space  $(S_2, \|\cdot\|_2, \preceq)$ ). Assume that

1.  $\int_{S_1} \|x\|_1 \mu(dx) < \infty, \int_{S_1} \|x\|_1 \nu(dx) < \infty$   
 $\left( \int_{S_2} \|x\|_2 \mu(dx) < \infty, \int_{S_2} \|x\|_2 \nu(dx) < \infty \right),$
2.  $\int h d\mu \leq \int h d\nu$  for all  $h$  continuous and convex with  $0 \leq h \leq 1 + \|\cdot\|_1$   
(for all  $h$  continuous, convex and increasing with  $0 \leq h \leq 1 + \|\cdot\|_2$ ).

Then there exists a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and random variables  $X$  and  $Y$  such that

1.  $X \sim \mu$  and  $Y \sim \nu$ ,
2.  $X = \mathbb{E}(Y|X)$  ( $X \preceq \mathbb{E}(Y|X)$ ).

*Proof.* For  $i = 1, 2$ , apply Lemma 2.2.2 to the measures  $\mu, \nu \in Pr_{\sigma_i}(S_i)$ , to obtain  $\gamma \in \Lambda_i$ . Then take  $\Omega = S_i \times S_i, \mathcal{F} = \mathbb{B}(S_1 \times S_1), \mathbb{P} = \gamma$ . Let  $X : S_i \times S_i \rightarrow S_i$  be the projection of the first coordinate and let  $Y : S_i \times S_i \rightarrow S_i$  be the projection of the second coordinate.  $\square$

In the case when  $S_i = \mathbb{R}$ , then Theorem 2.2.3 takes the next easier form.

**Theorem 2.2.4.** Let  $\mu$  and  $\nu$  be integrable probability measures on  $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ .<sup>3</sup> Then the condition

$$\int_{\mathbb{R}} f d\mu \leq \int_{\mathbb{R}} f d\nu \quad \forall \text{ convex } f : \mathbb{R} \rightarrow \mathbb{R} \quad (2.2.5)$$

is equivalent to the existence of a probability space with measure  $\mathbb{P}$  and random variables  $X \sim \mu$  and  $Y \sim \nu$  defined on  $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$  such that

$$X = \mathbb{E}(Y|X) \text{ under } \mathbb{P}.$$

*Proof.* Suppose that  $X = \mathbb{E}(Y|X)$  is valid. Then, Jensen's inequality for conditional expectation implies that that

$$f(X) = f(\mathbb{E}(Y|X)) \leq \mathbb{E}(f(Y)|X);$$

computing the expected value in the previous inequality implies (2.2.5) so necessity is proved. Theorem 2.2.3 implies sufficiency.  $\square$

## 2.3 A review of convex order in $\mathbb{R}$

### 2.3.1 An actuarial view.

Suppose that a person suffers some accident which results in a (random) total loss of  $X > 0$  units of money. If said damage was insured, then in an “ideal” and easy setting the insurance company would pay the total loss, that is, it would pay  $X$  units of money to cover the loss. In real life, there exist more complex contracts in which the main insurance company (called the **cedant**) cedes part of its obligation (or risk) to other insurance company (called the **reinsurer**), that is, two companies would pay for the loss  $X$ : such contracts are commonly called **reinsurance agreements**. There exist several types of reinsurance agreements, however we will only be interested in the next one.

**Definition 2.3.1.** Consider a reinsurance agreement in which the cedant retains a risk of  $d > 0$  units of money and lets the reinsurer pay for the remainder, that is, if the loss  $X$  is less or equal than  $d$ , then the cedant

---

<sup>3</sup>Any probability measure over the Borel sets of any metric space is automatically a Radon measure (see Theorem 1.2 (pp. 27) in Parthasarathy [14]), so it is not necessary to specify that  $\mu$  and  $\nu$  are Radon measures.

will fully pay for that loss (and the reinsurer will not pay anything); if  $X$  is greater than  $d$ , then the cedant will pay  $d$  units of money and the reinsurer will pay  $X - d$  units of money, so that both contributions fully cover the loss. This kind of agreement is called **stop-loss reinsurance**. Consider the payment corresponding to the reinsurer for that particular claim and let  $\pi_X(d)$  be its mean; that is, let

$$\pi_X(d) := \mathbb{E}((X - d)^+) = \mathbb{E}(\max\{X - d, 0\});$$

we call  $\pi_X(d)$  the **stop-loss premium**.

From now on, let us assume that the total loss  $X$  is integrable.

**Proposition 2.3.1.** *1. If  $F_X$  is the distribution function of the random variable  $X$ , then for all  $d \in \mathbb{R}$*

$$\pi_X(d) = \int_d^\infty (1 - F_X(s))ds. \quad (2.3.1)$$

*2.  $\pi_X : \mathbb{R} \rightarrow \mathbb{R}$  is a non-increasing convex function such that  $\lim_{d \rightarrow -\infty} \pi_X(d) + d = \mathbb{E}(X)$  and  $\lim_{d \rightarrow \infty} \pi_X(d) = 0$ .*

*Proof.* 1. Just notice that

$$\begin{aligned} \int_d^\infty (1 - F_X(s))ds &= \int_{-\infty}^\infty \int_{-\infty}^\infty \mathbb{1}_{d < s < x} dF(x)ds \\ &= \int_{-\infty}^\infty \int_{-\infty}^\infty \mathbb{1}_{d < s < x} ds dF(x) \quad (\text{By Tonelli's Theorem}) \\ &= \int_d^\infty (x - d)dF(x) \\ &= \mathbb{E}((X - d)^+) = \pi_X(d). \end{aligned}$$

2. From (2.3.1) we have got that

$$\pi'_X(d) = F_X(d) - 1,$$

which is non-decreasing and non-positive, so  $\pi_X(\cdot)$  must be convex and

non-decreasing. Next, notice that for each  $d \leq 0$ ,

$$\begin{aligned}
\mathbb{E}((X - d)^+) + d &= \int_d^\infty (1 - F_X(s))ds - \int_d^0 ds \\
&= \int_0^\infty (1 - F_X(s))ds - \int_d^0 F_X(s)ds \\
&\xrightarrow{d \rightarrow -\infty} \mathbb{E}(X^+) - \int_{-\infty}^0 F_X(s)ds \\
&= \mathbb{E}(X^+) - \mathbb{E}(X^-) = \mathbb{E}(X).
\end{aligned}$$

Finally,

$$\begin{aligned}
\lim_{d \rightarrow \infty} \pi_X(d) &= \lim_{d \rightarrow \infty} \mathbb{E}((X - d)^+) \\
&= \mathbb{E}(\lim_{d \rightarrow \infty} (X - d)^+) \quad (\text{By DCT, since } \mathbb{E}(X) < \infty) \\
&= \mathbb{E}((-\infty)^+) = 0.
\end{aligned}$$

□

Moreover, we have the next result:

**Proposition 2.3.2.** *If  $\theta : \mathbb{R} \rightarrow \mathbb{R}$  is non-increasing, convex,  $\lim_{d \rightarrow \infty} \theta(d) = 0$  and  $\lim_{d \rightarrow -\infty} \pi_X(d) + d$  exists and is finite, then  $\theta(\cdot)$  is the stop-loss premium of some unique integrable distribution whose mean is  $\lim_{d \rightarrow -\infty} \pi_X(d) + d$ : we will call such a function  $\theta$  a **stop-loss function**.*

*Proof.* It is straightforward to prove that  $1 + \partial_+ \theta^4$  is the only distribution function that accepts  $\theta(\cdot)$  as its associated stop-loss premium. □

It can be shown that the stop-loss reinsurance scheme is the least variable agreement possible, from the point of view of the cedant (see Theorem 1.4.3 in Kaas et al. [9]). Nevertheless, we will be more interested in see what happens with this type of reinsurance from the point of view of the reinsurer:

Suppose we are part of an insurance company which has the option to choose between being a stop-loss reinsurer for the cedant  $A$  or being a stop-loss reinsurer for the cedant  $B$  (assume there is no pay for us). Suppose that the claims that  $A$  receives have the same distribution as some random

---

<sup>4</sup>In this context,  $\partial_+$  denotes the right derivative of  $\theta(\cdot)$ .

variable  $X$  and the claims that  $B$  receives have the same distribution as some random variable  $Y$ . Also, suppose that  $\mathbb{E}(X) = \mathbb{E}(Y)$  but

$$\mathbb{E}((X - d)^+) \leq \mathbb{E}((Y - d)^+) \quad \text{for all } d \in \mathbb{R}. \quad (2.3.2)$$

Which would be the best option to choose from,  $A$  or  $B$ ? At first we would think that since  $\mathbb{E}(X) = \mathbb{E}(Y)$ , then it really does not matter which option we choose; nevertheless, (2.3.2) tells us that the most desirable one is option  $A$ , since we are expected to pay less. But what else can we say about  $X$  and  $Y$ ? Notice that for fixed  $d$ , we (as the reinsurers) ignore claims which are less than  $d$  units of money: that is, if (2.3.2) is true then  $Y$  must have a higher probability than  $X$  of taking “extremely” larger values than  $d$ , but since  $\mathbb{E}(X) = \mathbb{E}(Y)$ , then  $Y$  must also have a higher probability than  $X$  of taking “extremely” small values (since we must compensate those larger values of  $Y$  with smaller ones to guarantee that the expected value is the same). This gives us the intuition that  $Y$  is more “variable” than  $X$  and suggests the following definition.

**Definition 2.3.2.** Let  $X$  and  $Y$  be positive and integrable random variables. We write  $X \leq_{SL} Y$  whenever  $\mathbb{E}((X - d)^+) \leq \mathbb{E}((Y - d)^+)$  for all  $d \in \mathbb{R}$ : we call the relation  $\leq_{SL}$  the **stop-loss order**.

**Remark 2.3.1.** The relation  $\leq_{SL}$  is not an order in the mathematical sense of the word: it is actually a partial order, since not every pair of random variables can be related by  $\leq_{SL}$ , even if their expected value is the same.

Now that we have given a constructive reasoning behind the stop-loss order, let us forget about the insurance setting and move on to “variability” arguments. We have seen that (2.3.2) hints that  $Y$  is more variable than  $X$ . Suppose that also

$$\mathbb{E}((d - X)^+) \leq \mathbb{E}((d - Y)^+) \quad \text{for all } d \in \mathbb{R}. \quad (2.3.3)$$

For fixed  $d$ , this means that  $Y$  takes values smaller than  $d$  in a more “extreme” way than  $X$ ; this is consistent with the explanation of (2.3.2), so both (2.3.2) and (2.3.3) hint that  $X$  is more “constant” than  $Y$ . Let us study both properties and let us lift the condition of  $X$  and  $Y$  being positive random variables.

**Proposition 2.3.3.** *Let  $X$  an integrable random variable with distribution function  $F_X$ . Then*

$$\mathbb{E}((X - d)^+) = \int_d^{+\infty} (1 - F_X(s))ds, \text{ and } \mathbb{E}((d - X)^+) = \int_{-\infty}^d F_X(s)ds.$$

*Proof.* It is analogue to the proof of Proposition 2.3.2. □

**Theorem 2.3.1.** *Let  $X$  and  $Y$  be two integrable random variables. Then the next statements are equivalent*

1.  $\mathbb{E}(X) = \mathbb{E}(Y)$  and

$$\mathbb{E}((X - d)^+) \leq \mathbb{E}((Y - d)^+) \text{ for all } d \in \mathbb{R}. \quad (2.3.4)$$

2.  $\mathbb{E}(X) = \mathbb{E}(Y)$ ,

$$\mathbb{E}((X-d)^+) \leq \mathbb{E}((Y-d)^+) \text{ and } \mathbb{E}((d-X)^+) \leq \mathbb{E}((d-Y)^+) \text{ for all } d \in \mathbb{R}. \quad (2.3.5)$$

3.  $\mathbb{E}(f(X)) \leq \mathbb{E}(f(Y))$  for all convex function  $f : \mathbb{R} \rightarrow \mathbb{R}$  in the case both expected values are finite.

*If either of them happens, we write  $X \leq_{cx} Y$  and we call the relation  $\leq_{cx}$  **convex order**.*

*Proof.* (1.  $\Rightarrow$  2.) Fix  $d \in \mathbb{R}$ . Since  $\mathbb{E}(X) = \mathbb{E}(Y)$  and  $\mathbb{E}((X - d)^+) \leq \mathbb{E}((Y - d)^+)$ , then

$$\mathbb{E}((d-X)^+) = \mathbb{E}((X-d)^+) - \mathbb{E}(X-d) \leq \mathbb{E}((Y-d)^+) - \mathbb{E}(Y-d) = \mathbb{E}((d-Y)^+),$$

so the result follows.

(2.  $\Rightarrow$  3.) Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a convex function such that  $\mathbb{E}(f(X))$  and  $\mathbb{E}(f(Y))$  are finite. It is known that convex functions are differentiable at all but countably many points: let  $a \in \mathbb{R}$  be such that  $f$  is differentiable in  $a$ . Define  $g(x) = f(x) - g(a) - (x - a)f'(a)$ . Since  $\mathbb{E}(X) = \mathbb{E}(Y)$ , the inequality  $\mathbb{E}(f(X)) \leq \mathbb{E}(f(Y)) < \infty$  is equivalent to the inequality

$\mathbb{E}(g(X)) \leq \mathbb{E}(g(Y)) < \infty$ . Also, notice that  $g(a) = g'(a) = 0$ . Let  $F_X$  be the distribution function of  $X$ . Then we have that

$$\begin{aligned} \mathbb{E}(g(X)) &= \int_{-\infty}^a g(x)F(dx) - \int_a^{+\infty} g(x)\overline{F}(dx) \\ &= - \int_{-\infty}^a g'(x)F(x)dx + \int_a^{+\infty} g'(x)\overline{F}(x)dx \end{aligned} \quad (2.3.6)$$

(By integration by parts)

$$= \int_{-\infty}^a \left( \int_{-\infty}^x F_X(s)ds \right) dg'(x) + \int_a^{+\infty} \left( \int_x^{+\infty} \overline{F}_X(s)ds \right) dg'(x)$$

(By integration by parts)

$$= \int_{-\infty}^a \mathbb{E}((x - X)^+)dg'(x) + \int_a^{+\infty} \mathbb{E}((X - x)^+)dg'(x) \quad (2.3.7)$$

(By Proposition 2.3.3).

Before continuing with the proof we must clarify a few details of the previous calculations: since  $f$  is convex it can be checked that  $g$  is also convex and thus differentiable at all but countably many points, say  $C \in \mathbb{R}$ . This means that  $g'$  is defined in  $\mathbb{R} \setminus C$ ; hence we can extend  $g'$  by defining  $g'(x) = \lim_{s \downarrow x} g'(s)$  for all  $x \in \mathbb{R}$  (where the limit is taken over  $s \in \mathbb{R} \setminus C$ ). This extended  $g'$  is right-continuous and we can see that this extension does no harm since the first time we introduced  $g'$  was in (2.3.6) where an integral with respect to the Lebesgue measure was being computed. Now, since  $g$  is convex then  $g'$  is non-decreasing, so the integrals at (2.3.7) are well-defined as Lebesgue-Stieltjes integrals. Moreover, this means that  $dg' \geq 0$ , so the result follows from (2.3.5).

(3.  $\Rightarrow$  1.) Both  $f_1(x) = x$  and  $f_2 = -x$  are convex functions, hence  $\mathbb{E}(X) \leq \mathbb{E}(Y)$  and  $\mathbb{E}(-X) \leq \mathbb{E}(-Y)$ , so  $\mathbb{E}(X) = \mathbb{E}(Y)$ . (2.3.4) follows because  $f_3(x) = (x - d)^+$  is convex for all  $d \in \mathbb{R}$ .  $\square$

**Remark 2.3.2.** As in Definition 2.3.2, the relation  $\leq_{cx}$  is not an order since not every pair of random variables can be compared, but it straightforward to prove that it is a partial order.

**Proposition 2.3.4.** *If  $X \leq_{cx} Y$  and  $X, Y$  are square integrable, then*

$$\text{Var}(X) \leq \text{Var}(Y) \quad (2.3.8)$$

*Proof.* Let  $a = \mathbb{E}(X) = \mathbb{E}(Y)$ . Since the real function  $f$  defined by  $f(x) = (x - a)^2$  is convex, then (2.3.8) follows from the definition of convex order and variance.  $\square$

The previous result ultimately proves that convex order effectively is a measure for variability, and is actually stronger than the order induced by the variance.

### 2.3.2 A mathematical view.

We have defined what does it mean for a pair of random variables to be convex ordered. Nevertheless, one can see that in Theorem 2.3.1 we were actually working with the distribution functions of that pair rather than with the random variables themselves. This motivates the next definition.

**Definition 2.3.3.** Let  $\mu$  and  $\nu$  be probability measures over  $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ . Then we say that  $\nu$  **majorizes**  $\mu$  (and we denote it by  $\mu \prec \nu$ ) if

$$\int_{\mathbb{R}} f d\mu \leq \int_{\mathbb{R}} f d\nu \quad (2.3.9)$$

for all real-valued convex function  $f$ .

Although it is a quite easy definition, most of the computations become very difficult. Nevertheless, we have developed a powerful tool in Subsection 2.2 which makes computations a lot easier and compact.

**Theorem 2.3.2.** *Let  $\mu$  and  $\nu$  be integrable probability measures over  $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ . Then  $\mu \prec \nu$  iff there exists a probability space and real random variables  $X \sim \mu$  and  $Y \sim \nu$  such that  $X = \mathbb{E}(Y|X)$  in that space.*

This result, which is a particular case of Theorem 2.2.4, is called the **Strassen's theorem for majorization**. From now on we will only work with integrable measures over  $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ , so in the remaining proofs of this chapter we will use the equivalence in Theorem 2.3.2 as an alternative definition of majorization.

**Proposition 2.3.5.** *Let  $\{Z_i\}_{i=0}^n$  be any collection of integrable random variables. If  $\mathbb{E}(Z_0|Z_1, \dots, Z_n) = g(Z_1, \dots, Z_n)$  for some measurable  $g : \mathbb{R}^n \rightarrow \mathbb{R}$ , then  $\mathbb{E}(Z_0|g(Z_1, \dots, Z_n)) = g(Z_1, \dots, Z_n)$ .*

*Proof.* Just apply the tower property of conditional expectation.  $\square$

**Lemma 2.3.1.** *Let  $X_1, X_2, \dots, X_n$  be i.i.d. with finite mean,  $a_1, \dots, a_n$  are real constants and define  $\bar{a} = \sum_{i=1}^n a_i/n$ . Then*

$$\mathcal{L}\left(\bar{a} \sum_{i=1}^n X_i\right) \prec \mathcal{L}\left(\sum_{i=1}^n a_i X_i\right).$$

*Proof.* Consider a random permutation, say  $\sigma$ , of  $1, 2, \dots, n$ , which is distributed uniformly on the set of all permutations, and is also independent from  $\{X_i\}_{i=1}^n$ . Then

$$\mathbb{E}\left(\sum_{i=1}^n a_{\sigma(i)} X_i \mid X_1, \dots, X_n\right) = \sum_{i=1}^n \mathbb{E}(a_{\sigma(i)}) X_i = \bar{a} \sum_{i=1}^n X_i,$$

so by Proposition 2.3.5 we have

$$\mathbb{E}\left(\sum_{i=1}^n a_{\sigma(i)} X_i \mid \bar{a} \sum_{i=1}^n X_i\right) = \bar{a} \sum_{i=1}^n X_i. \quad (2.3.10)$$

Since the collection  $X_1, \dots, X_n$  is exchangeable, then

$$\mathcal{L}\left(\sum_{i=1}^n a_{\sigma(i)} X_i\right) = \mathcal{L}\left(\sum_{i=1}^n a_{\sigma(i)} X_i\right)$$

so the result follows from (2.3.10).  $\square$

**Lemma 2.3.2.** *Let  $X$  and  $Y$  be independent random variables with finite means, and let  $a = \mathbb{E}(Y)$ . Then  $\mathcal{L}(aX) \prec \mathcal{L}(XY)$ .*

*Proof.*  $\mathbb{E}(YX|X) = \mathbb{E}(Y|X)X = aX$ ; using Proposition 2.3.5 we have that  $\mathbb{E}(YX|aX) = aX$  and the proof is complete.  $\square$

**Lemma 2.3.3.** *For  $i = 1, 2$ , suppose that  $\mu_i \prec \nu_i$  where  $\mu_i$  and  $\nu_i$  are integrable probability measures, and suppose that  $p \in [0, 1]$ . Then we have*

1.  $\mu_1 * \mu_2 \prec \nu_1 * \nu_2$  and,
2.  $p\mu_1 + (1-p)\mu_2 \prec p\nu_1 + (1-p)\nu_2$ .

*Proof.* Direct from definition.  $\square$

**Theorem 2.3.3.** Let  $\{X_i\}_{i=1}^\infty, \{Y_i\}_{i=1}^\infty, \{N_i\}_{i=1}^2$  be three independent i.i.d. sequences of non-negative random variables with finite means, with  $N_1$  and  $N_2$  being integer valued. Then

$$\mathcal{L}\left(\sum_{i=1}^{N_1} X_i + \sum_{i=1}^{N_2} Y_i\right) \prec \mathcal{L}\left(\sum_{i=1}^{N_1} (X_i + Y_i)\right). \quad (2.3.11)$$

*Proof.* According to Theorem 2.2.4, it is enough to construct a probability space on which there is defined a martingale  $(W_1, W_2)$  such that the marginal distribution of  $W_1$  is the l.h.s. of (2.3.11) and the marginal distribution of  $W_2$  is the r.h.s. of (2.3.11).

We are going to work on a fixed probability space that supports the random variables  $X_1, \dots, Y_1, \dots, N_1, N_2$  (described as in the theorem) plus an uniformly distributed random variable on  $[0, 1]$ , say  $U$ , which is independent from all the previous random variables. Now, define  $W_1$  by

$$W_1 = \sum_{i=1}^{N_1} X_i + \sum_{i=1}^{N_2} Y_i;$$

it clearly has the desired distribution. Next, define a new random variable  $N$  as follows. Let

$$C = \frac{\sum_{L+1}^M X_i}{\sum_{L+1}^M (X_i + Y_i)}, \quad (2.3.12)$$

where  $L = \min(N_1, N_2)$  and  $M = \max(N_1, N_2)$ . If the denominator of (2.3.12) is ever 0, we arbitrarily define  $C$  to be  $1/2$ . Now, define

$$N = \begin{cases} N_1 & \text{if } N_1 = N_2 & \text{or,} \\ & N_1 > N_2, \quad U \leq C & \text{or,} \\ & N_1 < N_2, \quad U > 1 - C; \\ N_2 & \text{if } N_1 > N_2, \quad U > C & \text{or,} \\ & N_1 < N_2, \quad U \leq 1 - C. \end{cases}$$

Define  $W_2$  by

$$W_2 = \sum_{i=1}^N (X_i + Y_i).$$

We must prove

1.  $\mathbb{E}(W_2|W_1) = W_1$ ,
2.  $N$  is independent of  $\{X_i\}_{i=1}^\infty, \{Y_i\}_{i=1}^\infty$  and has the same distribution as  $N_1$ .

Let  $\mathcal{F} := \sigma(\{X_i\}_{i=1}^\infty, \{Y_i\}_{i=1}^\infty)$ . According to the definition of  $N$ , we have:

For  $n_1 = n_2$ ,

$$\mathbb{E}(W_2|\mathcal{F}, N_1 = n_1, N_2 = n_2) = \sum_{i=1}^{n_1} (X_i + Y_i).$$

For  $n_1 > n_2$ ,

$$\begin{aligned} \mathbb{E}(W_2|\mathcal{F}, N_1 = n_1, N_2 = n_2) &= C \sum_{i=1}^{n_1} (X_i + Y_i) + (1 - C) \sum_{i=1}^{n_2} (X_i + Y_i) \\ &= \left( \frac{\sum_{n_2+1}^{n_1} X_i \sum_{i=1}^{n_1} (X_i + Y_i) + \sum_{n_2+1}^{n_1} Y_i \sum_{i=1}^{n_2} (X_i + Y_i)}{\sum_{n_2+1}^{n_1} (X_i + Y_i)} \right) \mathbb{1}_{\{\sum_{n_2+1}^{n_1} (X_i + Y_i) > 0\}} \end{aligned} \quad (2.3.13)$$

$$+ \left( \frac{1}{2} \sum_{i=1}^{n_1} (X_i + Y_i) + \frac{1}{2} \sum_{i=1}^{n_2} (X_i + Y_i) \right) \mathbb{1}_{\{\sum_{n_2+1}^{n_1} (X_i + Y_i) = 0\}} \quad (2.3.14)$$

$$= \sum_{i=1}^{n_1} X_i + \sum_{i=1}^{n_2} Y_i,$$

where the last equality was obtained by making

$$\sum_{i=1}^{n_1} (X_i + Y_i) = \sum_{i=1}^{n_2} (X_i + Y_i) + \sum_{i=n_2+1}^{n_1} (X_i + Y_i)$$

in (2.3.13) and by noting that

$$\sum_{i=n_2+1}^{n_1} X_i = \sum_{i=n_2+1}^{n_1} Y_i = 0$$

in (2.3.14).

For  $n_1 < n_2$ , applying similar procedures as before, we have that

$$\begin{aligned}\mathbb{E}(W_2|\mathcal{F}, N_1 = n_1, N_2 = n_2) &= (1 - C) \sum_{i=1}^{n_2} (X_i + Y_i) + C \sum_{i=1}^{n_1} (X_i + Y_i) \\ &= \sum_{i=1}^{n_1} X_i + \sum_{i=1}^{n_2} Y_i.\end{aligned}$$

Hence we get that

$$\mathbb{E}(W_2|\mathcal{F}, N_1, N_2) = \sum_{i=1}^{N_1} X_i + \sum_{i=1}^{N_2} Y_i = W_1,$$

so point 1 follows by applying Proposition 2.3.5.

To prove point 2, define  $\pi_i = \mathbb{P}(N_1 = i)$  for  $i = 0, 1, 2, \dots$  and let

$$C_{j,k} := \begin{cases} \frac{\sum_{j \vee k}^{j \wedge k} X_i}{\sum_{j \wedge k}^{j \vee k} X_i + Y_i} & \text{if } \sum_{j \vee k}^{j \wedge k} X_i + Y_i > 0 \\ 1/2 & \text{if } \sum_{j \wedge k}^{j \vee k} X_i + Y_i = 0 \end{cases};$$

notice that  $C_{j,k} = C_{k,j}$ . Then

$$\begin{aligned}\mathbb{P}(N = n|\mathcal{F}) &= \mathbb{P}(N = n|\mathcal{F}, N_1 = N_2 = n)\pi_n^2 \\ &\quad + \sum_{a=0}^{n-1} [\mathbb{P}(N = n|\mathcal{F}, N_1 = n, N_2 = a) + \mathbb{P}(N = n|\mathcal{F}, N_1 = n, N_2 = a)]\pi_n\pi_a \\ &\quad + \sum_{b=n+1}^{\infty} [\mathbb{P}(N = n|\mathcal{F}, N_1 = b, N_2 = n) + \mathbb{P}(N = n|\mathcal{F}, N_1 = n, N_2 = b)]\pi_b\pi_n \\ &= \pi_n^2 + \sum_{a=0}^{n-1} [C_{n,a} + (1 - C_{a,n})]\pi_n\pi_a + \sum_{b=n+1}^{\infty} [(1 - C_{b,n}) + C_{n,b}]\pi_b\pi_n \\ &= \pi_n \sum_{l=0}^{\infty} \pi_l = \pi_n;\end{aligned}$$

this proves point 2 and thus the proof is complete.  $\square$

### On a short proof of Strassen's theorem in $\mathbb{R}$ .

With the aid of the results of stop-loss premiums for distributions over  $\mathbb{R}$  given in the current section, we can provide a much simpler proof of Strassen's theorem for majorization over  $\mathbb{R}$ : said proof is due to Müller et. al. [11].

**Lemma 2.3.4.** *Let  $\theta_1$  be a stop-loss function and let  $l : \mathbb{R} \rightarrow \mathbb{R}$  be some affine function with  $l(x) = ax + b$  for some  $a \in (-1, 0)$  and some  $b \in \mathbb{R}$ . Define  $\theta_2(x) = \max\{\theta_1(x), l(x)\}$ . Then there exists a probability space with random variables  $X_1$  and  $X_2$  where  $\pi_{X_1} = \theta_1$ ,  $\pi_{X_2} = \theta_2$  and  $\mathbb{E}(X_2|X_1) = X_1$ .*

*Proof.* If  $l(x) \leq \theta_1(x)$  for all  $x \in \mathbb{R}$ , then  $\theta_1 = \theta_2$  and hence the result is trivial. So let us assume that this is not the case: since  $\theta_1$  is a stop-loss function then there must exist two points in  $\mathbb{R}$ , say  $x_1$  and  $x_2$ , where  $l$  and  $\theta_1$  coincide.<sup>5</sup> Notice that the maximum of two non-increasing convex functions is convex and non-increasing. Furthermore, notice that

$$\lim_{d \rightarrow -\infty} \theta_2(d) + d = \lim_{d \rightarrow -\infty} \theta_1(d) + d$$

exists and is finite, and

$$\lim_{d \rightarrow \infty} \max\{\theta_1(d), l(d)\} = \max\{\lim_{d \rightarrow \infty} \theta_1(d), \lim_{d \rightarrow \infty} l(d)\} = \max\{0, -\infty\} = 0.$$

Thus, by Proposition 2.3.2,  $\theta_2$  must be a stop-loss function. Let  $F_1$  and  $F_2$  be the respective associated distributions of the stop-loss functions  $\theta_1$  and  $\theta_2$ . By the definition of  $\theta_2$  and by Theorem 2.3.1, it follows that  $F_2$  majorizes  $F_1$ . Next, define the Markov kernel

$$Q(x, \cdot) = \begin{cases} \delta_x & \text{if } x \in \mathbb{R} \setminus (t_1, t_2) \\ \frac{x-t_1}{t_2-t_1} \delta_{t_2} + \frac{t_2-x}{t_2-t_1} \delta_{t_1} & \text{if } x \in (t_1, t_2) \end{cases}.$$

It is straightforward to check that  $\int yQ(x, dy) = x$ . It is also easy to verify that  $\theta_2$  is the stop-loss premium of  $\int Q(x, dy)F_1(dx)$ <sup>6</sup> and thus  $F_2(\cdot) = \int Q(x, \cdot)F_1(dx)$ . Within this context, if we consider a Markov chain  $\{X_n\}_{n=1}^2$  with Markov kernel given by  $Q$  and initial distribution  $X_1 \sim F_1$ , we will have that  $\mathbb{E}(X_2|X_1) = X_1$  and  $X_2 \sim F_2$ , proving the result.  $\square$

<sup>5</sup>Notice that this is the only way  $l$  and  $\theta_1$  can coincide since the right derivative of  $\theta_1$  is non-decreasing, at  $-\infty$  is  $-1$  and at  $+\infty$  is  $0$ . See Proposition 2.3.2.

<sup>6</sup>We only need to check that if  $A \in (t_1, t_2)$  then  $\int Q(x, A)F_1(dx) = 0$ , if  $A \in \mathbb{R} \setminus [t_1, t_2]$  then  $\int Q(x, A)F_1(dx) = F_1(A)$ . After doing this, everything will follow from the definition of  $\theta_1$  and  $\theta_2$ .

With the previous lemma in mind we are ready to provide a much shorter proof of Strassen's theorem for majorization over  $\mathbb{R}$ , which we state again for the sake of clarity.

**Theorem 2.3.4.** *Let  $\mu$  and  $\nu$  be probability measures over  $\mathbb{R}$ . Then  $\mu \prec \nu$  iff there exists a probability space with random variables  $X_1$  and  $Y$  where  $X_1 \sim \mu$ ,  $Y \sim \nu$  and  $\mathbb{E}(Y|X_1) = X_1$ .*

*Proof.* Let  $\theta_1$  and  $\theta_2$  be the stop-loss premium functions associated to  $\mu$  and  $\nu$ , respectively. Since  $\theta_2$  is a stop-loss function, it can be written as the supremum of some affine functions  $l_1, l_2, l_3, \dots$ , where  $l_i(x) = a_i x + b_i$  for  $a_i \in (-1, 0)$  and  $b_i \in \mathbb{R}$  for all  $i \in \mathbb{N}$ . Now, recursively define the functions  $\phi_1 = \theta_1$  and  $\phi_{n+1} = \max\{\phi_n, l_n\}$ . Notice that  $\{\phi_n\}_{n \geq 1}$  is a non-decreasing sequence of stop-loss functions, and that the pair  $(\phi_n, \phi_{n+1})$  satisfies the assumptions of Lemma 2.3.4 for every  $n \geq 1$ : within that context, let  $Q_n$  be the Markov kernel associated to the pair  $(\phi_n, \phi_{n+1})$ . Following the proof of Lemma 2.3.4, if we consider the non-homogeneous Markov chain  $\{X_n\}_{n \geq 1}$  with initial distribution  $X_1 \sim \mu$  and successive Markovian kernels  $\{Q_n\}_{n \geq 1}$ , we will have that  $\mathbb{E}(X_{n+1}|X_n) = X_n$ . Since

$$\begin{aligned} \mathbb{E}(|X_n|) &= 2\mathbb{E}(X_n^+) - E(X_n) \\ &= 2\phi_n - \left( \lim_{d \rightarrow -\infty} \phi_n(d) - d \right) \\ &\leq 2\theta_2 - \left( \lim_{d \rightarrow -\infty} \theta_1(d) - d \right) < \infty, \end{aligned}$$

$X_1, X_2, \dots$  is an  $L_1$ -bounded martingale, and hence it converges to a random variable  $Y$  such that  $\mathbb{E}(Y|X_n) = X_n$  for all  $n \in \mathbb{N}$ , and in particular  $\mathbb{E}(Y|X_1) = X_1$ . Moreover, it is obvious that  $\pi_Y = \theta_2$ , so the distribution of  $Y$  must be  $\nu$ .  $\square$

# Chapter 3

## Phase-type distributions theory.

In this chapter we will review some of the most used tools in Applied Probability, which will be vital in further chapters.

In Section 3.1 we state some known properties of Markov jump processes, which are defined to be continuous time Markov processes with an at most countable state space; here we will deal with the time-homogeneous and finite state space case.

In Section 3.2 we study the pillar of this manuscript: phase-type distributions. A random variable which is phase-type distributed is defined as the random time it takes to a certain Markov jump process to get absorbed in some fixed state. We will show how to manipulate this class of distributions and some important results about it.

In Section 3.3 we review some basic results for renewal processes in the case their interarrival times are phase-type distributed; these are extremely desirable since we get explicit quantities in an easier way than in the general case.

In Section 3.4 we investigate the method of uniformization for Markov jump processes, which basically is the idea of simulating trajectories of a given Markov jump process by generating a Markov chain which changes state at each arrival time of a Poisson process. Later, we investigate an ex-

tension of this method which will be useful in Section 4.1.

In Section 3.5 we will prove that Erlang distributions are the least “variable” ones within the class of phase-type distributions. This classic result was originally proved by Aldous and Shepp [1] by demonstrating that Erlang distributions of order  $n$  have the minimum coefficient of variance within the class of phase-type distributions of order  $n$ . However, we give a much stronger result based on O’Cinneide [13], which states that Erlang distributions of order  $n$  are majorized by any  $n$ -dimensional phase-type distribution of the same mean. This will be the founding argument of the erlangization method, which consists in approximating certain fixed positive value with Erlang distributed random variables: this method will be used in Chapters 5 and 6.

### 3.1 Markov jump processes.

Let  $X = \{X_t\}_{t \geq 0}$  be a Markov process which takes values in the state space  $E = \{1, 2, \dots, n\}$ .<sup>1</sup> Then  $X$  is called a **Markov jump process**: some of its properties are being presented next (see Section 2.3 in Peralta [15] for a full proof of the following statements).

Define  $T_0 = 0$  and let  $T_1, T_2, \dots$  denote the successive times at which  $X$  switches states, that is, its jump times. Then the discrete-time process  $Y = \{Y_n\}_{n \in \mathbb{N} \cup \{0\}}$  where  $Y_n = X_{T_n}$ , is a Markov chain (with some transition matrix  $\mathbf{Q} = \{q_{ij}\}_{i,j \in E}$ ) which takes note of the states that the process  $X$  sequentially visits; if  $X$  ever gets absorbed in some state, say at the  $m$ -th jump, then we may define  $Y_{m+1} = Y_{m+2} = \dots = Y_m$ . It can be shown that the conditional distribution of  $T_{n+1} - T_n$  given the event  $\{Y_n = i\}$  is exponential with a certain parameter that depends on  $i$ , say  $\lambda_i \geq 0$ .<sup>2</sup> This is equivalent to say that  $\mathbb{P}(X_{t+dt} \neq i | X_t = i) = \lambda_i dt$ . Hence, we may call  $\lambda_i$  the **intensity of jump from the state  $i$** . Now, given that there exists a jump in the interval  $[t, t + dt)$ , it will land in  $j$  with probability  $q_{ij}$ , according to the definition of  $\mathbf{Q} = \{q_{ij}\}_{i,j \in E}$ ; this means that  $\mathbb{P}(X_{t+dt} = j | X_t = i) = \lambda_i q_{ij} dt$ . Thus, for  $i \neq j$  we may define  $\lambda_{ij} := \lambda_i q_{ij}$  and call it the **intensity of jump from state  $i$  to state  $j$** . Define  $\lambda_{ii} = -\lambda_i$  and  $\mathbf{\Lambda} = \{\lambda_{ij}\}_{i,j \in E}$ . Then  $\mathbf{\Lambda}$  is called the **intensity matrix of the process  $X$** . If we let  $p_{ij}^t := \mathbb{P}(X_t = j | X_0 = i)$  and  $\mathbf{P}^t := \{p_{ij}^t\}_{i,j \in E}$ , then it can be shown that

$$\mathbf{P}^t = \exp(\mathbf{\Lambda}t),$$

where  $\exp(\mathbf{A})$  denotes the exponential of the matrix  $\mathbf{A}$ , defined by

$$\exp(\mathbf{A}) := \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!}.$$

We say that a state  $i \in E$  is **absorbing (transient)** for  $X$  iff  $i$  is absorbing (transient) for  $Y$ ; particularly  $i$  is absorbing iff  $\lambda_i = 0$  or  $q_{ij} = \delta_{ij}$ .

---

<sup>1</sup>Actually, in our setting, it needs to be homogeneous in time and standard. We also suppose that we are working with the càdlàg version of the process. See Section 2.3 in Peralta [15] for more details on this.

<sup>2</sup>When  $\lambda_i = 0$  this means that a jump from  $i$  never occurs, which is compatible with the idea of its jump time being an exponentially distributed of parameter 0.

## 3.2 Phase-type distributions.

Let  $X = \{X_t\}_{t \geq 0}$  be a Markov jump process with finite state space  $E = \{1, \dots, p, p+1\}$  such that the states  $1, \dots, p$  are transient and  $p+1$  is absorbing. Then  $X$  has an intensity matrix

$$\Lambda = \left( \begin{array}{c|c} \mathbf{T} & \mathbf{t} \\ \hline \mathbf{0} & 0 \end{array} \right), \quad (3.2.1)$$

where  $\mathbf{T} = \{t_{ij}\}_{i,j \in \{1, \dots, p\}}$  and  $\mathbf{t} = (t_1, \dots, t_p)'$ .<sup>3</sup> We call  $\mathbf{T}$  the **sub-intensity matrix** and  $\mathbf{t}$  the **vector of absorption rates (or intensities)**. Since the elements of each row of any intensity matrix must sum 0, we have that  $\mathbf{t} = -\mathbf{T}\mathbf{e}$  where  $\mathbf{e} = (1, 1, \dots, 1)'$ ,<sup>4</sup> so the matrix  $\Lambda$  is completely characterized by the matrix  $\mathbf{T}$ .

**Remark 3.2.1.** The sum of the elements of each row of  $\mathbf{T}$  is non-positive and the elements of its diagonal are also non-positive; we refer to any matrix which satisfies this properties as a sub-intensity matrix. We can associate such a matrix to a possibly **terminating Markov jump process with states  $\{1, \dots, p\}$** .<sup>5,6</sup>

Another condition we impose on  $X$  is that it cannot start in the state  $p+1$ , that is  $\mathbb{P}(X_0 = p+1) = 0$ .<sup>7</sup> Let  $\pi_i := \mathbb{P}(X_0 = i)$ ,  $i = 1, \dots, p$  and let  $\boldsymbol{\pi} := (\pi_1, \dots, \pi_p)$ . We call  $\boldsymbol{\pi}$  the **initial vector distribution of  $X$** .

**Definition 3.2.1.** Let  $X$  be a Markov jump process as the one we described previously. Then, the time until absorption

$$\tau = \inf\{t \geq 0 : X_t = p+1\}$$

---

<sup>3</sup>We will use this kind of matrix-block notation constantly. To get the proper dimension of the matrix  $\mathbf{0}$ , we need to notice that its height must coincide with the height of the block which is on its right, and its width must coincide with the width of the block which is above; this in this case,  $\mathbf{0}$  is a row vector of dimension  $p$ . From here on, we shall omit these details, unless it is necessary a clarification.

<sup>4</sup>In this case, the necessary dimension of  $\mathbf{e}$  for  $\mathbf{T}\mathbf{e}$  to make sense is  $p$ : again, we shall omit these details, unless it is necessary a clarification.

<sup>5</sup>Being terminating means that  $\sigma := \sup\{t \geq 0 : X_t \in \{1, \dots, p\}\}$  is a.s. finite.  $\sigma$  is known as the **lifetime of the process  $X$** . At time  $\sigma$  one usually sends the process  $X$  to a **cemetery state** where it remains forever.

<sup>6</sup>This can be done by defining the lifetime of the process to be the time of absorption to  $p+1$  and the cemetery state to be  $p+1$ , all of this within the setting defined at the beginning of this section.

<sup>7</sup>This condition will be lifted in further pages, when we consider phase-type distributions which have an atom at 0; see Section 3.5.

is said to have a  $p$ -**dimensional phase-type distribution** or  $PH_p$ -**distribution** with representation  $(\boldsymbol{\pi}, \mathbf{T})$  and we write

$$\tau \sim PH_p(\boldsymbol{\pi}, \mathbf{T}).$$

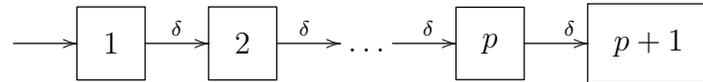
In the case we do not want to specify its dimension, we simply call it **phase-type distribution** or  $PH$ -**distribution**.

**Remark 3.2.2.** Given a phase-type distribution, neither its dimension nor representation is unique.

Here are two examples of phase-type distributions which will be used throughout this manuscript.

**Example 3.2.1.** (Exponential Distribution) *Let  $\tau \sim PH_1(\boldsymbol{\pi}, \mathbf{T})$ , that is, there only exists one transient state in the associated process  $X$ . Notice that the time of absorption is just the time it takes  $X$  to make a jump from 1, that is,  $\tau \sim \text{Exp}(-t_{11})$ . Thus, the family of  $PH_1$ -distribution coincides with the family of exponential distributions.*

**Example 3.2.2.** (Erlang distribution) *Consider the particular case of the distribution  $\text{Gamma}(p, \delta)$  when  $p \in \mathbb{N}$ . This is known as **Erlang distribution of parameters**  $(p, \delta)$  and it is denoted by  $\text{Erl}(p, \delta)$ . Since this distribution corresponds to the convolution of  $p$  exponential distributions of parameter  $\delta$ , the distribution  $\text{Erl}(p, \delta)$  can be represented by the following diagram*



which corresponds to the  $PH_p$ -distribution where  $\boldsymbol{\pi} = (1, 0, \dots, 0, 0)$ ,

$$\mathbf{T} = \begin{pmatrix} -\delta & \delta & 0 & \dots & 0 & 0 \\ 0 & -\delta & \delta & \dots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \dots & -\delta & \delta \\ 0 & 0 & 0 & \dots & 0 & -\delta \end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \delta \end{pmatrix}.$$

It is easy to check that for every  $s \geq 0$  and for  $\boldsymbol{\Lambda}$  as in (3.2.1), we have

$$\mathbf{P}^t = \exp(\boldsymbol{\Lambda}s) = \begin{pmatrix} \exp(\mathbf{T}s) & \mathbf{e} - \exp(\mathbf{T}s)\mathbf{e} \\ 0 & 1 \end{pmatrix};$$

see Section 3.1 in Peralta [15] for a full proof. This means that the transition probabilities (from time 0 to time  $s$ ) between transient states are given by the matrix  $\exp(\mathbf{T}s)$ . This observation allows us to obtain several distributional properties of  $\tau \sim PH_p(\boldsymbol{\pi}, \mathbf{T})$ .

**Theorem 3.2.1.** *Let  $\tau \sim PH_p(\boldsymbol{\pi}, \mathbf{T})$ . Then for all  $s \geq 0$ ,*

1.  $F_\tau(s) = 1 - \boldsymbol{\pi}e^{\mathbf{T}s}\mathbf{e}$ , and
2.  $f_\tau(s) = \boldsymbol{\pi}e^{\mathbf{T}s}\mathbf{t}$ .

*Proof.* Let  $X$  be the Markov jump process associated to  $\tau$ .

1.

$$\begin{aligned}
F_\tau(s) &= \mathbb{P}(\tau \leq s) \\
&= 1 - \mathbb{P}(\tau > s) \\
&= 1 - \mathbb{P}(X_s \in \{1, \dots, p\}) \\
&= 1 - \sum_{i=1}^p \mathbb{P}(X_0 = i) \mathbb{P}(X_s \in \{1, \dots, p\} | X_0 = i) \\
&= 1 - \sum_{i=1}^p \mathbb{P}(X_0 = i) \sum_{j=1}^p \mathbb{P}(X_s = j | X_0 = i) \\
&= 1 - \sum_{i=1}^p \pi_i \sum_{j=1}^p p_{ij}^s \\
&= 1 - \boldsymbol{\pi}e^{\mathbf{T}s}\mathbf{e}.
\end{aligned}$$

2.

$$\begin{aligned}
f_\tau(s)ds &= \mathbb{P}(\tau \in [s, s + ds)) \\
&= \sum_{j=1}^p \mathbb{P}(X_s = j) \mathbb{P}(\tau \in [s, s + ds) | X_s = j) \\
&= \sum_{j=1}^p \sum_{i=1}^p \mathbb{P}(X_0 = i) \mathbb{P}(X_s = j | X_0 = i) \mathbb{P}(\tau \in [s, s + ds) | X_s = j) \\
&= \sum_{j=1}^p \sum_{i=1}^p \pi_i p_{ij}^s t_j ds \\
&= \boldsymbol{\pi}e^{\mathbf{T}s}\mathbf{t}.
\end{aligned}$$

□

The proof of Theorem 3.2.1 shows the standard procedure for working with phase-type distributions, which heavily relies on probabilistic arguments.

**Theorem 3.2.2.** *Let  $\mathbf{T}$  be a sub-intensity matrix and  $\delta \geq 0$ . Then*

1.  $\mathbf{T} - \delta \mathbf{I}$  is a sub-intensity matrix.
2.  $\mathbf{T}$  is invertible iff  $\mathbf{T}$  is associated to an effectively terminating Markov jump process.
3. If either condition of point 2 is true, then  $\mathbf{T} - \delta \mathbf{I}$  is also invertible.

*Proof.* See Theorem 2.1.3 (pp. 37) and Lemma 2.1.4 (pp. 40) in Peralta [15].  $\square$

**Theorem 3.2.3.** *Let  $\tau \sim PH_p(\boldsymbol{\pi}, \mathbf{T})$ ; then its Laplace transform  $L_\tau(s) := \mathbb{E}(e^{-s\tau})$  (for  $s \geq 0$ ) is given by*

$$L_\tau(s) = \boldsymbol{\pi}(s\mathbf{I} - \mathbf{T})^{-1}\mathbf{t}.$$

*Proof.* See Theorem 2.1.5 (pp. 41) in Peralta [15].  $\square$

**Theorem 3.2.4.** *Let  $X \sim PH_d(\boldsymbol{\pi}, \mathbf{T})$  and  $Y \sim PH_p(\boldsymbol{\alpha}, \mathbf{S})$  be independent. Then*

$$\min(X, Y) \sim PH_{dp}(\boldsymbol{\pi} \otimes \boldsymbol{\alpha}, \mathbf{T} \oplus \mathbf{S}), \quad (3.2.2)$$

where  $\otimes$  denotes the Kronecker product and  $\oplus$  the Kronecker sum.

*Proof.* Let us denote by  $E^T$  the set of transient states associated to  $\mathbf{T}$  and by  $E^S$  the transient states associated to  $\mathbf{S}$ . Furthermore, let them be such that  $E^T \cap E^S = \emptyset$ . Let  $\mathbf{T} = \{t_{ij}\}$ ,  $\mathbf{t} = -\mathbf{T}\mathbf{e}$  and  $\mathbf{s} = -\mathbf{S}\mathbf{e}$ . By definition of Kronecker product and Kronecker sum,

$$\boldsymbol{\pi} \otimes \boldsymbol{\alpha} = (\pi_1\boldsymbol{\alpha} \quad \pi_2\boldsymbol{\alpha} \quad \cdots \quad \pi_d\boldsymbol{\alpha}),$$

and

$$\begin{aligned} \mathbf{T} \oplus \mathbf{S} &= \mathbf{T} \otimes \mathbf{I}_p + \mathbf{I}_d \otimes \mathbf{S} \\ &= \begin{pmatrix} t_{11}\mathbf{I}_p + \mathbf{S} & t_{12}\mathbf{I}_p & t_{13}\mathbf{I}_p & \cdots & t_{1d}\mathbf{I}_p \\ t_{21}\mathbf{I}_p & t_{22}\mathbf{I}_p + \mathbf{S} & t_{23}\mathbf{I}_p & \cdots & t_{2d}\mathbf{I}_p \\ t_{31}\mathbf{I}_p & t_{32}\mathbf{I}_p & t_{33}\mathbf{I}_p + \mathbf{S} & \cdots & t_{3d}\mathbf{I}_p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{d1}\mathbf{I}_p & t_{d2}\mathbf{I}_p & t_{d3}\mathbf{I}_p & \cdots & t_{dd}\mathbf{I}_p + \mathbf{S} \end{pmatrix}, \end{aligned} \quad (3.2.3)$$

where  $\mathbf{I}_d$  and  $\mathbf{I}_p$  denote the identity matrix of dimension  $d$  and  $p$ , respectively. By inspection, it is really easy to see that  $\boldsymbol{\pi} \otimes \boldsymbol{\alpha}$  and  $\mathbf{T} \oplus \mathbf{S}$  are indeed an initial vector distribution and a sub-intensity matrix, respectively. The form of both matrices suggest that our new state space is  $E^{T \times S} = \{(i, j) : i \in E^T, j \in E^S\}$ , whose elements are ordered in a **row-major** fashion.<sup>8</sup> With this in mind, the structure of  $\boldsymbol{\pi} \otimes \boldsymbol{\alpha}$  tells us that at time 0 we are picking a pair  $(i, j) \in E^{T \times S}$  with probability  $\pi_i \alpha_j$ ; that is, we are picking the first entry according to  $\boldsymbol{\pi}$  and independently we are choosing the second entry according to  $\boldsymbol{\alpha}$ . Now let  $(i_0, j_0)$  be that initial state. Notice that according to the structure in (3.2.3), the first jump can happen in one of the following three ways: it lands in some  $(k, j_0)$  with intensity  $t_{i_0 k}$ , it lands in some  $(i_0, l)$  with intensity  $s_{j_0 l}$ , or either it gets absorbed with intensity  $t_i + s_i$ . The important thing to notice is that jumps between states in  $E^{T \times S}$  maintain one of their entries fixed, and the one that changes, will do it according to the intensities of  $\mathbf{T}$  or  $\mathbf{S}$  (depending if it is the first or second entry the one that changes); also, we can think of absorption as coming from the first entry (with intensity  $\mathbf{t}$ ) or from the second entry (with intensity  $\mathbf{s}$ ). This is basically the same as letting two independent Markov jump process with parameters  $(\boldsymbol{\pi}, \mathbf{T})$  and  $(\boldsymbol{\alpha}, \mathbf{S})$  evolve in a parallel fashion; in the moment either of them gets absorbed, we stop both processes. Thus, the time until absorption of the process with parameters  $(\boldsymbol{\pi} \otimes \boldsymbol{\alpha}, \mathbf{T} \oplus \mathbf{S})$  is the minimum between the times of absorption of parallel (and independent) Markov jump processes with parameters  $(\boldsymbol{\pi}, \mathbf{T})$  and  $(\boldsymbol{\alpha}, \mathbf{S})$ ; this is equivalent to (3.2.2).  $\square$

Now, consider a sub-intensity matrix  $\mathbf{T}$  of dimension  $p$  and a defective vector distribution  $\boldsymbol{\pi}$  of the same dimension: that is,  $0 < \sum_{i=1}^p \pi_i < 1$ . We can associate the pair  $(\boldsymbol{\pi}, \mathbf{T})$  to an absorbing Markov jump process with space state  $E = \{1, \dots, p, p+1, \Delta\}$  in the following way: it starts in some transient state  $i \in \{1, \dots, p\}$  with probability  $\pi_i$  and will evolve according to the sub-intensity matrix  $\mathbf{T}$ ; otherwise, it starts in the cemetery state  $\Delta$  with probability  $1 - \sum_{i=1}^p \pi_i$  and will stay in that state forever.<sup>9</sup> In this setting we say that the random time  $\tau = \inf\{t \geq 0 : X_t = p+1\}$  has a **defective  $p$ -dimensional phase-type distribution**, and is denoted also

<sup>8</sup>That is, the first  $p$  elements correspond to the first row of  $E^T \times E^S$ , the next  $p$  elements correspond to the second row of  $E^T \times E^S$ , and so on. Whenever we encounter a problem of ordering the elements of a matrix in a linear way, we will use this type of sorting.

<sup>9</sup>In this setting, we need that  $\Delta \neq p+1$ , so that absorption to  $p+1$  and termination will be two entirely different things.

by  $PH_p^*(\boldsymbol{\pi}, \mathbf{T})$ . Note that the event  $\{\tau = \infty\} = \{X_0 = \Delta\}$  happens with probability  $1 - \sum_{i=1}^p \pi_i$ , so  $\tau$  is effectively a defective random variable.

**Theorem 3.2.5.** *Let  $\tau \sim PH_p^*(\boldsymbol{\pi}, \mathbf{T})$ . Then*

1.  $F_\tau(s) = \boldsymbol{\pi} \mathbf{e} - \boldsymbol{\pi} e^{\mathbf{T}s} \mathbf{e}$ , and
2.  $f_\tau(s) = \boldsymbol{\pi} e^{\mathbf{T}s} \mathbf{t}$ .

*Proof.* The same arguments as in the proof of Theorem 3.2.1 apply. However, in the first one we must take into account that the event  $\{\tau = \infty\}$  happens with probability  $1 - \boldsymbol{\pi} \mathbf{e}$ . □

### 3.3 Renewal theory for phase-type distributions.

**Definition 3.3.1.** Let  $\{T_i\}_{i \in \mathbb{N}}$  be a sequence of i.i.d. random variables whose support is  $[0, \infty)$ . Then the process  $\{\sigma_i\}_{i \in \mathbb{N} \cup \{0\}}$  where

$$\sigma_n := \sum_{i=1}^n T_i$$

is called a **renewal process**, each  $\sigma_i$  ( $i = 0, 1, 2, \dots$ ) is called an **arrival (or renovation) time** and each  $T_i$  ( $i = 0, 1, 2, \dots$ ) is called an **interarrival time**. If the distribution of  $T_1$  is such that the event  $\{T_1 = \infty\}$  happens with positive probability (that is,  $T_1$  has a defective distribution), then the process  $\{\sigma_i\}$  will be called a **terminating renewal process**.

Basically, the process  $\{\sigma_i\}$  is a concatenation of random times. One of the fundamental interests in renewal theory is knowing the probability that renewals occur in a certain interval. Moreover, it is interesting to explicitly know the renewal density (in the case it exists at all), which is defined as the function  $u : \mathbb{R}_+ \rightarrow \mathbb{R}$  such that

$$u(x)dx = \mathbb{P}(\text{There exists a renewal in } [x, x + dx)).$$

It turns out that this density is not easy to calculate, even when we assume that the interarrival times have an absolutely continuous distribution; however, when  $T_1 \sim PH_d(\boldsymbol{\pi}, \mathbf{T})$ ,  $u(\cdot)$  can be explicitly calculated with the help of some probabilistic arguments.

**Theorem 3.3.1.** *Consider a renewal process with non-defective interarrival times  $\{T_i\}$  which have a  $PH_p(\boldsymbol{\pi}, \mathbf{T})$  distribution. Then the renewal density  $u(\cdot)$  exists and is given by*

$$u(x) = \boldsymbol{\pi} e^{(\mathbf{T} + \mathbf{t}\boldsymbol{\pi})x} \mathbf{t} \text{ for all } x \geq 0. \quad (3.3.1)$$

*Proof.* Let  $\{J_t^{(k)}\}$  be the underlying Markov jump process of  $T_k$  and define  $\{\widehat{J}_t\}$  to be the ordered concatenation of the sequence  $\{J_t^{(k)}\}$ ; that is,

$$J_t = \begin{cases} J_t^{(1)} & \text{for } 0 \leq t < T_1 \\ J_{t-\sigma_1}^{(2)} & \text{for } \sigma_1 \leq t < \sigma_2 \\ \vdots & \vdots \\ J_{t-\sigma_n}^{(n+1)} & \text{for } \sigma_n \leq t < \sigma_{n+1} \\ \vdots & \vdots \end{cases}$$

Then  $\{\widehat{J}_t\}$  is a Markov jump process which has two kinds of jumps; the jumps “contained” in each  $J_t^{(k)}$ , and the ones corresponding to the transitions from  $J_t^{(k)}$  to  $J_t^{(k+1)}$  for  $k \in \mathbb{N}$ . Jumps of the first type occur according to the intensity matrix  $\mathbf{T}$ , while a jump of the second type, from  $i$  to  $j$ , occurs with intensity  $t_i \pi_j$ .<sup>10</sup> Then, the intensity matrix of  $\widehat{J}$  is  $\mathbf{T} + \mathbf{t}\boldsymbol{\pi}$  and the state-distribution of  $\widehat{J}_x$  is  $\boldsymbol{\pi} e^{(\mathbf{T} + \mathbf{t}\boldsymbol{\pi})x}$ . Since

$$\mathbb{P}(\text{A renovation occurs in } [x, x + dx] | \widehat{J}_x = i) = t_i dx,$$

then

$$\begin{aligned} u(x)dx &= \mathbb{P}(\text{A renovation occurs in } [x, x + dx]) \\ &= \sum_{i=1}^p \mathbb{P}(\widehat{J}_x = i) \mathbb{P}(\text{A renovation occurs in } [x, x + dx] | \widehat{J}_x = i) \\ &= \sum_{i=1}^p (\boldsymbol{\pi} e^{(\mathbf{T} + \mathbf{t}\boldsymbol{\pi})x})_i t_i dx \\ &= \boldsymbol{\pi} e^{(\mathbf{T} + \mathbf{t}\boldsymbol{\pi})x} \mathbf{t} dx, \end{aligned}$$

so the result follows. □

---

<sup>10</sup>Indeed, since we need to get absorbed from  $i$  (which happens with intensity  $t_i$ ) and we also need that the next concatenated process starts in state  $j$  (which happens with probability  $\pi_j$ ). Independence yields the result.

The argument used in the previous proof illustrates the method which we call **concatenation**; that is, a method in which we sequentially merge several processes into one.

**Definition 3.3.2.** Let  $\{\sigma_i\}$  be a terminating renewal process with defective interarrival times  $\{T_i\}_{i \in \mathbb{N}}$ . Let

$$N := \inf\{n \in \mathbb{N} : T_n = \infty\}, \text{ and } \xi := \sum_{i=1}^N T_i;$$

then we call  $\xi$  the **lifetime of the terminating renewal process**  $\{\sigma_i\}$ .

**Theorem 3.3.2.** Consider a terminating renewal process with defective interarrival times  $\{T_i\}$  which have a  $PH_p^*(\boldsymbol{\pi}, \mathbf{T})$  distribution. Then

$$\mathbb{P}(\xi > x) = \boldsymbol{\pi} e^{(\mathbf{T} + \mathbf{t}\boldsymbol{\pi})x} \mathbf{e}.$$

*Proof.* We will use the same idea of concatenation introduced in the proof of Theorem 3.3.1. That is, take each underlying process of  $\{T_i\}$  and glue them together:<sup>11</sup> this way we have that the sub-intensity matrix of this concatenated process is given by  $\mathbf{T} + \mathbf{t}\boldsymbol{\pi}$  and its initial distribution is given by  $\boldsymbol{\pi}$ . Thus

$$\begin{aligned} \mathbb{P}(\xi > x) &= \mathbb{P}(\text{The concatenated process is in } E \text{ at time } x) \\ &= \sum_{i=1}^p \mathbb{P}(\text{The concatenated process is in } i \text{ at time } x) \\ &= \sum_{i=1}^p (\boldsymbol{\pi} e^{(\mathbf{T} + \mathbf{t}\boldsymbol{\pi})x})_i \\ &= \boldsymbol{\pi} e^{(\mathbf{T} + \mathbf{t}\boldsymbol{\pi})x} \mathbf{e}. \quad \square \end{aligned}$$

**Remark 3.3.1.** The particular arguments given in the previous proof will be heavily used throughout Part 3; in there, the proofs may not be as detailed as this one, but the idea is basically the same: merging processes into one and computing the distribution of the lifetime of this concatenated process.

---

<sup>11</sup>Here we have that  $T_1$  is defective and although the concatenated process is terminating, its construction is basically the same as the one given in the proof of Theorem 3.3.1

### 3.4 Uniformization.

Consider a (terminating) Markov jump process  $X = \{X_t\}_{t \geq 0}$  with finite state space  $E = \{1, \dots, n\}$ , initial distribution  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)$  and (sub-)intensity matrix  $\mathbf{T} = \{t_{ij}\}_{1 \leq i, j \leq n}$ . Fix  $\theta \geq \max_{0 \leq i \leq n} -t_{ii}$  and consider the matrix  $\mathbf{S} = \mathbf{I} + (1/\theta)\mathbf{T}$ . Then we have the following:

**Lemma 3.4.1.** *If  $\mathbf{T}$  is a (sub-) intensity matrix, then  $\mathbf{S}$  is a (sub-) stochastic matrix.*

*Proof.* Let  $\mathbf{T}$  be an intensity matrix. Fix  $i, j \in \{1, \dots, n\}$ : if  $i \neq j$  then  $s_{ij} = (1/\theta)t_{ij} \geq 0$ , and if  $i = j$  then  $s_{ij} = 1 + (1/\theta)t_{ij} = (\theta + t_{ij})/\theta \geq 0$ . Also,  $\sum_j s_{ij} = 1 + \sum_j t_{ij}/\theta = 1$ . The case when  $\mathbf{T}$  is a sub-intensity matrix is similar.  $\square$

Now, notice that for  $x \geq 0$ ,

$$\begin{aligned} \boldsymbol{\pi} \exp(\mathbf{T}x) &= \boldsymbol{\pi} \exp\left(\theta \left(-\mathbf{I} + \mathbf{I} + \frac{1}{\theta}\mathbf{T}\right)x\right) \\ &= \boldsymbol{\pi} \exp(\theta(-\mathbf{I} + \mathbf{S})x) \\ &= \boldsymbol{\pi} \exp(\theta\mathbf{S}x)e^{-\theta x} \\ &= \sum_{i=0}^{\infty} \boldsymbol{\pi} \mathbf{S}^i e^{-\theta x} \frac{(\theta x)^i}{i!}; \end{aligned}$$

we may recognize the terms  $\boldsymbol{\pi} \mathbf{S}^i$  as the row vector that provides the probability distribution (over  $E$ ) after  $i$  steps of the discrete-time Markov chain whose initial distribution is given by  $\boldsymbol{\pi}$  and whose (sub-)stochastic transition probability matrix is given by  $\mathbf{S}$ . The term  $e^{-\theta x}(\theta x)^i/i!$  may be recognized as the probability of  $i$  arrivals from a Poisson process with rate  $\theta$  up to the time  $x$ . That is, the (terminating) process  $\{X_t\}_{t \geq 0}$  may be regarded as the process  $\{Y_{N_t}\}_{t \geq 0}$ , where  $\{Y_n\}_{n \in \{0, 1, 2, \dots\}}$  is a (terminating) Markov chain with (sub-)stochastic matrix  $\mathbf{S}$  and  $\{N_t\}_{t \geq 0}$  is an independent Poisson process with rate  $\theta$ . This is the idea behind **uniformization**, which was introduced in Grassmann [7]. It is called this way because now the jumps happen at a constant rate; nevertheless, a jump can happen to the same state it came from.

Later in this manuscript we shall need the following extension of the uniformization method. Consider the bivariate process  $\{Z_n, M_t\}_{n \in \{0, 1, \dots\}, t \geq 0}$ ,

where  $\{Z_n\}_{n \in \{0,1,\dots\}}$  is a Markov chain with initial distribution  $\boldsymbol{\pi}$  and whose (sub-)stochastic matrix is given by  $\mathbf{U} = \mathbf{I} + \boldsymbol{\Delta}_{\mathbf{1}/\boldsymbol{\eta}_i} \mathbf{T}$  where  $\boldsymbol{\Delta}_{\mathbf{1}/\boldsymbol{\eta}_i} = \text{diag}\{1/\eta_i : 1 \leq i \leq n\}$  for some fixed  $\eta_i > -t_{ii}$ ,  $i \in \{1, \dots, n\}$ , and  $\{M_t\}_{t \geq 0}$  is an arrival process with exponential interarrival times whose successive parameters are given by the sequence  $\{\eta_{Z_{i-1}}\}_{i \in \{1,\dots\}}$ . Then we have the following:

**Theorem 3.4.1.** *The process  $\{X_t\}_{t \geq 0}$  is equal in distribution to the process  $\{Z_{M_t}\}_{t \geq 0}$ .*

*Proof.* It suffices to study the distributional properties of the first jump. Let  $\tau_0 = \inf\{s > 0 : Z_{M_s} \neq Z_0\}$  and assume that  $Z_0 = k$ . Then

$$\begin{aligned}
\mathbb{P}(\tau_0 > t) &= \mathbb{P}(Z_{M_s} = k \text{ for every } s \in [0, t]) \\
&= \sum_{j=0}^{\infty} \mathbb{P}(M_t = j, Z_1 = \dots = Z_j = k) \\
&= \sum_{j=0}^{\infty} \mathbb{P}(M_t = j | Z_1 = \dots = Z_j = k) \mathbb{P}(Z_1 = \dots = Z_j = k) \\
&= \sum_{j=0}^{\infty} e^{-\eta_k t} \frac{(\eta_k t)^j}{j!} (1 + t_{kk}/\eta_k)^k \\
&= e^{-\eta_k t} e^{\eta_k t(1+t_{kk}/\eta_k)} \\
&= e^{-t(-t_{kk})};
\end{aligned}$$

this means that its first “real” jump occurs at an exponential time of parameter  $-t_{kk}$ . Now, for  $l \neq k$ ,

$$\mathbb{P}(Z_{M_{\tau_0}} = l) = \mathbb{P}(Z_1 = l | Z_1 \neq k) = \frac{\mathbb{P}(Z_1 = l)}{1 - \mathbb{P}(Z_1 = k)} = \frac{t_{kl}/\eta_k}{1 - (1 + t_{kk}/\eta_k)} = -\frac{t_{kl}}{t_{kk}}.$$

Since the distribution of time between “real” jumps and the distribution of where said jumps land are the same as in the process  $X$ , then we have that  $\{X_t\}_{t \geq 0}$  and  $\{Z_{M_t}\}_{t \geq 0}$  are equal in distribution.  $\square$

Notice that the previous proof formalizes the intuitive idea given at the beginning of this section about the method of uniformization. Also notice that the uniformization scheme creates “non-real” jumps: that is why we are going to refer to them as **transitions** instead of jumps.

### 3.5 The least variable phase-type distribution of order $n$ .

Recall from Section 3.2 the definition of a defective phase-type distribution, that is, a phase-type distribution which has an atom at  $+\infty$  of size  $1 - \boldsymbol{\pi}\mathbf{e}$ , where  $\boldsymbol{\pi}$  is the initial distribution parameter of said  $PH^*$ -distribution. In several settings we wish to consider phase-type distributions which have atoms at 0 instead of at  $+\infty$ : to do, this we extend (without any effort) the definition of a phase-type distribution  $PH(\boldsymbol{\pi}, \mathbf{T})$  to allow an atom at 0 of size  $1 - \boldsymbol{\pi}\mathbf{e}$ .<sup>12</sup> Hence, whenever we encounter a pair  $(\boldsymbol{\pi}, \mathbf{T})$  such that  $\boldsymbol{\pi}$  is a non-negative row vector of size  $n$ ,  $\boldsymbol{\pi}\mathbf{e} < 1$  and  $\mathbf{T}$  is an invertible sub-intensity  $n \times n$ -matrix, it can be associated to either a  $PH_n^*$ -distribution (that is, a defective one) or to a  $PH_n$ -distribution (that is, one that has an atom at 0). In this section we are only interested in working with non-defective  $PH$ -distributions with (possible) atoms at 0.

**Remark 3.5.1.** Within the previous framework, the point mass distribution at 0 is considered a  $PH^n$ -distribution for any  $n \in \mathbb{N}$ . Moreover, if we extend the definition of an Erlang distribution with parameters  $(n, \lambda)$  to allow  $\lambda \in (0, +\infty]$ , then the point mass distribution at 0 is also an Erlang distribution (of any order): we are working with this extended definition during this section.

**Corollary 3.5.1.** *Consider the collection of distributions  $\{Erl(n, \lambda_i)\}_{i=1}^m$  for  $m, n \in \mathbb{N}$  and  $\lambda_i \in (0, +\infty]$ . Then any mixture of those distributions majorizes the Erlang distribution of order  $n$  which has the same mean as said mixture.*

*Proof.* Consider the mixture  $\sum_{i=1}^m \alpha_i Erl(n, \lambda_i)$  for  $a_1, \dots, a_m \geq 0$  and  $\sum a_i = 1$ . If  $X \sim Erl(n, 1)$  and  $Y$  is an independent discrete random variable that takes the value  $1/\lambda_i$  with probability  $\alpha_i$ ,  $i \in \{1, \dots, m\}$ , then  $XY \sim \sum_{i=1}^m \alpha_i Erl(n, \lambda_i)$ , so the result follows from Lemma 2.3.2.  $\square$

For every measure  $\mu$  over  $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ , let  $\mathbb{E}_\mu := \int x\mu(dx)$ : this notation will be used throughout the next lemma and theorem. Also, we are going to denote by  $\mathbf{1}_i$  the column vector whose  $i$ -th entry is 1 and the remaining ones are 0.<sup>13</sup>

<sup>12</sup>One direct way to do this is by taking  $\mathbb{P}(X_0 = p + 1) = 1 - \boldsymbol{\pi}\mathbf{e}$  in Section 3.2: it is easy to see that this leaves us with a phase-type distribution with an atom at 0.

<sup>13</sup>As in previous sections, its dimension will be implicit in the situation it is being used.

**Lemma 3.5.1.** *Let  $n \geq 2$  and suppose that for any  $PH_{n-1}$  distribution, say  $\mu^{n-1}$ , we have that  $\nu^{n-1} \prec \mu^{n-1}$ , where  $\nu^{n-1} \sim \text{Erl}(n-1, (n-1)/\mathbb{E}_{\mu^{n-1}})$ . If  $\mu_1 \sim PH_n(\mathbf{1}'_1, \mathbf{T})$ , we have that  $\nu_1 \prec \mu_1$ , where  $\nu_1 = \text{Erl}(n, n/\mathbb{E}_{\mu_1})$ .*

*Proof.* Let  $Z$  be an absorbing Markov jump process with sub-intensity matrix  $\mathbf{T} = \{t_{ij}\}_{ij \in \{1, \dots, n\}}$  on its transient states  $\{1, \dots, n\}$ , which initializes in the state 1, and let  $\tau$  be the time of its absorption. Define  $\tau_1$  to be the time the process spends in the state 1 before the absorption and  $\tau_{>1}$  to be the time the process spends in the states  $\{2, \dots, n\}$ . Clearly  $\tau_1 + \tau_{>1} = \tau$ , but notice that  $\tau_1$  and  $\tau_{>1}$  are not independent. Let  $N$  be the number of visits  $Z$  makes to state 1. For  $i = 1, \dots, N-1$ , let  $X_i$  be the time the process spends in state 1 during the  $i$ -th visit and let  $U$  be the time it spends in its last visit to 1. Let  $Y_i$  be the time the process spends in the states  $\{2, \dots, n\}$  between the  $i$ -th and  $i+1$ -th visit to 1, and let  $V$  be the time the process spends in its last visit to the states  $\{2, \dots, n\}$  right before absorption (define  $V$  to be 0 if the absorption comes directly from state 1). By the strong Markov property, given  $N = m \geq 1$ ,  $X_1, \dots, X_{m-1}$  are i.i.d.,  $Y_1, \dots, Y_{m-1}$  are i.i.d.<sup>14</sup> and these sequences, together with  $U$  and  $V$  are independent between them; also, their distributions are independent from  $m$ . We extend these sequences to  $\{X_i\}_{i=1}^\infty$  and  $\{Y_i\}_{i=1}^\infty$  which are i.i.d., independent between them and from  $U$  and  $V$ . Let us introduce a new r.v.,  $N'$ , which is independent from the previous r.v.'s and has the same distribution as  $N$ . Then, by Theorem 2.3.3, we have

$$\mathcal{L} \left( \sum_{i=1}^{N-1} X_i + \sum_{i=1}^{N'-1} Y_i \right) \prec \mathcal{L} \left( \sum_{i=1}^{N-1} (X_i + Y_i) \right).$$

---

<sup>14</sup>In case they exist,  $X_i$  is just an exponential distribution whose parameter is the exit rate from state 1; and  $Y_i$  is a  $PH_{n-1}$ -distribution which has an initial distribution that is given by  $(-t_{12}/t_{11}, -t_{13}/t_{11}, \dots, -t_{1n}/t_{11})$ , or in other words, the exit probabilities from 1 to  $(2, \dots, n)$ .

Using Lemma 2.3.3, we have

$$\begin{aligned}
\mathcal{L}(\tau'_1 + \tau'_{>1}) &= \mathcal{L} \left( \left( \sum_{i=1}^{N-1} X_i + U \right) + \left( \sum_{i=1}^{N'-1} Y_i + V \right) \right) \\
&\prec \mathcal{L} \left( \sum_{i=1}^{N-1} (X_i + Y_i) + U + V \right) \\
&= \mathcal{L}(\tau_1 + \tau_{>1}) \\
&= \mathcal{L}(\tau),
\end{aligned}$$

where  $\tau'_1 \sim \tau_1$  and  $\tau'_{>1} \sim \tau_{>1}$ , but now  $\tau'_1 \perp \tau'_{>1}$ . Notice that  $\tau'_{>1}$  is the distribution of the occupation time of  $Z$  in the states  $\{2, \dots, n\}$ ; hence it is  $PH_{n-1}$ -distributed;<sup>15</sup> thus by hypothesis,  $\mathcal{L}(\sigma_{n-1}) \preceq_2 \mathcal{L}(\tau'_{>1})$  where  $\sigma_{n-1} \sim \text{Erl}(n-1, (n-1)/\mathbb{E}(\tau'_{>1}))$  and  $\sigma_{n-1}$  is independent from  $\tau'_1$ . Using the same reasoning as before,  $\tau'_1$  must be an exponentially distributed random variable. Then  $\tau'_1 + \sigma_{n-1}$  is basically a weighted sum of  $n$  i.i.d. exponential random variables so according to Lemma 2.3.1,  $\mathcal{L}(\sigma_n) \prec \mathcal{L}(\tau'_1 + \sigma_{n-1})$  where  $\sigma_n \sim \text{Erl}(n, n/\mathbb{E}(\tau'_1 + \sigma_{n-1})) = \text{Erl}(n, n/\mathbb{E}_{\mu_1})$ . In conclusion, we have that

$$\nu_1 = \mathcal{L}(\sigma_n) \prec \mathcal{L}(\tau'_1 + \sigma_{n-1}) \prec \mathcal{L}(\tau'_1 + \tau'_{>1}) \prec \mathcal{L}(\tau) = \mu_1,$$

and the proof is complete.  $\square$

**Remark 3.5.2.** By relabelling, we can replace “ $PH_n(\mathbf{1}'_1, \mathbf{T})$ ” with “ $PH_n(\mathbf{1}'_i, \mathbf{T})$ ” for any  $i \in \{2, \dots, n\}$  in the statement of Lemma 3.5.1 and the result still holds.

**Theorem 3.5.1.** *Let  $\mu = PH_n(\pi, T)$ . Then  $\nu \prec \mu$ , where  $\nu = \text{Erl}(n, n/\mathbb{E}_\mu)$ .*

*Proof.* The proof is by induction on representation order. In the case  $n = 1$ ,  $\mu = (1 - \pi_1)\delta_0 + \pi_1\mu_1$  where  $\mu_1$  is an exponential distribution, that is,  $\mu$  is the mixture of Erlang distributions of dimension 1,<sup>16</sup> so by Corollary 3.5.1 the statement is true in this case.

<sup>15</sup>Other explanation is that  $\tau'_{>1}$  is a geometric sum of i.i.d.  $PH_{n-1}$  random variables, so it is also a  $PH_{n-1}$  random variable.

<sup>16</sup>Recall that  $\delta_0$  can be considered as an Erlang distribution.

Now suppose that the statement holds for  $PH_{n-1}$ -distributions; by Lemma 3.5.1 and the previous Remark, it must also hold for  $PH_n$ -distributions whose initial distribution is non-random; that is,

$$\mu_i \succ \nu_i \text{ for all } i \in \{1, \dots, n\}, \quad (3.5.1)$$

where  $\mu_i \sim PH_n(\mathbf{1}'_i, \mathbf{T})$  and  $\nu_i \sim Erl(n, n/\mathbb{E}_{\mu_i})$ . Then, we would have the next string of equalities and “inequalities”:

$$\mu = \sum_{i=1}^n \pi_i \mu_i + \left(1 - \sum_{i=1}^n \pi_i\right) \delta_0 \succ \sum_{i=1}^n \pi_i \nu_i + \left(1 - \sum_{i=1}^n \pi_i\right) \delta_0 \succ \nu,$$

where  $\nu \sim Erl(n, n/\mathbb{E}_{\sum_{i=1}^n \pi_i \mathbb{E}_{\mu_i}}) = Erl(n, n/\mathbb{E}_{\mu})$ . The first equality follows from a standard decomposition of  $PH_n$ -distributions. The middle “inequality” follows from (3.5.1) and from the first part of Lemma 2.3.3. The last “inequality” follows from Corollary 3.5.1. This proves the main result.  $\square$

We have shown that within the class of phase-type distributions of dimension  $n$  and fixed mean, the Erlang distribution is the least variable one in the convex order sense.

### What is erlangization and why we use it?

Suppose that we have a complex stochastic model which has some random component which is Erlang-distributed and suppose that we are able to compute the probability of some event of this stochastic model.<sup>17</sup> What if we are more interested in having a non-random component of length  $c > 0$  instead of an Erlang-distributed one?<sup>18</sup> The approach that makes the most sense (and is actually the easiest in most cases) is to consider the sequence  $\{Erl(n, n/c)\}_{n=1}^{\infty}$ : it is fairly easy to prove that the mean of each one of those distributions is  $c$  and that actually  $Erl(n, n/c)$  converges weakly to  $\delta_c$  as  $n \rightarrow \infty$ . This way, if we compute the desired probability in the case the component is  $Erl(n, n/c)$ -distributed for a sufficiently large  $n$ , we can

<sup>17</sup>For example, suppose that we are able to compute the probability that some risk process gets ruined before an Erlang-distributed time (we will explain this example carefully in Chapter 5).

<sup>18</sup>It makes sense that we are more interested in computing the finite-horizon probability of ruin rather than an Erlang-horizon probability of ruin, or at least insurance companies are more interested in that.

get an approximation for the non-random case.<sup>19</sup> This method is known as **erlangization**.

Now that we have given an heuristic idea of the erlangization method, we need to ask ourselves another question: why do we use this method in particular? To answer this, it is important to state that a large portion of stochastic models are computationally tractable whenever we assume that its components have a markovian-like behaviour, and in particular, are tractable whenever we work with phase-type distributions.<sup>20</sup> Going back to the situation of the previous paragraph, it is a rule of thumb that whenever “something” can be computed in the Erlang case, it may also be computable in the phase-type case: suppose that this is actually true, so now we are able to compute the probability of some event associated to a stochastic model with some  $PH$ -distributed component. Can we choose an alternative approach similar to erlangization which consists in approximating  $\delta_c$  with another sequence of  $PH$ -distributed random variables of increasing dimension and fixed mean  $c$ ? The answer is yes, but we have a major drawback: if we take any other sequence, it will have more variability (in the convex order sense) than the erlangization method, so the convergence to  $\delta_c$  could be slower. This is the whole point of this chapter: to convince the reader that the erlangization method is the best one amongst any other  $PH$ -method we can think of.

In the following chapters we will work with a pair of risk processes and we will study several types of ruin for them. It turns out that the probability of these ruins are easily computable whenever we assume that one specific component is  $PH$ -distributed, however, we are actually more interested in the case when that component is non-random: thus, we can apply the erlangization method to compute an approximation.

---

<sup>19</sup>Only in some specific cases the convergence of these probabilities can be rigorously proved: for complex systems one can just hope for the best, but it is difficult to imagine practical cases in which this convergence could fail.

<sup>20</sup>It is not that whenever we have phase-type distributed components, explicit quantities can always be computed: it is just that phase-type distributions have been shown to be very useful to compute exact probabilities in areas like queueing and risk theory, so one hopes that having these “phase-type” components in a model will make things easier (at least this is the case throughout this whole manuscript).

# Chapter 4

## Fluid-flow and risk models.

In this chapter we study some classic stochastic models.

In Section 4.1 we will review the fluid-flow process, which is a model used to describe the fluid level in a reservoir subject to randomly determined periods of fillings and emptying; in this case, those random periods will be determined by a Markov jump process with finite state space. We will find a method for getting the distribution of the maximum of a fluid-flow model via an iterative algorithm, based on Asmussen [3].

In Section 4.2 we will study two of the most important classic risk models: the *PH*-Cramér-Lundberg and *PH*-Sparre-Andersen processes. Both processes model the reservoir of an insurance company; the difference between them is that the *PH*-Cramér-Lundberg process is a Lévy process while the *PH*-Sparre-Andersen process has an underlying renewal process. Several results about the probability of ruin are presented for both processes: in this manuscript the probability of ruin for the latter is calculated by using fluid-flow arguments; the usual approach is to use the theory of random walks.

## 4.1 The fluid-flow model.

Let  $\{J_t\}_{t \geq 0}$  denote a non-absorbing Markov jump process over a finite state-space  $E = \{1, \dots, d\}$  with intensity matrix  $\boldsymbol{\lambda} = \{\lambda_{ij}\}_{i,j \in E}$ . Also, consider some fixed constants  $r_i = r(i) \neq 0$  for  $i = 1, \dots, d$ . We define the **fluid-flow process**  $\{V_t\}_{t \geq 0}$  by

$$V_t = \int_0^t r(J_u) du.$$

This means that in an interval where  $J_t = i$ , the process  $\{V_t\}$  increases linearly with slope  $r_i$ . Let us consider an interval of this kind and denote by  $X$  its length: we know that  $X$  is exponentially distributed with parameter  $\lambda_i := -\lambda_{ii} > 0$ . Then, the fluid-flow process increases or decreases (depending on  $\text{sgn}(r_i)$ )  $|r_i|X$  units in total during the length of said interval. Since  $|r_i|X \sim \text{Exp}(\lambda_i/|r_i|)$ , then we conclude that the sequence of piecewise increments and decrements of  $\{V_t\}_{t \geq 0}$  are exponentially distributed of successive parameters  $\{\lambda_{J_{\sigma_i}}/r_{J_{\sigma_i}}\}_{i=1}^\infty$ , where  $\sigma_1, \sigma_2, \dots$  correspond to the jump times of the process  $J$ . This leads us to define

$$\mathbf{T} := \boldsymbol{\Delta}_{\mathbf{1}/|\mathbf{r}|} \boldsymbol{\lambda}, \quad (4.1.1)$$

where

$$\boldsymbol{\Delta}_{\mathbf{1}/|\mathbf{r}|} := \text{diag} \left( \frac{1}{|r_1|}, \dots, \frac{1}{|r_d|} \right).$$

Hence,

$$\mathbf{T} = \begin{pmatrix} \frac{\lambda_{11}}{|r_1|} & \frac{\lambda_{12}}{|r_1|} & \dots & \frac{\lambda_{1d}}{|r_1|} \\ \frac{\lambda_{21}}{|r_2|} & \frac{\lambda_{22}}{|r_2|} & \dots & \frac{\lambda_{2d}}{|r_2|} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\lambda_{d1}}{|r_d|} & \frac{\lambda_{d2}}{|r_d|} & \dots & \frac{\lambda_{dd}}{|r_d|} \end{pmatrix}.$$

**Remark 4.1.1.** From here on, we shall consider the bivariate stochastic process  $\{(J_t, V_t)\}_{t \geq 0}$ , and even if we are only talking about  $\{V_t\}$ , we should keep in mind that the process  $\{J_t\}$  is associated to it. Another detail that needs to be pointed out is that  $\{V_t\}$  by itself is not an homogeneous Markov process, but  $\{(J_t, V_t)\}$  is:<sup>1</sup> this property will be heavily used throughout this manuscript, usually without explicitly mentioning it.

---

<sup>1</sup>It is actually a strong Markov process, since it inherits its properties from  $\{J_t\}$ .

At any given time  $t$ ,  $V$  will be either increasing or decreasing, depending on  $\text{sgn}(r_{J_t})$ : hence we may partition the state space  $E = E^+ \cup E^-$ ,<sup>2</sup> where

$$E^+ = \{i \in E : r_i > 0\}, \quad \text{and} \quad E^- = \{i \in E : r_i < 0\}.$$

With this partition in mind, we need to rearrange the elements of  $\mathbf{T}$  into

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}^{++} & \mathbf{T}^{+-} \\ \mathbf{T}^{-+} & \mathbf{T}^{--} \end{pmatrix},$$

where  $\mathbf{T}^{++}$  is the block which contains the transitions between the states in  $E^+$ ,  $\mathbf{T}^{+-}$  is the block which contains the transitions from the states in  $E^+$  to the states in  $E^-$ ,  $\mathbf{T}^{-+}$  is the block which contains the transitions from the states in  $E^-$  to the states in  $E^+$ , and  $\mathbf{T}^{--}$  is the block which contains the transitions between the states in  $E^-$ .

For a moment, let us consider the process  $\{(J_t, V_t)\}$  such that  $V_0 = 0$  and  $J_0 = i$  for some  $i \in E^-$ . Define  $\alpha_{ij}^{-+}$  to be the probability that the process  $V$  up-crosses for the first time the level 0 while the process  $J$  is in state  $j \in E^+$ . Let  $\boldsymbol{\alpha}^{-+} := \{\alpha_{ij}^{-+}\}_{i \in E^-, j \in E^+}$ . This matrix of dimension  $|E^-| \times |E^+|$  will be referred as the **matrix of up-crossing probabilities**.

Next, define  $\tau(x) = \inf\{t > 0 : V_t = x\}$  and consider the (possibly terminating) stochastic process  $\{m(x)\}$  defined by  $m(x) = J_{\tau(x)}$  for  $x \geq 0$  such that  $\tau(x) < \infty$ . That is,  $\{m(x)\}$  is the ascending ladder process (taking values in  $E^+$ ) that takes note of the state in which the process  $J$  was at the instant at which the process  $V$  reached the level  $x$  for the first time. If the main process is such that  $V_0 = 0$  and  $J_0 = i$  for some  $i \in E$ , then the initial distribution of  $\{m(x)\}$  (that is, the distribution of  $m(0)$ ) is  $\mathbf{1}'_i$  if  $i \in E^+$ , and  $\boldsymbol{\alpha}_i^{-+} := \{\alpha_{ij}^{-+}\}_{j \in E^+}$  if  $i \in E^-$ .

The process  $\{m(x)\}$  can jump from a state  $i \in E^+$  to  $j \in E^+$  in two ways: either directly (which happens with intensity  $\lambda_{ij}/r_i$ , according to the discussion which leads to the definition in (4.1.1)), or by jumping from  $i$  to another state  $k \in E^-$  (which happens with intensity  $\lambda_{ik}/r_i$ ) and then up-crossing for the first time the previous maximum level in state  $j$  (which

---

<sup>2</sup>In further pages we will want to consider the case in which  $E_0 := \{i \in E : r_i = 0\}$  is not empty. However, working in this setting for now does no harm.

happens with probability  $\alpha_{kj}^{-+}$ ). That is,

$$\mathbb{P}(m(x + dx) = j | m(x) = i) = \frac{\lambda_{ij}}{r_i} dx + \sum_{k \in E^-} \frac{\lambda_{ik}}{r_i} \alpha_{kj}^{-+} dx.$$

This discussion implies that  $m(x)$  is a (possibly terminating) Markov jump process with intensity matrix given by

$$\mathbf{U} = \mathbf{T}^{++} + \mathbf{T}^{+-} \boldsymbol{\alpha}^{-+}. \quad (4.1.2)$$

If we were capable of knowing the exact values of  $\boldsymbol{\alpha}^{-+}$ , then we would be able to describe the distributional behaviour of  $m(x)$  explicitly. Unfortunately, computing directly the values of  $\boldsymbol{\alpha}^{-+}$  is not an easy task: nevertheless, in the following pages we are giving an iterative algorithm whose limit is said matrix of probabilities.

Let us focus first on obtaining the probability  $\alpha_{ij}^{-+}$  for  $i \in E^-$  and  $j \in E^+$ : consider the case when  $V_0 = 0$  and  $J_0 = i$ . We have discussed before that  $V_{\sigma_1}$  is exponentially distributed with rate  $\lambda_i/|r_i|$ ; then at time  $\sigma_1$  the process  $J$  will change to some other state  $k \in E$ . Consider an independent and identically distributed fluid-flow process  $\{(J', V')\}$  such that  $J'_0 = k$  and  $V'_0 = 0$ . By the strong Markov process and homogeneity, the conditional probability that  $V$  up-crosses the level 0 while  $J$  is in state  $j \in E^+$  (given that  $V_{\sigma_1} = -x$  ( $x \geq 0$ ) and the first jump of  $J$  is to state  $k$ ) is equal to the probability that  $V'$  up-crosses the level  $x$  for the first time while  $J$  is in state  $j \in E^+$ . But this probability has an explicit form (see the discussion after (4.1.2)), which is

$$\begin{aligned} & \mathbb{P}(V \text{ up-crosses } 0 \text{ while } J \text{ is in the state } j | \text{First jump of } J \text{ is to } k, X = x) \\ &= (\mathbf{1}'_k \delta_{E^+}(k) + \boldsymbol{\alpha}_k^{-+} \delta_{E^-}(k)) e^{Ux} \mathbf{1}_j. \end{aligned}$$

Integrating over the possible values of  $V_{\sigma_1}$ , summing over all  $k \in E^+ \cup E^-$  and recalling that the transition probabilities between the states of  $J$  are

given by  $\{p_{ik} := (1 - \delta_{ik})\lambda_{ik}/\lambda_i\}_{i,k \in E}$ , we have that

$$\begin{aligned} \alpha_{ij}^{-+} &= \mathbb{P}(V \text{ up-crosses } 0 \text{ while } J \text{ is in the state } j) \\ &= \int_0^\infty \left( \sum_{k \in E^+ \cup E^-} (\mathbf{1}'_k \delta_{E^+}(k) + \boldsymbol{\alpha}_k^{-+} \delta_{E^-}(k)) e^{Ux} \mathbf{1}_j p_{ik} \right) \frac{\lambda_i}{|r_i|} e^{-\frac{\lambda_i}{|r_i|}x} dx \end{aligned} \quad (4.1.3)$$

$$= \sum_{k \in E^+ \cup E^-} \frac{\lambda_{ik}}{|r_i|} (1 - \delta_{ik}) (\mathbf{1}'_k \delta_{E^+}(k) + \boldsymbol{\alpha}_k^{-+} \delta_{E^-}(k)) \left( \int_0^\infty e^{(U - \frac{\lambda_i}{|r_i|} \mathbf{I})x} dx \right) \mathbf{1}_j. \quad (4.1.4)$$

From here on, we assume that  $\{m(x)\}$  is a terminating process.<sup>3</sup> Then,  $\mathbf{U}$  is an invertible sub-intensity matrix; by Theorem 3.2.2,  $\mathbf{U} - \lambda_i/|r_i| \mathbf{I}$  ( $x \geq 0$ ) is also an invertible sub-intensity matrix. Moreover,

$$\int_0^\infty e^{(U - \frac{\lambda_i}{|r_i|} \mathbf{I})x} dx = \left( \frac{\lambda_i}{|r_i|} \mathbf{I} - \mathbf{U} \right)^{-1}. \quad (4.1.5)$$

If we let  $\boldsymbol{\alpha}_i^{-+}$  denote the  $i$ -th row of  $\boldsymbol{\alpha}^{-+}$ , then (4.1.4) and (4.1.5) imply that

$$\begin{aligned} \boldsymbol{\alpha}_i^{-+} &= \left( \sum_{k \in E^+ \cup E^-} \frac{\lambda_{ik}}{|r_i|} (1 - \delta_{ik}) (\mathbf{1}'_k \delta_{E^+}(k) + \boldsymbol{\alpha}_k^{-+} \delta_{E^-}(k)) \right) \left( \frac{\lambda_i}{|r_i|} \mathbf{I} - \mathbf{U} \right)^{-1} \\ &= \left( \sum_{k \in E^+} \frac{\lambda_{ik}}{|r_i|} \mathbf{1}'_k + \sum_{k \in E^-} \frac{\lambda_{ik}}{|r_i|} \boldsymbol{\alpha}_k^{-+} + \frac{\lambda_i}{|r_i|} \boldsymbol{\alpha}_i^{-+} \right) \left( \frac{\lambda_i}{|r_i|} \mathbf{I} - \mathbf{U} \right)^{-1}. \end{aligned} \quad (4.1.6)$$

Notice that

1.  $\sum_{k \in E^+} (\lambda_{ik}/|r_i|) \mathbf{1}'_k$  corresponds to the  $i$ -th row of  $\mathbf{T}^{-+}$ ,
2.  $\sum_{k \in E^-} (\lambda_{ik}/|r_i|) \boldsymbol{\alpha}_k^{-+}$  corresponds to the  $i$ -th row of  $\mathbf{T}^{--} \boldsymbol{\alpha}^{-+}$ , and
3.  $(\lambda_i/|r_i|) \boldsymbol{\alpha}_i^{-+}$  corresponds to the  $i$ -th row of  $\boldsymbol{\Delta}_{\lambda/|r|} \boldsymbol{\alpha}^{-+}$  (where  $\boldsymbol{\Delta}_{\lambda/|r|} = \text{diag}(\lambda_i/|r_i| : i \in E^-)$ ).

---

<sup>3</sup>This can happen in two cases: either  $V$  has a negative drift or if the process  $J$  eventually gets absorbed in some state  $\Delta$  such that  $r_\Delta = 0$ . The latter is the case which we will study in Chapters 5 and 6.

This implies that

$$\begin{aligned} \alpha_i^{-+} &= \left( (\mathbf{T}^{-+} + \mathbf{T}^{--}\alpha^{-+} + \Delta_{\lambda/|r|}\alpha^{-+}) \left( \frac{\lambda_i}{|r_i|} \mathbf{I} - \mathbf{U} \right)^{-1} \right)_i & (4.1.7) \\ &= \left( (\mathbf{T}^{-+} + \mathbf{T}^{--}\alpha^{-+} + \Delta_{\lambda/|r|}\alpha^{-+}) \left( \frac{\lambda_i}{|r_i|} \mathbf{I} - (\mathbf{T}^{++} + \mathbf{T}^{+-}\alpha^{-+}) \right)^{-1} \right)_i, & (4.1.8) \end{aligned}$$

where again, the subindex  $i$  in the matrices of the previous equalities means that we are referring to the  $i$ -th row of said matrix.

This is not a direct solution to the problem of finding the values of  $\alpha^{-+}$ , but notice that the equality in (4.1.8) means that we have found an operator  $f$  on  $M_{|E^-| \times |E^+|}(\mathbb{R})^4$  such that  $\alpha^{-+} = f(\alpha^{-+})$ , that is,  $\alpha^{-+}$  is a fixed point of said operator. If we were able to show that  $f$  is a contraction, we would be done; however, it seems easier to prove that  $\alpha^{-+}(n)$  converges (entry-wise) to  $\alpha^{-+}$  as  $n \rightarrow \infty$ , where  $\alpha^{-+}(0)$  is some fixed  $|E^-| \times |E^+|$ -matrix and  $\alpha^{-+}(n) := f(\alpha^{-+}(n-1))$  for  $n \in \mathbb{N}$ , so we are doing it this way.

**Definition 4.1.1.** (Order between matrices) From here on, if  $\mathbf{S}$  and  $\mathbf{R}$  are two matrices of the same dimensions, then we say that  $\mathbf{S} \geq \mathbf{R}$  whenever  $\mathbf{S} - \mathbf{R}$  is a matrix with non-negative elements.

Notice that  $\alpha^{+-}$  must have certain properties, such as  $\alpha^{+-} \geq \mathbf{0}$  and  $\alpha^{+-}\mathbf{e} \leq \mathbf{e}$ , so we may work only in the space of  $|E^-| \times |E^+|$ -matrices that attain those properties: let us denote this space by  $\Gamma$ . Then the operator  $f$  restricted to  $\Gamma$  is such that

$$f(\beta)_i = \left( (\mathbf{T}^{-+} + \mathbf{T}^{--}\beta + \Delta_{\lambda/|r|}\beta) \left( \frac{\lambda_i}{|r_i|} \mathbf{I} - (\mathbf{T}^{++} + \mathbf{T}^{+-}\beta) \right)^{-1} \right)_i, \quad (4.1.9)$$

---

<sup>4</sup>That is, the space of matrices of dimension  $|E^-| \times |E^+|$  over  $\mathbb{R}$ .

or rearranging the terms in (4.1.3), we get that

$$f(\boldsymbol{\beta})_i = \left( \sum_{k \in E^+} \int_0^\infty p_{ik} \mathbf{1}'_k e^{(\mathbf{T}^{++} + \mathbf{T}^{+-} \boldsymbol{\beta})x} \frac{\lambda_i}{|r_i|} e^{-\frac{\lambda_i}{|r_i|}x} dx \right. \quad (4.1.10)$$

$$\left. + \sum_{k \in E^- \setminus \{i\}} \int_0^\infty p_{ik} \boldsymbol{\beta}_k e^{(\mathbf{T}^{++} + \mathbf{T}^{+-} \boldsymbol{\beta})x} \frac{\lambda_i}{|r_i|} e^{-\frac{\lambda_i}{|r_i|}x} dx \right)_i,$$

where  $\boldsymbol{\beta}_k := \{\beta_{kj}\}_{j \in E^+}$ .

**Lemma 4.1.1.** *Let  $\boldsymbol{\beta} \in \Gamma$ . Then  $\mathbf{T}^{++} + \mathbf{T}^{+-} \boldsymbol{\beta}$  is an intensity (or sub-intensity) matrix and moreover, the mapping  $\boldsymbol{\beta} \rightarrow \exp((\mathbf{T}^{++} + \mathbf{T}^{+-} \boldsymbol{\beta})x)$  is non-decreasing for each  $x \geq 0$ .*

*Proof.* First, recall that

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}^{++} & \mathbf{T}^{+-} \\ \mathbf{T}^{-+} & \mathbf{T}^{--} \end{pmatrix}$$

is an intensity matrix. Since  $\mathbf{T}^{+-} \geq \mathbf{0}$  and  $\boldsymbol{\beta} \geq \mathbf{0}$ , then  $\mathbf{T}^{+-} \boldsymbol{\beta} \geq \mathbf{0}$ . This implies that  $\mathbf{T}^{++} + \mathbf{T}^{+-} \boldsymbol{\beta}$  is non-negative in its non-diagonal entries. Next, using again that  $\mathbf{T}$  is an intensity matrix, we have that for each  $i \in E^+$ ,

$$0 \geq t_{ii} + \sum_{j \in E^+ \setminus \{i\}} t_{ij} + \sum_{j \in E^-} t_{ij} \geq \sum_{j \in E^+ \setminus \{i\}} t_{ij} + \sum_{j \in E^-} t_{ij} \beta_{ji},$$

proving that the sum of the elements of each row of  $\mathbf{T}^{++} + \mathbf{T}^{+-} \boldsymbol{\beta}$  is not greater than 0. Moreover, it is easy to see that the mapping  $\boldsymbol{\beta} \rightarrow \mathbf{T}^{++} + \mathbf{T}^{+-} \boldsymbol{\beta}$  is non-decreasing.

If  $\boldsymbol{\beta} \in \Gamma$  is such that  $\mathbf{T}^{++} + \mathbf{T}^{+-} \boldsymbol{\beta}$  is an intensity matrix, then any  $\boldsymbol{\beta}' \in \Gamma$  such that  $\boldsymbol{\beta}' \geq \boldsymbol{\beta}$  will render  $\mathbf{T}^{++} + \mathbf{T}^{+-} \boldsymbol{\beta} = \mathbf{T}^{++} + \mathbf{T}^{+-} \boldsymbol{\beta}'$ ; the reason is that the sum of the elements of the  $i$ -th row of  $\mathbf{T}^{++} + \mathbf{T}^{+-} \boldsymbol{\beta}'$  (which is not greater than 0, because of the previous paragraph) is equal to the sum of the elements of the  $i$ -th row of  $\mathbf{T}^{++} + \mathbf{T}^{+-} \boldsymbol{\beta}$  (which is 0, because it is an intensity matrix) plus the sum of the elements of the  $i$ -th row of  $\mathbf{T}^{+-}(\boldsymbol{\beta}' - \boldsymbol{\beta})$ , and since  $\boldsymbol{\beta}' - \boldsymbol{\beta} \geq \mathbf{0}$ , the only valid option is that  $\mathbf{T}^{+-}(\boldsymbol{\beta}' - \boldsymbol{\beta}) = \mathbf{0}$ . In this case, the second statement of the Lemma is trivially true.

Next, we shall study the case when  $\beta \in \Gamma$  renders  $\mathbf{T}^{++} + \mathbf{T}^{+-}\beta$  a sub-intensity matrix, say  $\mathbf{R}$ . Fix  $\beta' \in \Gamma$  such that  $\beta' \geq \beta$  and let  $\mathbf{S} = \mathbf{T}^{++} + \mathbf{T}^{+-}\beta'$ ;  $\mathbf{S}$  must be an intensity or sub-intensity matrix. Let  $\{X_t\}_{t \geq 0}$  be a Markov jump process with sub-intensity matrix  $\mathbf{T}^{++}$  and let  $\{Y_t\}_{t \geq 0}$  be a Markov jump process with sub-intensity matrix  $\mathbf{R}$ : what we are doing next is to construct a Markov jump process  $X'$  which depends on  $X$  and which has the same distribution as  $Y$ . Denote by  $\tau_X$  the time at which the process  $X$  terminates and define the new process  $\{X'_t\}_{t \geq 0}$  such that  $X'_t = X_t$  for all  $t \in [0, \tau_X)$ , and at time  $\tau_X$ ,  $X'$  either effectively terminates or jumps to any state in  $E^+$  (dependant of the last non-cemetery state of  $X$ ; this will be explained later) and continues its path repeating this dynamic: that is, we may have to concatenate a random number of independent terminating copies of  $X$  until an effective termination of  $X'$  exists. Let<sup>5</sup>

$$\mathbb{P}(X'_{\tau_X} = j | X'_{\tau_X^-} = i) := \sum_{k \in E^-} t_{ik} \beta_{kj} / t_{ii}$$

for  $i, j \in E^+$ , and let  $1 - \sum_{i \in E^+} \sum_{k \in E^-} t_{ik} \beta_{kj} / t_{ii}$  be the probability of effectively terminating the process  $X'$  at the moment  $\tau_X$ . It is an easy exercise to verify that the intensity matrix of the process  $X'$  is given by  $\mathbf{T}^{++} + \mathbf{T}^{+-}\beta$ ; that is,  $X'$  has the same distribution as  $Y$ . Consequently we have that

$$\begin{aligned} \mathbb{P}(Y_t = j | Y_0 = i) &= \mathbb{P}(X'_t = j | X'_0 = i) \\ &\geq \mathbb{P}(X'_t = j, \tau_X > t | X'_0 = i) \\ &= \mathbb{P}(X_t = j | X_0 = i); \end{aligned}$$

this implies that  $\exp(\mathbf{T}^{++}) \leq \exp(\mathbf{T}^{++} + \mathbf{T}^{+-}\beta)$ . To prove that  $\exp(\mathbf{T}^{++} + \mathbf{T}^{+-}\beta) \leq \exp(\mathbf{T}^{++} + \mathbf{T}^{+-}\beta')$ , we only need to replace  $\mathbf{T}^{++}$  with  $\mathbf{T}^{++} + \mathbf{T}^{+-}\beta$  and  $\mathbf{T}^{+-}\beta$  with  $\mathbf{T}^{+-}(\beta' - \beta)$  in the previous argumentation.  $\square$

From this lemma and from the equality in (4.1.10) it easily follows that  $f$  is a non-decreasing operator. Now, let  $\alpha^{-+}(0) := \mathbf{0}$ . Since  $\alpha^{-+} \geq \mathbf{0} = \alpha^{-+}(0)$  and  $\alpha^{-+}$  is a fixed point of  $f$ , then we have that

$$\alpha^{-+} = f(\alpha^{-+}) \geq f(\alpha^{-+}(0)) = \alpha^{-+}(1).$$

Thus, by induction it follows that

$$\alpha^{-+} \geq \alpha^{-+}(n) \text{ for all } n \in \mathbb{N}.$$

---

<sup>5</sup>Notice that  $t_{ii} > 0$  for all  $i \in E^+$ , since there are no absorbing states in  $E^+$ .

Since  $f$  is non-decreasing, the sequence  $\{\alpha^{-+}(n)\}_{n \in \mathbb{N}}$  is also non-decreasing, so this sequence must converge (entry-wise) to some  $\gamma \in \Gamma$  such that  $\gamma \leq \alpha^{-+}$ . We must prove that this inequality is actually an equality. To this end, let us define for each  $n \in \mathbb{N} \cup \{0\}$  the event

$$G_n = \{V \text{ up-crosses } 0 \text{ before the } (n+1)\text{-th jump of } J\},$$

and for each  $i \in E^-$  and  $j \in E^+$  define

$$\kappa_{ij}^{-+}(n) := \mathbb{P}(\{V \text{ up-crosses } 0 \text{ while } j \text{ is in } j\} \cap G_n | J_0 = i). \quad (4.1.11)$$

Naturally, let  $\kappa^{-+}(n) := \{\kappa_{ij}^{-+}(n)\}_{i \in E^-, j \in E^+}$ . From (4.1.11), it follows trivially that  $\kappa^{-+}(0) = \mathbf{0}$ , since it is impossible that a fluid-flow up-crosses level 0 before any jump of  $J$  occurred. According to the definition of  $f$  in (4.1.10),  $f(\kappa^{-+}(n))_{ij}$  could be (probabilistically) interpreted in the following way: we wait until the first jump of  $J$  occurs, leaving  $V$  with a negative height which is exponentially distributed of intensity  $\lambda_i/|r_i|$ . Said jump will land in either  $E^+$  or  $E^-$ : let us name them case 1 and case 2, and let us study what happens after this jump.

- Case 1) We let the jump process evolve according the intensity matrix  $\mathbf{T}^{++} + \mathbf{T}^{+-}\kappa^{-+}(n)$  and let it up-cross the (original) level 0 in state  $j$ . The intensity matrix  $\mathbf{T}^{++} + \mathbf{T}^{+-}\kappa^{-+}(n)$  means that the process has two types of jumps: the ones corresponding to the direct jumps from states in  $E^+$  to states in  $E^+$ , and the ones corresponding to the case when a excursion away from the  $V$ -maximum occurs, but this time, we let each excursion have at most  $n$   $J$ -jumps.
- Case 2) We let a excursion away from the  $V$ -maximum of at most  $n$  associated  $J$ -jumps occur, then we let the process evolve with the same intensity matrix as the one explained in case 1, and let it up-cross the (original) level 0 in state  $j$ .

This finalizes the probabilistic interpretation of each element of the matrix  $f(\kappa^{-+}(n))$ .

Notice that any trajectory of the process  $V$  associated to the probability  $\kappa_{ij}^{-+}(n+1)$  is contained in either case 1 or 2 of  $f(\kappa^{-+}(n))$ . This means that

$$\kappa^{-+}(n+1) \leq f(\kappa^{-+}(n)).$$

All this led us to the equation:

$$\begin{aligned}\kappa^{-+}(1) &\leq f(\kappa^{-+}(0)) \\ &\leq f(\alpha^{-+}(0)) \quad (\text{Since } \kappa^{-+}(0) \leq \alpha^{-+}(0) \text{ and } f \text{ is non-decreasing}) \\ &= \alpha^{-+}(1).\end{aligned}$$

Applying  $f$  iteratively to the previous equation leads us to

$$\kappa^{-+}(n) \leq \alpha^{-+}(n) \quad \text{for all } n \in \mathbb{N} \cup \{0\}.$$

By the definition in (4.1.11) we have that  $\kappa^{-+}(n) \uparrow \alpha^{-+}$ , so we can conclude that  $\gamma = \lim_{n \rightarrow \infty} \alpha^{-+}(n) = \alpha^{-+}$ . All this discussion is summarized in the next theorem (we shall use the notation introduced in the previous part).

**Theorem 4.1.1.** *Let  $\{V_t, J_t\}_{t \geq 0}$  be a fluid-flow process such that  $V_0 = 0$  and  $J_0 = i$  for some  $i \in E$ . Then the ascending ladder process  $\{m(x)\}_{x \geq 0}$  is a possibly terminating Markov jump process with (sub-)intensity matrix*

$$U = \mathbf{T}^{++} + \mathbf{T}^{+-} \alpha^{-+},$$

and initial distribution given by  $\mathbf{1}'_i$  if  $i \in E^+$ , or  $\alpha_i^{-+}$  if  $i \in E^-$ . If the process  $\{m(x)\}$  is terminating, then the matrix  $\alpha^{-+}$  can be calculated by

$$\alpha^{-+} = \lim_{n \rightarrow \infty} \alpha^{-+}(n),$$

where  $\alpha^{-+}(0) = \mathbf{0}$ ,  $\alpha^{-+}(n+1) = f(\alpha^{-+}(n))$ , and  $f$  is an operator on  $M_{|E^-| \times |E^+|}(\mathbb{R})$  given by (4.1.9).

**Theorem 4.1.2.** *The iterative scheme proposed in Theorem 4.1.1 can be changed by making  $\alpha^{-+} = \lim_{n \rightarrow \infty} \alpha^{-+}(n)$  where  $\alpha^{-+}(0) = \mathbf{0}$ ,  $\alpha^{-+}(n+1) = g(\alpha^{-+}(n))$  and  $g$  is the operator on  $M_{|E^-| \times |E^+|}(\mathbb{R})$  given by*

$$\begin{aligned}g(\beta) &= (\mathbf{T}^{-+} + \mathbf{T}^{--} \beta + \eta \beta)(\eta \mathbf{I} - \mathbf{U})^{-1} \\ &= (\mathbf{T}^{-+} + \mathbf{T}^{--} \beta + \eta \beta)(\eta \mathbf{I} - (\mathbf{T}^{++} + \mathbf{T}^{+-} \beta))^{-1},\end{aligned}$$

for any  $\eta > \sup_{i \in E^-} \lambda_i / |r_i|$ .

*Proof.* Let us consider the uniformization scheme described in Theorem 3.4.1 for the process  $J$  by taking  $\eta_i = \eta |r_i| > \lambda_i$ . Within that scheme and if  $J_0 = i$ , the next clock rings according to an exponential distribution with

rate  $\eta|r_i|$ , and its transition<sup>6</sup> is governed by the sub-stochastic matrix  $\mathbf{Q} = \mathbf{I} + \mathbf{\Delta}_{\mathbf{1}/(\eta|r_i|)}\boldsymbol{\lambda}$ , where  $\mathbf{\Delta}_{\mathbf{1}/(\eta|r_i|)} = \text{diag}\{1/(\eta|r_i|) : i \in E^-\}$ . This means that the value of  $|V|$  at the time the first clock rings is exponentially distributed with rate  $\eta$ . Then, along the same argumentation given before (4.1.3), we have that

$$\begin{aligned}\alpha_{ij}^{-+} &= \mathbb{P}(V \text{ up-crosses } 0 \text{ while } J \text{ is in the state } j) \\ &= \int_0^\infty \left( \sum_{k \in E^+ \cup E^-} q_{ik} (\mathbf{1}'_k \delta_{E^+}(k) + \boldsymbol{\alpha}_k^{-+} \delta_{E^-}(k)) e^{\mathbf{U}x} \mathbf{1}_j \right) \eta e^{-\eta x} dx \\ &= \sum_{k \in E^+ \cup E^-} \eta \left( \delta_{ik} + \frac{\lambda_{ik}}{\eta|r_i|} \right) (\mathbf{1}'_k \delta_{E^+}(k) + \boldsymbol{\alpha}_k^{-+} \delta_{E^-}(k)) \left( \int_0^\infty e^{(\mathbf{U} - \eta \mathbf{I})x} dx \right) \mathbf{1}_j.\end{aligned}$$

By the same reasons as the ones preceding (4.1.6), we get that

$$\begin{aligned}\boldsymbol{\alpha}_i^{-+} &= \left( \sum_{k \in E^+ \cup E^-} \eta \left( \delta_{ik} + \frac{\lambda_{ik}}{\eta|r_i|} \right) (\mathbf{1}'_k \delta_{E^+}(k) + \boldsymbol{\alpha}_k^{-+} \delta_{E^-}(k)) \right) (\eta \mathbf{I} - \mathbf{U})^{-1} \\ &= \left( \sum_{k \in E^+} \frac{\lambda_{ik}}{|r_i|} \mathbf{1}'_k + \sum_{k \in E^-} \frac{\lambda_{ik}}{|r_i|} \boldsymbol{\alpha}_k^{-+} + \eta \boldsymbol{\alpha}_i^{-+} \right) (\eta \mathbf{I} - \mathbf{U})^{-1};\end{aligned}$$

applying the same arguments given before (4.1.7) we have that

$$\begin{aligned}\boldsymbol{\alpha}_i^{-+} &= ((\mathbf{T}^{-+} + \mathbf{T}^{--} \boldsymbol{\alpha}^{-+} + \eta \boldsymbol{\alpha}^{-+}) (\eta \mathbf{I} - \mathbf{U})^{-1})_i \\ &= \left( (\mathbf{T}^{-+} + \mathbf{T}^{--} \boldsymbol{\alpha}^{-+} + \eta \boldsymbol{\alpha}^{-+}) (\eta \mathbf{I} - (\mathbf{T}^{++} + \mathbf{T}^{+-} \boldsymbol{\alpha}^{-+}))^{-1} \right)_i.\end{aligned}\tag{4.1.12}$$

The rest of the proof is analogous to what we have done before stating Theorem 4.1.1.  $\square$

## 4.2 Risk models.

Consider an insurance company which has an initial capital  $u > 0$ , that continuously receives some premium rate (for now, suppose it is constant and it is 1; if we wish to increase or decrease that rate, we can use time-scaling, which will be explained shortly in this section) and whose positive

<sup>6</sup>Recall that “transitions” are different from “jumps”; see Section 3.4.

claims (say  $Y_1, Y_2, Y_3, \dots$ ) arrive according to some counting process  $\{N_t\}_{t \geq 0}$ . The amount of capital the insurance company having through time is called a **classic risk (or reserve) model**, and we shall denote it by  $\{R_t\}_{t \geq 0}$  where

$$R_t = u + t - \sum_{i=1}^{N_t} Y_i, \quad t \geq 0. \quad (4.2.1)$$

Although we have just defined this model in a very general setting, there exist further extensions, some of which will be discussed in the conclusion of this manuscript.

Sometimes it is more useful (or at least less difficult) to work with the **surplus process** associated to  $\{R_t\}_{t \geq 0}$ , which is denoted by  $\{S_t\}_{t \geq 0}$  and is defined by

$$S_t = u - R_t = -t + \sum_{i=1}^{N_t} Y_i, \quad t \geq 0.$$

Let  $\tau(u) = \inf\{s > 0 : R_s < 0\} = \inf\{s > 0 : S_s > u\}$ . We call  $\tau(u)$  the **time of ruin**. Basically we are interested in (at least) three things:

1.  $\psi(u) := \mathbb{P}(\tau(u) < \infty)$ , which is the probability that  $\{R_t\}$  ever gets ruined,
2.  $\psi(u, H) := \mathbb{P}(\tau(u) < H)$ , which is the probability that that  $\{R_t\}$  gets ruined before some time  $H$ ,
3. the value of  $-R_{\tau(u)}$  (on the event  $\{\tau(u) < \infty\}$ ), which is the negative quantity of money the insurance company has at the moment it is ruined.

These three elements shall be called, respectively, **infinite-horizon probability of ruin** (also called **probability of ruin**), **finite-horizon probability of ruin**, and **severity of ruin**.

**Remark 4.2.1.** In the beginning of this section we mentioned that we would only study the case in which the reserve model  $R$  increases at constant rate 1 between arrivals. Now, suppose we need to model a reserve that increases at some other rate  $c > 0$  and let us denote that model by  $R^c$ ; that is, let

$$R_t^c = u + ct - \sum_{i=1}^{N_t} Y_i.$$

In particular

$$R_{t/c}^c = u + t - \sum_{i=1}^{N_{t/c}} Y_i = u + t - \sum_{i=1}^{N'_t} Y_i, \quad (4.2.2)$$

where  $N' := \{N'_t\}_{t \geq 0}$  is the (accelerated) counting process  $N$  by a rate of  $1/c$  (that is,  $N'_t := N_{t/c}$  for each  $t \geq 0$ ). Clearly the probability of ruin and the distribution of the severity of ruin for the process  $\{R_t^c\}_{t \geq 0}$  coincide with the ones from the process  $\{R_{t/c}^c\}_{t \geq 0}$ , and the finite-horizon probability of ruin before time  $H$  of  $\{R_t^c\}_{t \geq 0}$  is the same as the finite-horizon probability of ruin before time  $H/c$  of  $\{R_{t/c}^c\}_{t \geq 0}$ , so basically we can study  $\{R_{t/c}^c\}_{t \geq 0}$  instead of studying  $\{R_t^c\}_{t \geq 0}$ : this is known as the **time-scaling method**. Furthermore, the form of (4.2.2) is exactly as in (4.2.1), so the process  $\{R_{t/c}^c\}_{t \geq 0}$  is a risk process which increases at a rate of 1 between arrivals: this is the reason why we stated that it was enough to study risk processes that were defined as in (4.2.1).

#### 4.2.1 PH-Cramér-Lundberg process.

The **Cramér-Lundberg process** is a risk model  $\{R_t\}$  as the one described in (4.2.1) with these further characteristics:

1.  $\{N_t\}$  is a Poisson process of intensity  $p > 0$ , and
2.  $Y_1, Y_2, Y_3, \dots$  are i.i.d. random variables which are also independent of  $\{N_t\}$ .

This makes  $\{S_t\}$  and  $\{R_t - u\}$  **Lévy processes**, that is, càdlàg processes with independent and stationary increments. It is a known fact that if  $\{X_t\}$  is a Lévy process with negative jumps, then its Laplace exponent, denoted and defined by  $\phi_X(s) = \log(\mathbb{E}(e^{sX_1}))$ , exists and is finite for every  $s \geq 0$ . In particular the next theorem follows.

**Theorem 4.2.1.** *Let  $R$  be a Cramér-Lundberg process. Then*

$$\phi_{R-u}(s) = s - p + pL_Y(s).$$

*Proof.* See Section 3.3 of Kyprianou [10]. □

The Cramér-Lundberg process has been widely studied, nevertheless, explicit formulae for the probability and severity of ruin have been found in a very few cases.

In our setting, we are further assuming that  $Y_1 \sim PH_{d_1}(\boldsymbol{\delta}, \mathbf{D})$ ; since phase-type distributions are dense in the class of distributions with positive support (see Theorem 9.14 (pp. 183) in Breuer et al. [5]), this sounds like a not-so-restrictive assumption. We will refer to this particular model as the **phase-type Cramér-Lundberg process** (or **PH-Cramér Lundberg process**). By assuming this model we can easily prove the next theorem using probabilistic arguments, which is not possible in the general case. For a proof of it see Section 3.1 in Peralta [15].

**Theorem 4.2.2.** *Let  $R_t$  be a PH-Cramér-Lundberg process. Then:*

1.  $\lambda \mathbb{E}(Y_1) \geq 1 \Leftrightarrow \liminf_t R_t = -\infty$  ; in that case, we have that

$$\mathbb{P}(\tau(u) < \infty) = 1.$$

2. If  $\lambda \mathbb{E}(Y_1) < 1$ , then  $R_t \rightarrow +\infty$  as  $t \rightarrow \infty$ ; in that case, we have that

$$\mathbb{P}(\tau(u) < \infty) = \boldsymbol{\nu} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\nu})u} \mathbf{e}, \text{ and } -R_{\tau(u)} \sim PH_{\mathbf{d}}(\boldsymbol{\nu} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\nu})u}, \mathbf{D}),$$

where  $\boldsymbol{\nu} = -p\boldsymbol{\delta}\mathbf{D}^{-1}$  and  $\mathbf{d} = -\mathbf{D}\mathbf{e}$ .

**Remark 4.2.2.** In most papers and books, the reasoning behind the distribution of the severity of ruin in the phase-type case is not mentioned, since it is fairly easy to guess how it is obtained; however, we are giving a brief explanation of it next. The Markov jump process associated to the claim that ruins the process will be in state  $i$  at the moment of the ruin with probability  $(\boldsymbol{\nu} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\nu})u})_i$ .<sup>7</sup> Then, given that said claim is in state  $i$  at that moment, we just have to let it evolve according to the sub-intensity matrix  $\mathbf{D}$  until absorption happens, which marks the end of the ruining claim and thus its severity is obtained. This same argument will work with further models, so we shall omit it from future proofs.

The finite-horizon probability of ruin is entirely a different story, since we no longer want to compute the distribution of global minimum of  $R$ , so the arguments behind Theorem 4.2.2 do not work: we will come back to this problem in the next chapter.

---

<sup>7</sup>This part also can be found in Section 3.1 of Peralta [15]

### 4.2.2 *PH*-Sparre-Andersen process.

The **Sparre-Andersen process** is a risk model  $\{R_t\}$  as the one described in (4.2.1) with these further characteristics:

1. The successive times between the arrivals in  $\{N_t\}$ , say  $T_1, T_2, T_3, \dots$ , are positive i.i.d. random variables, and
2.  $Y_1, Y_2, Y_3, \dots$  are i.i.d. random variables which are also independent of  $\{T_i\}_{i \in \mathbb{N}}$ .

In this case  $\{N_t\}$  (generally) stops being a Poisson process, but it is still a renewal process. Notice that we can recover the Cramér-Lundberg process if we set the sequence  $\{T_i\}$  to be exponentially distributed.

The explicit probability of ruin is obtained in very rare cases. Once again, we shall use the fact that phase-type distributions are dense amongst the distributions with positive support to argue that it is valid to work in the cases where both sequences  $\{Y_i\}$  and  $\{T_i\}$  are phase-type distributed (using a different distribution for each sequence). Classically, the probability of ruin of this particular case of Sparre-Andersen process, which will be referred as the **phase-type Sparre-Andersen process** (or *PH*-Sparre-Andersen process), is obtained using general random walks theory. Here we shall give an equivalent result using fluid-flow arguments.

Let  $\{S_t\}$  be the surplus process associated to the *PH*-Sparre-Andersen process  $\{R_t\}$  whose claims  $\{Y_i\}$  are  $PH_{d_1}(\boldsymbol{\delta}, \mathbf{D})$  distributed and the times between claims  $\{T_i\}$  are  $PH_{d_2}(\boldsymbol{\rho}, \mathbf{P})$  distributed. Denote the space state associated to each phase-type distribution by  $E^Y$  and  $E^T$ , respectively, which we assume to be disjoint.<sup>8</sup> Let

$$\sigma_n = \sum_{i=1}^n T_i, \quad n \in \mathbb{N},$$

and let  $Z^{(i)} := \{Z_t^{(i)}\}_{t \geq 0}$  and  $W^{(i)} := \{W_t^{(i)}\}_{t \geq 0}$  be the terminating Markov jump processes associated to  $Y_i$  and  $T_i$ , respectively, for each  $i \in \mathbb{N}$ .

---

<sup>8</sup>For example, we can take  $E^Y = \{1, \dots, d_1\}$  and  $E^T = \{d_1 + 1, \dots, d_1 + d_2\}$ . It really does not matter which choice of states we use, as long as  $E^Y \cap E^T = \emptyset$ .

We shall construct a fluid-model  $(J, V)$  in which we can embed the original process  $S$ ; the idea is to construct it in a way such that each upward jump of  $S$  can be identified with a piece-wise deterministic interval in  $V$  with slope 1 such that each one of these piece-wise deterministic intervals has the same length as the height of its associated jump, and such that the times between arrivals in which premium is recollected in  $S$  can be identified with piece-wise deterministic intervals of the same length and slope in  $V$ .<sup>9</sup> Let us explain it carefully by characterizing the process  $J$  first. Let  $E = E^- \cup E^+$ , where  $E^- = E^T$  and  $E^+ = E^Y$ . Then, define  $J$  by making

$$J_t = \left\{ \begin{array}{lll} W_t^{(1)} & \text{for} & t \in [0, \sigma_1) \\ Z_{t-\sigma_1}^{(1)} & \text{for} & t \in [\sigma_1, \sigma_1 + Y_1) \\ W_{t-(\sigma_1+Y_1)}^{(2)} & \text{for} & t \in [\sigma_1 + Y_1, \sigma_2 + Y_1) \\ Z_{t-(\sigma_2+Y_1)}^{(2)} & \text{for} & t \in [\sigma_2 + Y_1, \sigma_2 + Y_1 + Y_2) \\ \vdots & \vdots & \vdots \\ W_{t-(\sigma_{k-1}+\sum_{i=1}^{k-1} Y_i)}^{(k)} & \text{for} & t \in [\sigma_{k-1} + \sum_{i=1}^{k-1} Y_i, \sigma_k + \sum_{i=1}^{k-1} Y_i) \\ Z_{t-(\sigma_k+\sum_{i=1}^{k-1} Y_i)}^{(k)} & \text{for} & t \in [\sigma_k + \sum_{i=1}^{k-1} Y_i, \sigma_k + \sum_{i=1}^k Y_i) \\ \vdots & \vdots & \vdots \end{array} \right.$$

What we just did was constructing a “linear” version of  $S$ , in the sense that we stretched the original process  $N$  by making at each of its arrivals a time-wise insertion of the Markov jump process associated to the corresponding claim. We may refer to these inserted pieces of paths as **artificial insertions** in further pages. Having clarified this, it is easy to see that the transition matrix of the process  $J$  is given by

$$\Lambda = \begin{pmatrix} \mathbf{P} & \mathbf{p}\delta \\ \mathbf{d}\rho & \mathbf{D} \end{pmatrix} =: \begin{pmatrix} \Lambda^{--} & \Lambda^{-+} \\ \Lambda^{+-} & \Lambda^{++} \end{pmatrix},$$

where  $\mathbf{p} = -\mathbf{P}\mathbf{e}$  and  $\mathbf{d} = -\mathbf{D}\mathbf{e}$ . It is also easy to see that its initial distribution is  $(\boldsymbol{\rho}, \mathbf{0})$ .

Let  $V_0 = 0$ . Now that we have given the distributional behaviour of the process  $J$ , to finalize the characterization of  $(J, V)$  it is enough to give the

---

<sup>9</sup>Recall that we are working with the process  $S$ , so the correct slope for these piece-wise deterministic intervals is  $-1$

correct slopes for each state in  $E$ . To do this we only need to take  $r_i = -1$  for  $i \in E^-$ ,  $r_j = 1$  for  $j \in E^+$ . The choice of  $r_i = -1$  for  $i \in E^-$  is to match (in  $V$ ) the original slope between claims of the process  $S$ . Now, recall that the length of each artificial insertion is exactly equal to the height of its corresponding claim, so by taking  $r_j = 1$  for  $j \in E^+$  we are matching the height of the claims and the height of each artificial insertion.<sup>10</sup> This way we have made a one-to-one mapping between the trajectories of  $S$  and  $V$ , so the next theorem follows directly.

**Theorem 4.2.3.**  $\mathbb{P}(\inf R_t < 0)$  coincides with the probability that  $V$  will ever exceed  $u$ . Furthermore, the severity of ruin is  $PH_{d_1}(\boldsymbol{\mu}, \mathbf{D})$  distributed, where

$$\boldsymbol{\mu} := (\mathbb{P}(J \text{ is in state } i \text{ while } V \text{ upcrosses } u) : i \in E^Y).$$

Thus, combining Theorems 4.1.1 and 4.1.2 with the previous theorem, we get the following.

**Theorem 4.2.4.** Using the previous assumptions and notation, the probability of ruin for the  $PH$ -Sparre-Andersen process is

$$\begin{aligned} \psi(u) &= \boldsymbol{\rho}\boldsymbol{\alpha}^{-+} e^{(\boldsymbol{\Lambda}^{++} + \boldsymbol{\Lambda}^{+-} \boldsymbol{\alpha}^{-+})u} \mathbf{e} \\ &= \boldsymbol{\rho}\boldsymbol{\alpha}^{-+} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\rho}\boldsymbol{\alpha}^{-+})u} \mathbf{e}, \end{aligned}$$

where  $\boldsymbol{\alpha}^{-+} = \lim_{n \rightarrow \infty} \boldsymbol{\alpha}^{-+}(n)$ ,  $\boldsymbol{\alpha}^{-+}(0) = \mathbf{0}$ ,  $\boldsymbol{\alpha}^{-+}(n+1) = g(\boldsymbol{\alpha}^{-+}(n))$  and  $g$  is the operator on  $M_{|E^-| \times |E^+|}(\mathbb{R})$  given by

$$\begin{aligned} g(\boldsymbol{\beta}) &= (\boldsymbol{\Lambda}^{-+} + \boldsymbol{\Lambda}^{--}\boldsymbol{\beta} + \eta\boldsymbol{\beta})(\eta\mathbf{I} - (\boldsymbol{\Lambda}^{++} + \boldsymbol{\Lambda}^{+-}\boldsymbol{\beta}))^{-1} \\ &= (\mathbf{p}\boldsymbol{\delta} + \mathbf{P}\boldsymbol{\beta} + \eta\boldsymbol{\beta})(\eta\mathbf{I} - (\mathbf{D} + \mathbf{d}\boldsymbol{\rho}\boldsymbol{\beta}))^{-1} \end{aligned}$$

where  $\eta > -\mathbf{P}_{ii}$  for every  $i \in E^-$ . Moreover,

$$-R_{\tau(u)} \sim PH_{d_1}(\boldsymbol{\rho}\boldsymbol{\alpha}^{-+} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\rho}\boldsymbol{\alpha}^{-+})u}, \mathbf{D}).$$

On later chapters, we will have to deal with the **delayed  $PH$ -Sparre-Andersen process**, which is just a  $PH$ -Sparre-Andersen process whose first arrival has a different distribution from the subsequent ones; we denote this process by  $R^d = \{R_t^d\}_{t \geq 0}$ . The specific case we will encounter in future pages is the next one.

---

<sup>10</sup>We are referring to the height of an artificial insertion as the difference between  $V$  evaluated at the finishing point of the artificial insertion and  $V$  evaluated at the starting point of the artificial insertion.

**Theorem 4.2.5.** *Let  $R^d$  be a delayed PH-Sparre-Andersen process with  $T_1 \sim PH_{d_2}(\boldsymbol{\varrho}, \mathbf{P})$  and whose claims and subsequent interarrival times are  $PH_{d_1}(\boldsymbol{\delta}, \mathbf{D})$  and  $PH_{d_2}(\boldsymbol{\rho}, \mathbf{P})$ -distributed, respectively. Then the probability of ruin of this delayed process, denoted by  $\psi^d(u)$ , is given by*

$$\begin{aligned}\psi^d(u) &= \boldsymbol{\varrho} \boldsymbol{\alpha}^{-+} e^{(\boldsymbol{\Lambda}^{++} + \boldsymbol{\Lambda}^{+-} - \boldsymbol{\alpha}^{-+})u} \mathbf{e} \\ &= \boldsymbol{\varrho} \boldsymbol{\alpha}^{-+} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\rho}\boldsymbol{\alpha}^{-+})u} \mathbf{e},\end{aligned}$$

where  $\boldsymbol{\alpha}^{-+}$  is obtained in the same way as in Theorem 4.2.4. Moreover,

$$-R_{\tau(u)}^d \sim PH_{d_1}(\boldsymbol{\varrho} \boldsymbol{\alpha}^{-+} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\rho}\boldsymbol{\alpha}^{-+})u}, \mathbf{D}).$$

*Proof.* The same arguments as in Theorem 4.2.4 follow, except that in this case the associated fluid-flow model has  $(\boldsymbol{\varrho}, \mathbf{0})$  as initial distribution.  $\square$

**Remark 4.2.3.** Notice that we have not solved the problem for general delayed phase-type Sparre-Andersen process, since the first interarrival time must be phase-type distributed with the same sub-intensity matrix as the other interarrival times. Nevertheless, this special kind of delayed phase-type Sparre-Andersen process is the one that arises in most problems; for example, the stationary phase-type Sparre-Andersen process fits in the setting of Theorem 4.2.5 (see Section 4.1 in Peralta [15]).

## Chapter 5

# Finite-horizon Ruin Probabilities.

Let  $R = \{R_t\}_{t \geq 0}$  be a risk model as the one described in (4.2.1); that is,

$$R_t = u - S_t, \quad t \geq 0,$$

where

$$S_t = \sum_{i=1}^{N_t} Y_i - t,$$

$Y_1, Y_2, \dots$  are positive random variables which represent the size of each claim and  $N = \{N_t\}_{t \geq 0}$  is a counting process. Finding the explicit finite-horizon probability of ruin in this general setting is not possible. Nevertheless, we are presenting a method for finding an approximation of it when  $R$  is a *PH*-Sparre-Andersen process following the next steps: we will find the solution for a phase-type-horizon probability of ruin, using a fluid-flow method similar to the one used in Subsection 4.2.2 which is based on Asmussen et al. [4]. Then, we will be able to use an erlangization argument to give an approximation of probability of ruin before a deterministic time. Later, we will state the same result for *PH*-Cramér-Lundberg processes, which has a much simpler and compact form.

## 5.1 The $PH$ -Sparre-Andersen case.

Let  $\{R_t\}$  be a  $PH$ -Sparre-Andersen process with claims  $Y_1, Y_2, \dots \sim PH_{d_1}(\boldsymbol{\delta}, \mathbf{D})$  and interarrival times  $T_1, T_2, \dots \sim PH_{d_2}(\boldsymbol{\rho}, \mathbf{P})$ . As we said earlier, we wish to calculate

$$\psi(u, H) := \mathbb{P}(\tau(u) < H).$$

The problem of calculating this when  $H$  is a fixed positive number is extremely difficult. The approach we are taking here is to suppose  $H$  is some phase-type distributed random time, independent of  $R$ ; we shall refer to this random variable as the **cemetery clock**. If we let  $H \sim PH_{d_3}(\boldsymbol{\lambda}, \mathbf{L})$ , then we refer to this particular problem as the **phase-type horizon probability of ruin**. We need to be careful, since we are going to be working with three pairs of matrix-parameters; one way to remember which is which, is to note that  $\mathbf{D}$  is associated to the **d**ecrements caused by claims,  $\mathbf{P}$  is associated to times where **p**remium is recollected, and  $\mathbf{L}$  is associated to the random **l**ength at which we wish to analyse if ruin has or has not happened. Also, we shall keep in mind the argumentations behind Theorem 4.2.4 and Theorem 3.2.4 to solve this problem, since we are using both techniques here.

As in Theorem 4.2.4, instead of working with  $R$ , we are going to work with the surplus process  $S = \{S_t\}_{t \geq 0}$ . Suppose that the state spaces  $E^Y$ ,  $E^T$  and  $E^H$  (associated to the claims, interarrival times and cemetery clock) are pairwise disjoint. Let

$$\sigma_n = \sum_{i=1}^n T_i, \quad n \in \mathbb{N} \cup \{0\},$$

let  $Z^{(i)} := \{Z_t^{(i)}\}_{t \geq 0}$  and  $W^{(i)} := \{W_t^{(i)}\}_{t \geq 0}$  be the terminating Markov jump processes associated to  $Y_i$  and  $T_i$ , respectively, for each  $i \in \mathbb{N}$ , and let  $C = \{C_t\}_{t \geq 0}$  be the terminating process associated to the cemetery clock  $H$ . Define  $M := \sup\{n \geq 0 : \sigma_n < H\}$ ; that is,  $M$  is the number of claims before the random time  $H$  occurs.

We shall construct a fluid-flow model  $(J, V)$  in which we can embed the surplus process  $S$ , just as in Theorem 4.2.4, but this time, we want to kill it at the moment the cemetery clock rings. Let us explain it carefully by characterizing the process  $J$  first. Let  $E = E^- \cup E^+ \cup E^0$ , where  $E^- =$

$E^H \times E^T$ ,  $E^+ = E^H \times E^Y$  and  $E^0$  contains only the absorbing state  $\Delta$ . Then define  $J$  by

$$J_t = \left\{ \begin{array}{lll} C_t \times W_t^{(1)} & \text{for} & t \in [0, \sigma_1) \\ C_{\sigma_1} \times Z_{t-\sigma_1}^{(1)} & \text{for} & t \in [\sigma_1, \sigma_1 + Y_1) \\ C_{t-Y_1} \times W_{t-(\sigma_1+Y_1)}^{(2)} & \text{for} & t \in [\sigma_1 + Y_1, \sigma_2 + Y_1) \\ C_{\sigma_2} \times Z_{t-(\sigma_2+Y_1)}^{(2)} & \text{for} & t \in [\sigma_2 + Y_1, \sigma_2 + Y_1 + Y_2) \\ \vdots & \vdots & \vdots \\ C_{t-\sum^{n-1} Y_i} \times W_{t-(\sigma_{n-1}+\sum^{n-1} Y_i)}^{(n)} & \text{for} & t \in [\sigma_{n-1} + \sum^{n-1} Y_i, \sigma_n + \sum^{n-1} Y_i) \\ C_{\sigma_n} \times Z_{t-(\sigma_n+\sum^{n-1} Y_i)}^{(n)} & \text{for} & t \in [\sigma_n + \sum^{n-1} Y_i, \sigma_n + \sum^n Y_i) \\ \vdots & \vdots & \vdots \\ C_{t-\sum^M Y_i} \times W_{t-(\sigma_M+\sum^M Y_i)}^{(M)} & \text{for} & t \in [\sigma_M + \sum^M Y_i, H + \sum^M Y_i) \\ \Delta & \text{for} & t \in [H + \sum^M Y_i, \infty) \end{array} \right. .$$

Notice that this process, unlike the one described in Theorem 4.2.4, has a state space whose components are all pairs; these pairs have on its first entry the process  $C$  which is either evolving or stopped. For example, from 0 until the time of the first arrival, we are letting both  $C$  and  $W^{(1)}$  evolve in a parallel fashion; then, when a  $W^{(1)}$  gets absorbed, we stop the process  $C$  and let the process  $Z^{(1)}$  evolve; when  $Z^{(1)}$  gets absorbed, we start the process  $W^{(2)}$  and let  $C$  continue from the state it was when it was stopped, again, in a parallel fashion; at the instant  $W^{(2)}$  gets absorbed, we stop the process  $C$  and we start the process  $Z^{(2)}$  and so on. Notice that those times where  $C$  is stopped are the ones that were artificially inserted (see Subsection 4.2.2); during these times  $C$ -absorption cannot happen. Actually,  $C$ -absorption will happen while both  $C$  and  $W^{(M)}$  are running, and when this happens, we send the process  $J$  to the cemetery state  $\Delta$ .

Let us order the states in  $E$  by  $E^- = E^H \times E^T$ ,  $E^+ = E^H \times E^Y$  and  $E^0 = \{\Delta\}$ , each one of them sorted linearly in a row-major order (see the proof of Theorem 3.2.4). Then we get that the intensity matrix of the process  $J$  is given by

$$\Lambda = \begin{pmatrix} \mathbf{L} \oplus \mathbf{P} & \mathbf{I}_{d_3} \otimes (\mathbf{p}\boldsymbol{\delta}) & \mathbf{1} \otimes \mathbf{e}_{d_2} \\ \mathbf{I}_{d_3} \otimes (\mathbf{d}\boldsymbol{\rho}) & \mathbf{I}_{d_3} \otimes \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0 \end{pmatrix} =: \begin{pmatrix} \Lambda^{--} & \Lambda^{-+} & \mathbf{1} \otimes \mathbf{e}_{d_3} \\ \Lambda^{+-} & \Lambda^{++} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0 \end{pmatrix},$$

where  $\mathbf{I}_{d_3}$  is the identity matrix of dimension  $d_3$ ,  $\mathbf{e}_{d_2}$  is the column vector of 1's with size  $d_3$ ,  $\mathbf{d} = -\mathbf{D}\mathbf{e}$ ,  $\mathbf{p} = -\mathbf{P}\mathbf{e}$  and  $\mathbf{l} = -\mathbf{L}\mathbf{e}$ . Indeed:

1.  $\mathbf{L} \oplus \mathbf{P}$  corresponds to the parallel evolution of  $C$  and a  $W$ -process (according to the reasoning in the proof of Theorem 3.2.4).
2. The intensity matrix  $\mathbf{p}\boldsymbol{\delta}$  corresponds to the jumps caused by an exit from a  $W$ -process and the beginning of a  $Z$ -process. Now, suppose that  $C$  is in some state  $i \in E^H$  while this jump occurred. Since we want to fix  $C$  in this state  $i$ , we only allow jumps from the states  $\{i\} \times E^T$  to the states in  $\{i\} \times E^Y$ , all this with the transition intensities  $\mathbf{p}\boldsymbol{\delta}$ . This explains the structure of the matrix  $\mathbf{I}_{d_3} \otimes (\mathbf{p}\boldsymbol{\delta})$ .
3.  $\mathbf{D}$  is the transition between states in  $E^Y$ . While this transitions happen, we want to keep  $C$  fixed, and to keep track of which state  $C$  was when the  $Z$ -process started; this explains the structure of  $\mathbf{I}_{d_3} \otimes \mathbf{D}$ .
4. The matrix  $\mathbf{d}\boldsymbol{\rho}$  corresponds to the jumps caused by an exit from a  $Z$ -process and the beginning of a  $W$ -process. Since we already kept track of which state  $C$  was, say  $i \in E^H$ , and now we just want to resume this  $C$ -process, we need to only allow jumps from  $\{i\} \times E^Y$  to the states in  $\{i\} \times E^T$ , all this with the transition intensities  $\mathbf{d}\boldsymbol{\rho}$ . This explains the structure of the matrix  $\mathbf{I}_{d_3} \otimes (\mathbf{d}\boldsymbol{\rho})$ .
5. As we said before, absorption to  $\Delta$  from the  $E^+$  state space is impossible, while absorption to  $\Delta$  from  $E^-$  is only caused by the termination of the process  $C$ , whose intensities are given by  $\mathbf{l}$ . The matrix  $\mathbf{l} \otimes \mathbf{e}_{d_2}$  is just the block-adjusted column vector of intensities.
6. Finally,  $\Delta$  is a cemetery state, so all transition rates corresponding to its row must be 0.

By a similar argument as in Theorem 4.2.4, it is clear that the initial distribution of  $J$  must be given by  $(\boldsymbol{\lambda} \otimes \boldsymbol{\rho}, \mathbf{0}, 0)$ .

Let  $V_0 = 0$ . Now that we have given the distributional behaviour of the process  $J$ , to finalize the characterization of  $(J, V)$  it is enough to give the correct slopes for each state. The solution is to take  $r_i = -1$  for  $i \in E^-$  and  $r_j = 1$  for  $j \in E^+$ ; the argumentation is exactly the same as in Theorem 3.2.4. Nevertheless, we want to ignore further ruin events that might occur

after the cemetery clock rings: what we are going to do is to take  $r_\Delta = 0$ . This way, the process  $V$  will stay fixed on the level  $S_H$  after the clock rings. This means that we have just transformed the problem of checking if  $S$  will up-cross the level  $u$  before some time  $H$ , to the problem of checking if  $V$  will up-cross the level  $u$  at all (in an infinite-horizon sense). Thus, we have the next main result.

**Theorem 5.1.1.**  $\mathbb{P}(\tau(u) < H)$  coincides with the probability that  $V$  will ever exceed  $u$ . Further, the severity of ruin is  $PH_{d_1}(\boldsymbol{\mu}, \mathbf{D})$  distributed, where

$$\boldsymbol{\mu} := (\mathbb{P}(J \text{ is in a state of the form } (\cdot, i) \text{ while } V \text{ upcrosses } u) : i \in E^Y); \quad (5.1.1)$$

Combining Theorem 4.1.1 and Theorem 4.1.1 with the previous theorem, we get the following.

**Theorem 5.1.2.** Using the previous assumptions and notation, the phase-type-horizon probability of ruin for the PH-Sparre-Andersen process is

$$\begin{aligned} \psi(u, H) &= (\boldsymbol{\lambda} \otimes \boldsymbol{\rho}) \boldsymbol{\alpha}^{-+} e^{(\boldsymbol{\Lambda}^{++} + \boldsymbol{\Lambda}^{+-} \boldsymbol{\alpha}^{-+})u} \mathbf{e} \\ &= (\boldsymbol{\lambda} \otimes \boldsymbol{\rho}) \boldsymbol{\alpha}^{-+} e^{(\mathbf{I}_{d_3} \otimes \mathbf{D} + (\mathbf{I}_{d_3} \otimes (\mathbf{d}\boldsymbol{\rho}) \boldsymbol{\alpha}^{-+})u} \mathbf{e}, \end{aligned}$$

where  $\boldsymbol{\alpha}^{-+} = \lim_{n \rightarrow \infty} \boldsymbol{\alpha}^{-+}(n)$ ,  $\boldsymbol{\alpha}^{-+}(0) = \mathbf{0}$ ,  $\boldsymbol{\alpha}^{-+}(n+1) = g(\boldsymbol{\alpha}^{-+}(n))$  and  $g$  is the operator on  $M_{|E^-| \times |E^+|}(\mathbb{R})$  given by

$$\begin{aligned} g(\boldsymbol{\beta}) &= (\boldsymbol{\Lambda}^{-+} + \boldsymbol{\Lambda}^{--} \boldsymbol{\beta} + \eta \boldsymbol{\beta})(\eta \mathbf{I} - (\boldsymbol{\Lambda}^{++} + \boldsymbol{\Lambda}^{+-} \boldsymbol{\beta}))^{-1} \\ &= (\mathbf{I}_{d_3} \otimes (\mathbf{p}\boldsymbol{\delta}) + (\mathbf{L} \oplus \mathbf{P}) \boldsymbol{\beta} + \eta \boldsymbol{\beta})(\eta \mathbf{I} - (\mathbf{I}_{d_3} \otimes \mathbf{D} + \mathbf{I}_{d_3} \otimes (\mathbf{d}\boldsymbol{\rho}) \boldsymbol{\beta}))^{-1} \end{aligned}$$

for  $\eta > \sup_{i \in E^T, j \in E^H} -\mathbf{P}_{ii} - \mathbf{L}_{jj}$ .

**Remark 5.1.1.** Notice that we used that  $(\boldsymbol{\lambda} \otimes \boldsymbol{\rho}) \boldsymbol{\alpha}^{-+} e^{(\boldsymbol{\Lambda}^{++} + \boldsymbol{\Lambda}^{+-} \boldsymbol{\alpha}^{-+})u}$  is the distribution vector of the up-crossings over  $E^H \times E^Y$ , which was ordered linearly in a row-major fashion. Thus, one way to obtain the row vector

$$\begin{aligned} \boldsymbol{\mu} &= (\mathbb{P}(J \text{ is in a state of the form } (\cdot, i) \text{ while } V \text{ upcrosses } u) : i \in E^Y) \\ &= \left( \sum_{l \in E^H} \mathbb{P}(J \text{ is in the state } (l, i) \text{ while } V \text{ upcrosses } u) : i \in E^Y \right) \end{aligned}$$

introduced in (5.1.1) is by making

$$\boldsymbol{\mu} = (\boldsymbol{\lambda} \otimes \boldsymbol{\rho}) \boldsymbol{\alpha}^{-+} e^{(\boldsymbol{\Lambda}^{++} + \boldsymbol{\Lambda}^{+-} \boldsymbol{\alpha}^{-+})u} (\mathbf{e}_{d_3} \otimes \mathbf{I}_{d_1}),$$

where  $\mathbf{I}_{d_1}$  is the identity matrix of dimension  $d_1$  and  $\mathbf{e}_{d_3}$  is the column vector of 1's with size  $d_3$ : it is straightforward to check this fact by making an inspection.

## 5.2 The $PH$ -Cramér-Lundberg case

As we said before, the Cramér Lundberg process is a very special case of the Sparre-Andersen process, and as we have seen in Section 4.2, it is usually a lot easier to work with  $PH$ -Cramér-Lundberg processes than with  $PH$ -Sparre-Andersen processes. In this manuscript we have already studied the phase-type-horizon probability of ruin in the  $PH$ -Sparre-Andersen case, so studying it in the  $PH$ -Cramér-Lundberg case is just a matter of substitution in the final formulae; nevertheless, we can go deeper in the case of Erlang-horizon probability of ruin and get a very simple formula for this probability, which does not require any limiting argument (unlike Theorem 5.1.2).

Let us recall our setting:  $R$  is a risk process where  $\{Y_i\}$  are i.i.d. random variables with distribution  $PH_{d_1}(\boldsymbol{\delta}, \mathbf{D})$ ,  $\{T_i\}$  is an independent sequence of i.i.d. with exponential distribution of rate  $p > 0$ ,<sup>1</sup> and the cemetery clock is  $H \sim PH_{d_3}(\boldsymbol{\lambda}, \mathbf{L})$ , independent from everything else. Then, we have the following.

**Theorem 5.2.1.** *Using the previous assumptions and notation, the phase-type-horizon probability of ruin for the  $PH$ -Cramér-Lundberg process is*

$$\psi(u, H) = \boldsymbol{\lambda} \boldsymbol{\alpha}^{-+} e^{(\mathbf{I}_{d_3} \otimes \mathbf{D} + (\mathbf{I}_{d_3} \otimes \mathbf{d}) \boldsymbol{\alpha}^{-+})u} \mathbf{e}, \quad (5.2.1)$$

where  $\mathbf{d} = -\mathbf{D}\mathbf{e}$ ,  $\boldsymbol{\alpha}^{-+} = \lim_{n \rightarrow \infty} \boldsymbol{\alpha}^{-+}(n)$ ,  $\boldsymbol{\alpha}^{-+}(0) = \mathbf{0}$ ,  $\boldsymbol{\alpha}^{-+}(n+1) = g(\boldsymbol{\alpha}^{-+}(n))$  and  $g$  is the operator on  $M_{|E^-| \times |E^+|}(\mathbb{R})$  given by

$$g(\boldsymbol{\beta}) = (p\mathbf{I}_{d_3} \otimes \boldsymbol{\delta} + (\mathbf{L} - p\mathbf{I}_{d_3})\boldsymbol{\beta} + \eta\boldsymbol{\beta})(\eta\mathbf{I} - (\mathbf{I}_{d_3} \otimes \mathbf{D} + (\mathbf{I}_{d_3} \otimes \mathbf{d})\boldsymbol{\beta}))^{-1}$$

where  $\eta > \sup_{j \in E^H} -\mathbf{L}_{jj} + p$ .

**Remark 5.2.1.** Recall that the recursion proposed in Theorem 5.1.2 came from a fixed point problem whose solution was  $\boldsymbol{\alpha}^{-+}$  (see Theorem 4.1.2 for more details). In other words, we had that

$$\boldsymbol{\alpha}^{-+} = (p\mathbf{I}_{d_3} \otimes \boldsymbol{\delta} + (\mathbf{L} - p\mathbf{I}_{d_3})\boldsymbol{\alpha}^{-+} + \eta\boldsymbol{\alpha}^{-+})(\eta\mathbf{I} - (\mathbf{I}_{d_3} \otimes \mathbf{D} + (\mathbf{I}_{d_3} \otimes \mathbf{d})\boldsymbol{\alpha}^{-+}))^{-1}.$$

<sup>1</sup>Recall that  $Exp(p) = PH_1(1, -p)$ .

Multiplying it by  $\eta \mathbf{I} - (\mathbf{I}_{d_3} \otimes \mathbf{D} + (\mathbf{I}_{d_3} \otimes \mathbf{d})\boldsymbol{\alpha}^{-+})$  to the right, the previous equation is equivalent to

$$\boldsymbol{\alpha}^{-+}(\mathbf{I}_{d_3} \otimes \mathbf{D} + (\mathbf{I}_{d_3} \otimes \mathbf{d})\boldsymbol{\alpha}^{-+}) + (\mathbf{L} - p\mathbf{I}_{d_3})\boldsymbol{\alpha}^{-+} + p\mathbf{I}_{d_3} \otimes \boldsymbol{\delta} = \mathbf{0}. \quad (5.2.2)$$

Having clarified this, it is easier to see that Theorem 5.2.1 is equivalent to the result gotten in Asmussen et al. [4].

### 5.2.1 Erlang case.

Now we are interested in calculating the Erlang-horizon probability of ruin for the *PH*-Cramér-Lundberg process. But first, let us calculate the exponential-horizon probability of ruin, that is, let  $H_1 \sim \text{Exp}(l)$  be the cemetery clock, where  $l > 0$ . In this case, we have that  $\boldsymbol{\alpha}_{(1)}^{-+}$  (defined as the matrix  $\boldsymbol{\alpha}^{-+}$  from Theorem 5.2.1 for this particular case) is a  $E^+ = E^Y$  row vector and (5.2.2) takes the form

$$\boldsymbol{\alpha}_{(1)}^{-+}(\mathbf{D} + \mathbf{d}\boldsymbol{\alpha}_{(1)}^{-+}) + (-l - p)\boldsymbol{\alpha}_{(1)}^{-+} + p\boldsymbol{\delta} = \mathbf{0}. \quad (5.2.3)$$

If we let  $s := l + p - \boldsymbol{\alpha}_{(1)}^{-+}\mathbf{d}$ , then (5.2.3) is equivalent to

$$\boldsymbol{\alpha}_{(1)}^{-+}\mathbf{D} - s\boldsymbol{\alpha}_{(1)}^{-+} + p\boldsymbol{\delta} = \mathbf{0},$$

so that

$$\boldsymbol{\alpha}_{(1)}^{-+} = p\boldsymbol{\delta}(s\mathbf{I}_{d_1} - \mathbf{D})^{-1}. \quad (5.2.4)$$

Now, we only need to compute the value of  $s$ : to do this, just multiply (5.2.4) by  $\mathbf{d}$  to the right, so we get that

$$l + p - s = \boldsymbol{\alpha}_{(1)}^{-+}\mathbf{d} = p\boldsymbol{\delta}(s\mathbf{I}_{d_1} - \mathbf{D})^{-1}\mathbf{d} = pL_Y(s), \quad (5.2.5)$$

where  $L_Y(\cdot)$  is the Laplace transform of  $Y_1$  (see Theorem 3.2.3). Combining (5.2.5) and Theorem 4.2.1 we get that

$$\phi_{R-u}(s) = s - p + pL_Y(s) = l; \quad (5.2.6)$$

that is,  $s > 0$  is the unique root of the equation  $\phi_{R-u}(s) = l^2$ , where  $\phi_{R-u}(\cdot)$  is the Laplace exponent of the Lévy process  $\{R_t - u\}_{t \geq 0}$ . This way, we have

---

<sup>2</sup>Uniqueness and positivity can be verified by Theorem 3.12 in Kyprianou [10].

found a method to calculate  $s$  and therefore  $\alpha_{(1)}^{-+}$ .

Next, consider the case when we let  $H_2 \sim Erl(2, l)$  be the cemetery clock and let  $\alpha_{(2)}^{-+}$  be defined as the matrix  $\alpha^{-+}$  from Theorem 5.2.1 for this particular setting. Then  $\alpha_{(2)}^{-+}$  is a  $2 \times 2d_1$  matrix whose  $(i, j)$ -th entry (with  $i \in E^- = E^{H_2}$  and  $j \in E^+ = E^H \times E^Y$ ) represents the probability that the associated fluid-flow process up-crosses 0 while being at state  $j$ , given that it started in state  $i$ . Hence

$$\alpha_{(2)}^{-+} = \begin{pmatrix} \alpha^{(1)} & \alpha^{(2)} \\ \mathbf{0} & \alpha^{(1)} \end{pmatrix}, \quad (5.2.7)$$

where  $\alpha^{(1)} := \alpha_{(1)}^{-+}$  and  $\alpha^{(2)}$  is (currently) unknown. Indeed, since the phase-type representation of  $H_2$  consists of two subsequent transient states, say  $1_H$  and  $2_H$ , then there exist 4 cases:

1.  $i = 1_H$  and  $j = (1_H, \cdot)$ : this means that from the time the process started until it up-crossed level 0, there has not been any jumps in the phases of  $H_2$ , that is, the exponential clock that marks the jump has not rung, and that is why this probabilities are given by the vector  $\alpha_{(1)}^{-+}$ ;
2.  $i = 2_H$  and  $j = (2_H, \cdot)$ : this also means that there has not been any jumps until the up-crossing of level 0, so this probabilities are also given by the vector  $\alpha_{(1)}^{-+}$ ;
3.  $i = 1_H$  and  $j = (2_H, \cdot)$ : this means that there was one (and only one) jump from  $1_H$  to  $2_H$  which happened somewhere between the beginning of the process and the moment it up-crossed level 0. For now we do not know anything about this vector of probabilities, that is why we just named it  $\alpha^{(2)}$ ;
4.  $i = 2$  and  $j = (1_H, \cdot)$ : since the associated process behind  $H_2$  does not allow jumps from  $2_H$  to  $1_H$ , then this vector of probabilities must be  $\mathbf{0}$ .

To completely calculate  $\alpha_{(2)}^{-+}$ , it is necessary that we calculate the row vector  $\alpha^{(2)}$ . To do this, we shall use formula (5.2.2), which in this setting takes the

form

$$\begin{aligned}
\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} &= \begin{pmatrix} \boldsymbol{\alpha}^{(1)} & \boldsymbol{\alpha}^{(2)} \\ \mathbf{0} & \boldsymbol{\alpha}^{(1)} \end{pmatrix} \left( \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} + \begin{pmatrix} \mathbf{d} & \mathbf{0} \\ \mathbf{0} & \mathbf{d} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}^{(1)} & \boldsymbol{\alpha}^{(2)} \\ \mathbf{0} & \boldsymbol{\alpha}^{(1)} \end{pmatrix} \right) \\
&+ \begin{pmatrix} -l-p & l \\ 0 & -l-p \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}^{(1)} & \boldsymbol{\alpha}^{(2)} \\ \mathbf{0} & \boldsymbol{\alpha}^{(1)} \end{pmatrix} + p \begin{pmatrix} \boldsymbol{\delta} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\delta} \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{\alpha}^{(1)} & \boldsymbol{\alpha}^{(2)} \\ \mathbf{0} & \boldsymbol{\alpha}^{(1)} \end{pmatrix} \begin{pmatrix} \mathbf{d}\boldsymbol{\alpha}^{(1)} + \mathbf{D} & \mathbf{d}\boldsymbol{\alpha}^{(2)} \\ \mathbf{0} & \mathbf{d}\boldsymbol{\alpha}^{(1)} + \mathbf{D} \end{pmatrix} \\
&+ \begin{pmatrix} -l-p & l \\ 0 & -l-p \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}^{(1)} & \boldsymbol{\alpha}^{(2)} \\ \mathbf{0} & \boldsymbol{\alpha}^{(1)} \end{pmatrix} + p \begin{pmatrix} \boldsymbol{\delta} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\delta} \end{pmatrix};
\end{aligned}$$

next, computing only the upper-right block of the previous equality we get that

$$\begin{aligned}
\mathbf{0} &= \boldsymbol{\alpha}^{(1)}\mathbf{d}\boldsymbol{\alpha}^{(2)} + \boldsymbol{\alpha}^{(2)}(\mathbf{d}\boldsymbol{\alpha}^{(1)} + \mathbf{D}) + (-l-p)\boldsymbol{\alpha}^{(2)} + l\boldsymbol{\alpha}^{(1)} \\
&= \boldsymbol{\alpha}^{(2)}((-l-p + \boldsymbol{\alpha}^{(1)}\mathbf{d})\mathbf{I} + \mathbf{d}\boldsymbol{\alpha}^{(1)} + \mathbf{D}) + l\boldsymbol{\alpha}^{(1)}. \tag{5.2.8}
\end{aligned}$$

Recall that  $s := l + p - \boldsymbol{\alpha}_{(1)}^{-+}\mathbf{d} = l + p - \boldsymbol{\alpha}^{(1)}\mathbf{d}$ , so (5.2.8) implies that

$$\boldsymbol{\alpha}^{(2)}(-s\mathbf{I} + \mathbf{d}\boldsymbol{\alpha}^{(1)} + \mathbf{D}) = -l\boldsymbol{\alpha}^{(1)}.$$

Following an argumentation similar to Theorem 3.3.1, one can easily see that  $\mathbf{d}\boldsymbol{\alpha}^{(1)} + \mathbf{D}$  is an invertible sub-intensity matrix, so by Theorem 3.2.2 we get that  $-s\mathbf{I} + \mathbf{d}\boldsymbol{\alpha}^{(1)} + \mathbf{D}$  is an invertible matrix. This means that

$$\boldsymbol{\alpha}^{(2)} = l\boldsymbol{\alpha}^{(1)}(s\mathbf{I} - \mathbf{d}\boldsymbol{\alpha}^{(1)} - \mathbf{D})^{-1},$$

and so, the case when  $H_2$  is the cemetery clock is completely solved since we have successfully found a way for computing  $\boldsymbol{\alpha}_{(2)}^{-+}$  explicitly.

Finally, let us examine what happens when  $H_n \sim Erl(n, l)$  is the cemetery clock. Then by similar argumentation we have that  $\boldsymbol{\alpha}_{(n)}^{-+}$  (defined as the matrix  $\boldsymbol{\alpha}^{-+}$  from Theorem 5.2.1 for this particular setting) has the form

$$\boldsymbol{\alpha}_{(n)}^{-+} = \begin{pmatrix} \boldsymbol{\alpha}^{(1)} & \boldsymbol{\alpha}^{(2)} & \boldsymbol{\alpha}^{(3)} & \dots & \boldsymbol{\alpha}^{(n)} \\ \mathbf{0} & \boldsymbol{\alpha}^{(1)} & \boldsymbol{\alpha}^{(2)} & \dots & \boldsymbol{\alpha}^{(n-1)} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\alpha}^{(1)} & \dots & \boldsymbol{\alpha}^{(n-2)} \\ \vdots & \vdots & \vdots & \ddots & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\alpha}^{(1)} \end{pmatrix}. \tag{5.2.9}$$

By plugging this matrix in (5.2.2) and computing the upper-right block of that equality, we get that

$$\begin{aligned} \mathbf{0} &= \sum_{i=1}^n \alpha^{(i)} \mathbf{d}\alpha^{(n-i+1)} + \alpha^{(n)} \mathbf{D} + (-l - p)\alpha^{(n)} + l\alpha^{(n-1)} \\ &= \sum_{i=2}^{n-1} \alpha^{(i)} \mathbf{d}\alpha^{(n-i+1)} + \alpha^{(n)}(-s\mathbf{I} + \mathbf{d}\alpha^{(1)} + \mathbf{D}) + l\alpha^{(n-1)}, \end{aligned}$$

so

$$\alpha^{(n)} = \left( \sum_{i=2}^{n-1} \alpha^{(i)} \mathbf{d}\alpha^{(n-i+1)} + l\alpha^{(n-1)} \right) (s\mathbf{I} - \mathbf{d}\alpha^{(1)} - \mathbf{D})^{-1};$$

all this can be summarized in the next result.

**Theorem 5.2.2.** *Let  $R$  be the PH-Cramér-Lundberg described in Theorem 5.2.1 and  $H_n \sim \text{Erl}(n, l)$ . Then the  $H_n$ -horizon probability of ruin for  $R$  is given by the formula (5.2.1), where*

$$\alpha_{(n)}^{-+} = \begin{pmatrix} \alpha^{(1)} & \alpha^{(2)} & \alpha^{(3)} & \dots & \alpha^{(n)} \\ \mathbf{0} & \alpha^{(1)} & \alpha^{(2)} & \dots & \alpha^{(n-1)} \\ \mathbf{0} & \mathbf{0} & \alpha^{(1)} & \dots & \alpha^{(n-2)} \\ \vdots & \vdots & \vdots & \ddots & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \alpha^{(1)} \end{pmatrix},$$

$\alpha^{(1)} := \alpha_{(1)}^{-+}$  is given by (5.2.4) and  $\alpha^{(k)}$  can be computed recursively by the formula

$$\alpha^{(k)} = \left( \sum_{i=2}^{k-1} \alpha^{(i)} \mathbf{d}\alpha^{(k-i+1)} + l\alpha^{(k-1)} \right) (s\mathbf{I} - \mathbf{d}\alpha^{(1)} - \mathbf{D})^{-1},$$

where  $s$  is the unique positive number that meets the condition  $\phi_{R-u}(s) = l$ .

## Chapter 6

# Parisian Ruin Probabilities.

Let  $R = \{R_t\}$  be a risk model as the one described in (4.2.1). We have previously defined ruin as the event in which the process  $R$  down-crosses level 0 in some finite time; for the sake of clarity, we shall refer to it as **classic ruin**. In this chapter, we will define another type of ruin, the so-called **parisian ruin** which we will explain next. Let us define a sequence of (possibly random) times  $\{\zeta_i\}_{i \in \mathbb{N}}$ , which we call **parisian clocks**. We attach these clocks to the beginning of each excursion below zero that  $R$  might possibly make. Then we declare  $R$  to be ruined in a parisian way iff at least one of those clocks rings before its associated excursion below 0 ends. More specifically, let  $\theta_0 = \iota_0 = 0$  and define for  $i \in \mathbb{N}$

$$\begin{aligned}\theta_i &:= \inf\{s > \iota_{i-1} : R_s < 0\}, \text{ and} \\ \iota_i &:= \inf\{s > \theta_i : R_s > 0\}.\end{aligned}$$

Then the sequence  $\{\theta_i\}$  corresponds to the successive times the process  $R$  starts an excursion below 0, and the sequence  $\{\iota_i\}$  corresponds to the successive times in which these excursions ended. Let

$$N := \inf\{i \in \mathbb{N} : \zeta_i < \iota_i - \theta_i\}.$$

Then, according to the previous explanation, parisian ruin occurs iff  $N < \infty$ , and on this case, we define the time of parisian ruin to be  $\theta_N + \zeta_N$ .

There is not a known method for calculating parisian ruin probabilities for a general risk model. However, we present one when we assume that  $R$  is a *PH*-Sparre-Andersen process and the parisian clocks  $\{\zeta_i\}$  are phase-type

i.i.d. random variables, independent from the  $PH$ -Sparre-Andersen process. To do this, we will use a fluid-flow argument, so we need to keep Section 4.1 in mind. Later, we will present the same result when  $R$  is a  $PH$ -Cramér-Lundberg process, which has a much simpler form than the  $PH$ -Sparre-Andersen case. After this, with analogous computations as the ones made in Section 5.2, we will get an even easier method to compute the probability of a parisian ruin for a  $PH$ -Cramér-Lundberg process whenever the parisian clocks are Erlang-distributed: with this in mind we would be able to apply an erlangization argument to approximate the probability of a parisian ruin with fixed deterministic clocks.

## 6.1 The $PH$ -Sparre-Andersen case.

Let  $\{R_t\}$  be a  $PH$ -Sparre-Andersen process with claims  $Y_1, Y_2, \dots \sim PH_{d_1}(\boldsymbol{\delta}, \mathbf{D})$  and interarrival times  $T_1, T_2, \dots \sim PH_{d_2}(\boldsymbol{\rho}, \mathbf{P})$ . Let the parisian clocks  $\{\zeta_i\}$  be i.i.d. with distribution  $PH_{d_3}(\boldsymbol{\kappa}, \mathbf{K})$ , independent from everything else. Suppose that the state spaces  $E^Y$ ,  $E^T$  and  $E^\zeta$  (associated to the claims, interarrival times and parisian clocks) are pairwise disjoint. Let

$$\sigma_n := \sum_{i=1}^n T_i, \quad n \in \mathbb{N} \cup \{0\},$$

let  $Z^{(i)} := \{Z_t^{(i)}\}_{t \geq 0}$ ,  $W^{(i)} := \{W_t^{(i)}\}_{t \geq 0}$  and  $O^{(i)} := \{O_t^{(i)}\}_{t \geq 0}$  be the terminating Markov jump processes associated to  $Y_i$ ,  $T_i$  and  $\zeta_i$ , respectively, for each  $i \in \mathbb{N}$ .

### 6.1.1 Recovery probability.

Consider the case in which the reserve process starts on some level  $-x < 0$ ; this means that we immediately let the clock  $\zeta_1$  start from time 0 and we check if the process  $R$  can **recover** (that is, if it can up-cross level 0) before  $\zeta_1$  rings: in this subsection we will compute the probability of this happening. Notice that checking for a recovery is the same problem as letting  $R_0 = 0$  and checking if the process  $R$  can up-cross the level  $x$  before the clock  $\zeta_1$  rings: this is the setting we will work with, since it is easier to work with a process that start at 0. With this in mind, there is an obvious analogy of the work we made in Section 5.1, since there we also had a clock running;<sup>1</sup> nevertheless, there we were interested in the down-crossings of  $R$ , and here we are interested in the up-crossings, so we need to make some adjustments.

Following the idea in Section 5.1, we need to construct a fluid-flow model  $(J, V)$  in which we can embed the  $PH$ -Sparre-Andersen process  $R$  killed at the moment  $\zeta_1$  rings. To explain it carefully, we need to characterize the process  $J$  first. Let

$$N_1 = \sup\{i \in \mathbb{N} \cup \{0\} : \sigma_i < \zeta_1\},$$

that is, let  $N_1$  be the number of claim arrivals before  $\zeta_1$  rings. Let  $E = E^- \cup E^+ \cup E^0$ , where  $E^- = E^\zeta \times E^Y$ ,  $E^+ = E^\zeta \times E^T$  and  $E^0$  only contains

---

<sup>1</sup>The phase-type horizon clock.

the absorbing state  $\Delta$ . Then define  $J$  by

$$J_t = \left\{ \begin{array}{lll} O_t^{(1)} \times W_t^{(1)} & \text{for} & t \in [0, \sigma_1) \\ O_{\sigma_1}^{(1)} \times Z_{t-\sigma_1}^{(1)} & \text{for} & t \in [\sigma_1, \sigma_1 + Y_1) \\ O_{t-Y_1}^{(1)} \times W_{t-(\sigma_1+Y_1)}^{(2)} & \text{for} & t \in [\sigma_1 + Y_1, \sigma_2 + Y_1) \\ O_{\sigma_2}^{(1)} \times Z_{t-(\sigma_2+Y_1)}^{(2)} & \text{for} & t \in [\sigma_2 + Y_1, \sigma_2 + Y_1 + Y_2) \\ \vdots & \vdots & \vdots \\ O_{t-\sum^{n-1} Y_i}^{(1)} \times W_{t-(\sigma_{n-1}+\sum^{n-1} Y_i)}^{(n)} & \text{for} & t \in [\sigma_{n-1} + \sum^{n-1} Y_i, \sigma_n + \sum^{n-1} Y_i) \\ O_{\sigma_n}^{(1)} \times Z_{t-(\sigma_n+\sum^{n-1} Y_i)}^{(n)} & \text{for} & t \in [\sigma_n + \sum^{n-1} Y_i, \sigma_n + \sum^n Y_i) \\ \vdots & \vdots & \vdots \\ O_{t-\sum^{N_1} Y_i}^{(1)} \times W_{t-(\sigma_{N_1}+\sum^{N_1} Y_i)}^{(N_1)} & \text{for} & t \in [\sigma_{N_1} + \sum^{N_1} Y_i, \zeta_1 + \sum^{N_1} Y_i) \\ \Delta & \text{for} & t \in [\zeta_1 + \sum^{N_1} Y_i, \infty) \end{array} \right. .$$

As in Section 5.1, this Markov jump process has a state space whose components are all pairs; these pairs have on its first entry the process  $O^{(1)}$  which is either evolving or stopped. From 0 until  $\sigma_1$ ,  $O^{(1)}$  and  $W^{(1)}$  evolve in a parallel fashion; when a  $W^{(1)}$  gets absorbed, we stop the process  $O^{(1)}$  and let the process  $Z^{(1)}$  evolve; when  $Z^{(1)}$  gets absorbed, we start the process  $W^{(2)}$  and let  $O^{(1)}$  continue from the state it was when it was stopped, again, in a parallel fashion; at the instant  $W^{(2)}$  gets absorbed, we stop the process  $O^{(1)}$  and we start the process  $Z^{(2)}$  and so on. Notice that those times where  $O^{(1)}$  is stopped are the ones that were artificially inserted (see Subsection 4.2.2); during these times the termination of  $O^{(1)}$  cannot happen.  $O^{(1)}$ -absorption happens while both  $O^{(1)}$  and  $W^{(N_1)}$  are running, and when this happens, we send the process  $J$  to the cemetery state  $\Delta$ .

Order the states in  $E$  by  $E^+ = E^H \times E^T$ ,  $E^- = E^H \times E^Y$  and  $E^0 = \{\Delta\}$ , each one of them sorted linearly in a row-major order.<sup>2</sup> By the exact same arguments as the ones given in Section 5.1 we get that the intensity matrix of the process  $J$  is given by

$$\Lambda = \begin{pmatrix} \mathbf{K} \oplus \mathbf{P} & \mathbf{I}_{d_3} \otimes (\mathbf{p}\delta) & \mathbf{k} \otimes \mathbf{e}_{d_2} \\ \mathbf{I}_{d_3} \otimes (\mathbf{d}\rho) & \mathbf{I}_{d_3} \otimes \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0 \end{pmatrix} =: \begin{pmatrix} \Lambda^{++} & \Lambda^{+-} & \mathbf{k} \otimes \mathbf{e}_{d_3} \\ \Lambda^{-+} & \Lambda^{--} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0 \end{pmatrix},$$

<sup>2</sup>See the proof of Theorem 3.2.4.

where  $\mathbf{I}_{d_3}$  is the identity matrix of dimension  $d_3$ ,  $\mathbf{e}_{d_2}$  is the column vector of 1's with size  $d_3$ ,  $\mathbf{d} = -\mathbf{D}\mathbf{e}$ ,  $\mathbf{p} = -\mathbf{P}\mathbf{e}$  and  $\mathbf{k} = -\mathbf{K}\mathbf{e}$ . Notice that the matrices  $\mathbf{\Lambda}^{++}$ ,  $\mathbf{\Lambda}^{+-}$ ,  $\mathbf{\Lambda}^{--}$  and  $\mathbf{\Lambda}^{-+}$  have switched places (with respect how they were in Section 5.1); this is because here we are working directly with the reserve process, while in Section 5.1 we were working with the surplus process. By the same argumentation as in Section 5.1, the initial distribution of  $J$  is given by  $(\boldsymbol{\kappa} \otimes \boldsymbol{\rho}, \mathbf{0}, 0)$ .

Let  $V_0 = 0$ . To conclude with the characterization of  $(J, V)$  it is enough to give the correct slopes for each state in  $E$ . The solution is to take  $r_i = -1$  for  $i \in E^-$  and  $r_j = 1$  for  $j \in E^+$ ; the argumentation is exactly the same as in Theorem 3.2.4. Now, we want to ignore further possible up-crossings of level  $x$  after  $\zeta_1$  has rung; we do this by taking  $r_\Delta = 0$ . This way, the process  $V$  will stay fixed on the level  $R_{\zeta_1}$  after the parisian clock rings. Thus, we have transformed the problem of checking if  $R$  has up-crossed the level  $x$  before  $\zeta_1$ , to the problem of checking if  $V$  has up-crossed the level  $x$  at all (in an infinite-horizon sense). This implies the next theorem.

**Theorem 6.1.1.**  $\mathbb{P}_0(R \text{ up-crosses level } x \text{ before } \zeta_1 \text{ rings})$  coincides with the probability that  $V$  will ever exceed  $x$ . Further, on the event it up-crossed the level  $x$  before  $\zeta_1$  rang, the time until the next claim arrives is  $PH_{d_2}(\boldsymbol{\nu}, \mathbf{P})$  distributed, where

$$\boldsymbol{\nu} := (\mathbb{P}(J \text{ is in a state of the form } (\cdot, i) \text{ while } V \text{ upcrosses } x) : i \in E^T). \quad (6.1.1)$$

Combining Theorem 4.1.1 and Theorem 4.1.2 with the previous theorem, we get the following.

**Theorem 6.1.2.** Using the previous assumptions and notation, the recovery probability of a PH-Sparre-Andersen process that starts in some level  $-x < 0$ , on the event that this recovery is smaller than the parisian clock  $\zeta_1$ , is given by

$$\begin{aligned} \varphi(x) &= (\boldsymbol{\kappa} \otimes \boldsymbol{\rho}) e^{(\mathbf{\Lambda}^{++} + \mathbf{\Lambda}^{+-} - \boldsymbol{\gamma}^{-+})x} \mathbf{e} \\ &= (\boldsymbol{\kappa} \otimes \boldsymbol{\rho}) e^{(\mathbf{K} \oplus \mathbf{P} + \mathbf{I}_{d_3} \otimes (\mathbf{p}\boldsymbol{\delta})\boldsymbol{\gamma}^{-+})x} \mathbf{e}, \end{aligned}$$

where  $\boldsymbol{\gamma}^{-+} = \lim_{n \rightarrow \infty} \boldsymbol{\gamma}^{-+}(n)$ ,  $\boldsymbol{\gamma}^{-+}(0) = \mathbf{0}$ ,  $\boldsymbol{\gamma}^{-+}(n+1) = g(\boldsymbol{\gamma}^{-+}(n))$  and

$g$  is the operator on  $M_{|E^-| \times |E^+|}(\mathbb{R})$  given by

$$\begin{aligned} g(\boldsymbol{\beta}) &= (\boldsymbol{\Lambda}^{-+} + \boldsymbol{\Lambda}^{--}\boldsymbol{\beta} + \eta\boldsymbol{\beta})(\eta\mathbf{I} - (\boldsymbol{\Lambda}^{++} + \boldsymbol{\Lambda}^{+-}\boldsymbol{\beta}))^{-1} \\ &= (\mathbf{I}_{d_3} \otimes (\mathbf{d}\boldsymbol{\rho}) + \mathbf{I}_{d_3} \otimes \mathbf{D}\boldsymbol{\beta} + \eta\boldsymbol{\beta})(\eta\mathbf{I} - (\mathbf{K} \oplus \mathbf{P} + \mathbf{I}_{d_3} \otimes (\mathbf{p}\boldsymbol{\delta})\boldsymbol{\beta}))^{-1} \end{aligned}$$

where  $\eta > \sup_{i \in E^T, j \in E^c} -\mathbf{P}_{ii} - \mathbf{K}_{jj}$ .

**Remark 6.1.1.** As in Section 5.1, the vector

$$\begin{aligned} \boldsymbol{\nu} &= (\mathbb{P}(J \text{ is in a state of the form } (\cdot, i) \text{ while } V \text{ upcrosses } x) : i \in E^T) \\ &= \left( \sum_{l \in E^c} \mathbb{P}(J \text{ is in a state of the form } (\cdot, i) \text{ while } V \text{ upcrosses } x) : i \in E^T \right) \end{aligned}$$

introduced in (6.1.1) can be obtained explicitly by making

$$\boldsymbol{\nu} = (\boldsymbol{\kappa} \otimes \boldsymbol{\rho})\boldsymbol{\gamma}^{-+} e^{(\mathbf{K} \oplus \mathbf{P} + \mathbf{I}_{d_3} \otimes (\mathbf{p}\boldsymbol{\delta})\boldsymbol{\gamma}^{-+})x} (\mathbf{e}_{d_3} \otimes \mathbf{I}_{d_2}),$$

where  $\mathbf{I}_{d_1}$  is the identity matrix of dimension  $d_2$  and  $\mathbf{e}_{d_3}$  is the column vector of 1's with size  $d_3$ .

## 6.1.2 Main result

In Subsection 4.2.2 we calculated the distribution of the severity of ruin for a phase-type Sparre-Andersen process. In the previous Subsection we calculated the probability of recovery before  $\zeta_1$  rings, given that  $R_0 = -x < 0$ . Thus, integrating over the severity of ruin, we get that  $\varrho_1(u)$ , defined by

$$\varrho_1(u) := \mathbb{P}(\theta_1 < \infty, \iota_1 - \theta_1 < \zeta_1 | R_0 = u)$$

has the analytical form

$$\begin{aligned} \varrho_1(u) &= \int_0^\infty \mathbb{P}(\theta_1 < \infty, \iota_1 - \theta_1 < \zeta_1 | R_{\theta_1} = -x) \mathbb{P}(-R_{\theta_1} \in dx) \\ &= \int_0^\infty ((\boldsymbol{\kappa} \otimes \boldsymbol{\rho}) e^{(\mathbf{K} \oplus \mathbf{P} + (\mathbf{I}_{d_3} \otimes (\mathbf{p}\boldsymbol{\delta}))\boldsymbol{\gamma}^{-+})x} \mathbf{e}_{d_3 \times d_2}) (\boldsymbol{\rho}\boldsymbol{\alpha}^{-+} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\rho}\boldsymbol{\alpha}^{-+})u} e^{\mathbf{D}x} \mathbf{d}) dx \\ &= ((\boldsymbol{\kappa} \otimes \boldsymbol{\rho}) \otimes (\boldsymbol{\rho}\boldsymbol{\alpha}^{-+} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\rho}\boldsymbol{\alpha}^{-+})u})) \\ &\quad \times \left( \int_0^\infty e^{((\mathbf{K} \oplus \mathbf{P} + (\mathbf{I}_{d_3} \otimes (\mathbf{p}\boldsymbol{\delta}))\boldsymbol{\gamma}^{-+}) \oplus \mathbf{D})x} dx \right) (\mathbf{e}_{d_3 \times d_2} \otimes \mathbf{d}) \\ &= -((\boldsymbol{\kappa} \otimes \boldsymbol{\rho}) \otimes (\boldsymbol{\rho}\boldsymbol{\alpha}^{-+} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\rho}\boldsymbol{\alpha}^{-+})u})) \\ &\quad \times ((\mathbf{K} \oplus \mathbf{P} + (\mathbf{I}_{d_3} \otimes (\mathbf{p}\boldsymbol{\delta}))\boldsymbol{\gamma}^{-+}) \oplus \mathbf{D})^{-1} (\mathbf{e}_{d_3 \times d_2} \otimes \mathbf{d}), \end{aligned}$$

where all the matrices, vectors and notation were taken from Subsection 4.2.2 and Subsection 6.1.1.<sup>3</sup> Moreover, we are capable of calculating

$$\boldsymbol{q}_1 := \left( \mathbb{P} \left( \begin{array}{l} \text{The corresponding } W\text{-process was in} \\ \text{state } i \text{ while the first recovery happened} \end{array} \right) : i \in E^T \right)$$

by making use of fluid-flow arguments, so that

$$\begin{aligned} \boldsymbol{q}_1 &= \int_0^\infty ((\boldsymbol{\kappa} \otimes \boldsymbol{\rho}) e^{(\mathbf{K} \oplus \mathbf{P} + (\mathbf{I}_{d_3} \otimes (\mathbf{p}\boldsymbol{\delta}))\boldsymbol{\gamma}^{-+})x} (\mathbf{e}_{d_3} \otimes \mathbf{I}_{d_2}) (\boldsymbol{\rho}\boldsymbol{\alpha}^{-+} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\rho}\boldsymbol{\alpha}^{-+})u} e^{\mathbf{D}x} \mathbf{d}) dx \\ &= -((\boldsymbol{\kappa} \otimes \boldsymbol{\rho}) \otimes (\boldsymbol{\rho}\boldsymbol{\alpha}^{-+} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\rho}\boldsymbol{\alpha}^{-+})u})) \\ &\quad \times ((\mathbf{K} \oplus \mathbf{P} + (\mathbf{I}_{d_3} \otimes (\mathbf{p}\boldsymbol{\delta}))\boldsymbol{\gamma}^{-+}) \oplus \mathbf{D})^{-1} ((\mathbf{e}_{d_3} \otimes \mathbf{I}_{d_2}) \otimes \mathbf{d}) : \end{aligned} \quad (6.1.2)$$

indeed,  $((\boldsymbol{\kappa} \otimes \boldsymbol{\rho}) e^{(\mathbf{K} \oplus \mathbf{P} + (\mathbf{I}_{d_3} \otimes (\mathbf{p}\boldsymbol{\delta}))\boldsymbol{\gamma}^{-+})x}$  is the distribution vector of

$$\{J \text{ is in state } i \text{ while } V \text{ upcrossed level } 0 \text{ in the first recovery} : i \in E^+\}$$

conditional to given  $\{R_{\theta_1} = -x\}$ . By multiplying to the right said vector by  $\mathbf{e}_{d_3} \otimes \mathbf{I}_{d_2}$  we recover the distribution vector of

$$\left\{ \begin{array}{l} \text{The corresponding } W\text{-process was in} \\ \text{state } i \text{ while the first recovery happened} \end{array} : i \in E^T \right\}$$

conditional to the event  $\{R_{\theta_1} = -x\}$ .

Now that we have the probability of the first recovery, all we need to do is to decompose the whole trajectory of  $R$  in recoveries of this kind. That is, after we recovered for the first time, we wish to calculate the probability of recovering a second time. The calculation in (6.1.2) is of great importance to accomplish this, since the moment at which  $R$  recovers (for the first time) is an instant between arrivals, so it is not possible to use a renewal argument. What we are doing to calculate the probability of a second recovery is to start

---

<sup>3</sup>Also, several identities of Kronecker product and Kronecker sum were used, which can be summarized into the next result:

$$\int_0^\infty (\mathbf{A} e^{\mathbf{B}x} \mathbf{C}) \otimes (\mathbf{D} e^{\mathbf{E}x} \mathbf{F}) dx = (\mathbf{A} \otimes \mathbf{D}) (\mathbf{B} \oplus \mathbf{E})^{-1} (\mathbf{C} \otimes \mathbf{F})$$

whenever  $\mathbf{B}$  and  $\mathbf{E}$  are invertible sub-intensity matrices. See Appendix 1 in Peralta [15] for details.

a delayed  $PH$ -Sparre-Andersen process, since at the moment of the (first) up-crossing, the time until the next arrival has distribution  $PH_{d_2}(\varrho_1, \mathbf{P})$ ; once this claim arrives, the next ones will arrive according to the original  $PH_{d_2}(\boldsymbol{\rho}, \mathbf{P})$  distribution. In Theorem 4.2.5 we learned to deal with this kind of delayed  $PH$ -Sparre-Andersen process, so by an analogous argumentation as the one made there, we get the next result.

**Theorem 6.1.3.** *For every  $n \geq 1$ , let*

$$\varrho_n := \mathbb{P}(\theta_n < \infty, \iota_1 - \theta_1 < \zeta_1, \dots, \iota_n - \theta_n < \zeta_n | R_0 = u),$$

that is,  $\varrho_n(u)$  is the probability that  $R$  recovered from its  $n$  first classic ruins, and each of this recoveries took no more than the length of its associated parisian clock. Also, let

$$\boldsymbol{\varrho}_n := \left( \mathbb{P} \left( \begin{array}{l} \text{The corresponding } W\text{-process was in} \\ \text{state } i \text{ while the } n\text{-th recovery happened} \end{array} \right) : i \in E^T \right), \quad (6.1.3)$$

in a row vector fashion. Then

$$\varrho_n = -(\mathbf{A} \otimes (\boldsymbol{\varrho}_{n-1} \mathbf{B})) \mathbf{C}_1, \quad (6.1.4)$$

and

$$\boldsymbol{\varrho}_n = -(\mathbf{A} \otimes (\boldsymbol{\varrho}_{n-1} \mathbf{B})) \mathbf{C}_2$$

where

$$\mathbf{A} := \boldsymbol{\kappa} \otimes \boldsymbol{\rho},$$

$$\mathbf{B} := \boldsymbol{\alpha}^{-+} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\rho}\boldsymbol{\alpha}^{-+})u},$$

$$\mathbf{C}_1 := ((\mathbf{K} \oplus \mathbf{P} + (\mathbf{I}_{d_3} \otimes (\mathbf{p}\boldsymbol{\delta}))\boldsymbol{\gamma}^{-+}) \oplus \mathbf{D})^{-1} (\mathbf{e}_{d_3 \times d_2} \otimes \mathbf{d}),$$

$$\mathbf{C}_2 := ((\mathbf{K} \oplus \mathbf{P} + (\mathbf{I}_{d_3} \otimes (\mathbf{p}\boldsymbol{\delta}))\boldsymbol{\gamma}^{-+}) \oplus \mathbf{D})^{-1} ((\mathbf{e}_{d_3} \otimes \mathbf{I}_{d_2}) \otimes \mathbf{d}), \text{ and}$$

$$\boldsymbol{\varrho}_0 := \boldsymbol{\rho}.$$

Let  $n \geq 1$ . We need to compute the probability that, given that  $R$  underwent  $n$  recoveries without being parisian ruined, no further classic ruins will occur: let  $\varphi_n$  be that conditional probability and let us work within that conditioned space. From the definition in (6.1.3), we know that the first interarrival time after the  $n$ -th recovery is  $PH_{d_2}((1/\boldsymbol{\varrho}_n \mathbf{e})\boldsymbol{\varrho}_n, \mathbf{P})$ -distributed.<sup>4</sup> Also, at the time of recovery, the process  $R$  will be at the level

---

<sup>4</sup>Notice that here we need to normalize the initial distribution  $\varrho_n$ , since it is conditional to the event that  $R$  up-crossed 0 (at the  $n$ -th recovery), the time of the next arrival must not have an atom at 0.

0. Thus  $\varphi_n$  corresponds to the probability of ruin not happening in a delayed  $PH$ -Sparre-Andersen which starts at level 0, whose first interarrival time is  $PH_{d_2}((1/\varrho_n \mathbf{e})\varrho_n, \mathbf{P})$ -distributed, and whose claims and subsequent interarrival times are  $PH_{d_1}(\boldsymbol{\delta}, \mathbf{D})$  and  $PH_{d_2}(\boldsymbol{\rho}, \mathbf{P})$ -distributed, respectively. Then, according to Theorem 4.2.5, we get that

$$\varphi_n = 1 - \frac{1}{\varrho_n \mathbf{e}} \varrho_n \boldsymbol{\alpha}^{-+} e^{(\mathbf{D} + \mathbf{d} \boldsymbol{\rho} \boldsymbol{\alpha}^{-+})^0 \mathbf{e}} \quad (6.1.5)$$

$$= 1 - \frac{1}{\varrho_n \mathbf{e}} \varrho_n \boldsymbol{\alpha}^{-+} \mathbf{e}. \quad (6.1.6)$$

Thus, if  $E_n := \{R \text{ undergoes } n \text{ recoveries without being parisian ruined}\}$ , then

$$\begin{aligned} & \mathbb{P}(\text{No parisian ruin, } \theta_n < \infty, \theta_n = \infty) \\ &= \mathbb{P} \left( \begin{array}{l} \text{The whole process } R \text{ undergoes } \mathbf{exactly} \\ n \text{ recoveries without being parisian ruined} \end{array} \right) \\ &= \mathbb{P}(E_n) \mathbb{P}(R \text{ does not gets classically ruined after the } n\text{-th recovery} | E_n) \\ &= \varrho_n \varphi_n \end{aligned}$$

**Theorem 6.1.4.** *Let  $\psi^p(u)$  be the probability of parisian ruin. Then*

$$\psi^p(u) = \psi(u) - \sum_{n=1}^{\infty} \varrho_n \varphi_n,$$

where  $\psi(u)$  is the classic probability of ruin (described in Theorem 4.2.4),  $\varrho_n$  is given by (6.1.4) and  $\varphi_n$  is given by (6.1.6).

*Proof.* Notice that

$$\begin{aligned} 1 - \psi^p(u) &= \mathbb{P}_u(\text{There is no parisian ruin for } R) \\ &= \mathbb{P}_u(\text{There is no classic ruin for } R) \\ &\quad + \sum_{n=1}^{\infty} \mathbb{P}(\text{No parisian ruin, } \theta_n < \infty, \theta_n = \infty) \\ &= 1 - \psi(u) + \sum_{n=1}^{\infty} \varrho_n \varphi_n, \end{aligned}$$

so the result follows. □

## 6.2 The $PH$ -Cramér-Lundberg case.

Just as in Section 5.2, parisian ruin probability for the  $PH$ -Cramér-Lundberg case becomes easier than in the  $PH$ -Sparre-Andersen. What we ought to do is to let the interarrival times  $\{T_i\}$  be exponentially distributed of parameter  $p$ . By simple substitution in the equations of Theorem 6.1.2, we have the next result concerning the recovery probability of a Cramér-Lundberg process.

**Theorem 6.2.1.** *Using the assumptions and notation in Subsection 4.2.1, the recovery probability of the Cramér-Lundberg process  $R$  that starts in some level  $-x < 0$ , on the event that this recovery is smaller than the parisian clock  $\zeta_1$ , is given by*

$$\varphi(x) = \kappa e^{(\mathbf{K} - p\mathbf{I}_{d_3} + p(\mathbf{I}_{d_3} \otimes \boldsymbol{\delta})\boldsymbol{\gamma}^{-+})x} \mathbf{e},$$

where  $\boldsymbol{\gamma}^{-+} = \lim_{n \rightarrow \infty} \boldsymbol{\gamma}^{-+}(n)$ ,  $\boldsymbol{\gamma}^{-+}(0) = \mathbf{0}$ ,  $\boldsymbol{\gamma}^{-+}(n+1) = g(\boldsymbol{\gamma}^{-+}(n))$  and  $g$  is the operator on  $M_{|E^-| \times |E^+|}(\mathbb{R})$  given by

$$g(\boldsymbol{\beta}) = (\mathbf{I}_{d_3} \otimes \mathbf{d} + (\mathbf{I}_{d_3} \otimes \mathbf{D})\boldsymbol{\beta} + \eta\boldsymbol{\beta})(\eta\mathbf{I} - (\mathbf{K} - p\mathbf{I}_{d_3} + p(\mathbf{I}_{d_3} \otimes \boldsymbol{\delta})\boldsymbol{\beta}))^{-1}$$

where  $\mathbf{d} = -\mathbf{D}\mathbf{e}$  and  $\eta > \sup_{j \in E^c} -\mathbf{K}_{jj} + p$ .

**Remark 6.2.1.** As in Theorem 5.2.1, the recursion proposed in Theorem 6.2.1 comes from a fixed point problem whose solution is  $\boldsymbol{\gamma}^{-+}$ . In other words, we had that

$$\boldsymbol{\gamma}^{-+} = (\mathbf{I}_{d_3} \otimes \mathbf{d} + (\mathbf{I}_{d_3} \otimes \mathbf{D})\boldsymbol{\gamma}^{-+} + \eta\boldsymbol{\gamma}^{-+})(\eta\mathbf{I} - (\mathbf{K} - p\mathbf{I}_{d_3} + p(\mathbf{I}_{d_3} \otimes \boldsymbol{\delta})\boldsymbol{\gamma}^{-+}))^{-1}. \quad (6.2.1)$$

Multiplying (6.2.1) by  $\eta\mathbf{I} - (\mathbf{K} - p\mathbf{I}_{d_3} + p(\mathbf{I}_{d_3} \otimes \boldsymbol{\delta})\boldsymbol{\gamma}^{-+})$  to the right we get that

$$\boldsymbol{\gamma}^{-+}(\mathbf{K} - p\mathbf{I}_{d_3} + p(\mathbf{I}_{d_3} \otimes \boldsymbol{\delta})\boldsymbol{\gamma}^{-+}) + (\mathbf{I}_{d_3} \otimes \mathbf{D})\boldsymbol{\gamma}^{-+} + \mathbf{I}_{d_3} \otimes \mathbf{d} = \mathbf{0}; \quad (6.2.2)$$

we will work with this equation later.

Now, the distribution of the severity of ruin for a phase-type  $PH$ -Cramér-Lundberg process was given in Theorem 4.2.2, and we have calculated the probability of recovery from any point  $-x < 0$ . Thus, integrating over the severity of ruin, we get that  $\varrho_C(u)$ , defined by

$$\varrho_C(u) := \mathbb{P}(\theta_1 < \infty, \iota_1 - \theta_1 < \zeta_1 | R_0 = u)$$

has the analytical form

$$\begin{aligned}
\varrho_C(u) &= \int_0^\infty \mathbb{P}(\theta_1 < \infty, \iota_1 - \theta_1 < \zeta_1 | R_{\theta_1} = -x) \mathbb{P}(-R_{\theta_1} \in dx) \\
&= \int_0^\infty (\boldsymbol{\kappa} e^{(\mathbf{K} - p\mathbf{I}_{d_3} + p(\mathbf{I}_{d_3} \otimes \boldsymbol{\delta})\boldsymbol{\gamma}^{-+})x} \mathbf{e}_{d_3}) (\boldsymbol{\nu} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\nu})u} e^{\mathbf{D}x} \mathbf{d}) dx \\
&= (\boldsymbol{\kappa} \otimes (\boldsymbol{\nu} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\nu})u})) \left( \int_0^\infty e^{((\mathbf{K} - p\mathbf{I}_{d_3} + p(\mathbf{I}_{d_3} \otimes \boldsymbol{\delta})\boldsymbol{\gamma}^{-+}) \oplus \mathbf{D})x} dx \right) (\mathbf{e}_{d_3} \otimes \mathbf{d}) \\
&= -(\boldsymbol{\kappa} \otimes (\boldsymbol{\nu} e^{(\mathbf{D} + \mathbf{d}\boldsymbol{\nu})u})) \left( (\mathbf{K} - p\mathbf{I}_{d_3} + p(\mathbf{I}_{d_3} \otimes \boldsymbol{\delta})\boldsymbol{\gamma}^{-+}) \oplus \mathbf{D} \right)^{-1} (\mathbf{e}_{d_3} \otimes \mathbf{d})
\end{aligned} \tag{6.2.3}$$

where all the matrices, vectors and notation were taken from Theorem 6.2.1 and Theorem 4.2.2.

As we mentioned in Subsection 4.2.1, the *PH*-Crámer-Lundberg process is a Lévy process. Using the strong Markov property for Lévy processes, we get that on the event  $\iota_m < \infty$  ( $m \geq 2$ ),  $\{(R_{t-\iota_{i-1}} : t \in [\iota_{i-1}, \iota_i])\}_{i=2}^m$  are independent and identically distributed processes. Moreover, each one of them is equal in distribution to the process  $(R_t : t \in [0, \iota_1])$  on the event  $R_0 = 0$  and  $\iota_1 < \infty$ . In other words, the decomposition used for the *PH*-Sparre-Andersen process works better in the *PH*-Crámer-Lundberg setting, since at each time  $\iota_i$ , we can work with a restarted *PH*-Cramér-Lundberg process which starts at level 0 and each one of these recoveries will be equally distributed. This idea results in the next theorem.

**Theorem 6.2.2.** *Let  $\psi_C^p(u)$  be the probability of parisian ruin for the Crámer-Lundberg process  $R$  that starts at level  $u$ . Then*

$$\psi^p(u) = \psi(u) - \frac{\varrho_C(u)(1 - \psi(0))}{1 - \varrho_C},$$

where  $\psi(u)$  is the classic probability of ruin computed in Theorem 4.2.2,  $\varrho^C(u)$  is given by (6.2.3) and  $\varrho_C := \varrho_C(0)$ .

*Proof.* Notice that

$$\begin{aligned}
1 - \psi^p(u) &= \mathbb{P}_u(\text{There is no parisian ruin for } R) \\
&= \mathbb{P}_u(\text{There is no classic ruin for } R) \\
&\quad + \sum_{n=1}^{\infty} \mathbb{P}(\text{No parisian ruin, } \theta_n < \infty, \theta_n = \infty) \\
&= 1 - \psi(u) + \sum_{n=1}^{\infty} \varrho_C(u) \varrho_C^{n-1}(1 - \psi(0)) \\
&= \psi(u) - \frac{\varrho_C(u)(1 - \psi(0))}{1 - \varrho_C},
\end{aligned}$$

so the result follows.  $\square$

Clearly, the result for the *PH*-Cramér-Lundberg case is far easier than the result for the *PH*-Sparre-Andersen one, since for the former there is no need to compute any kind of series.

### 6.2.1 Erlang case.

Now we are interested in calculating the parisian ruin probability for the *PH*-Cramér-Lundberg process with Erlang distributed parisian clocks. To do this, first let us study the case when the parisian clocks are exponentially distributed; that is, let  $\zeta_1 \sim \text{Exp}(k)$  where  $k > 0$ . In this case, we have that  $\gamma_{(1)}^{-+}$  (defined as the matrix  $\gamma^{-+}$  from Theorem 6.2.1 for this particular setting) is a  $E^+ = E^Y$  column vector and (6.2.2) takes the form

$$\gamma_{(1)}^{-+}(-k - p + p\delta\gamma_{(1)}^{-+}) + \mathbf{D}\gamma_{(1)}^{-+} + \mathbf{d} = \mathbf{0}. \quad (6.2.4)$$

If we let

$$z := k + p - p\delta\gamma_{(1)}^{-+}, \quad (6.2.5)$$

then (6.2.4) is equivalent to

$$(\mathbf{D} - z\mathbf{I}_{d_1})\gamma_{(1)}^{-+} + \mathbf{d} = \mathbf{0}, \quad (6.2.6)$$

so that

$$\gamma_{(1)}^{-+} = (z\mathbf{I}_{d_1} - \mathbf{D})^{-1}\mathbf{d}. \quad (6.2.7)$$

Now, we only need to compute the value of  $z$ : to do this, just multiply (6.2.7) by  $p\delta$  to the right, so we get that

$$k + p - z = p\delta\gamma_{(1)}^{-+} = p\delta(z\mathbf{I}_{d_1} - \mathbf{D})^{-1}\mathbf{d} = pL_Y(z),$$

where  $L_Y(\cdot)$  is the Laplace transform of  $Y_1$  (see Theorem 3.2.3). Just as in (5.2.6) we get that

$$\phi_{R-u}(z) = z - p + pL_Y(z) = k;$$

that is,  $z > 0$  is the unique root of the equation  $\phi_{R-u}(z) = k$  and this way, we have found a method to calculate  $z$  and  $\gamma_{(1)}^{-+}$ .

Next, consider the case when we let  $\zeta_1 \sim \text{Erl}(2, p)$  and let  $\gamma_{(2)}^{-+}$  be defined as the matrix  $\gamma^{-+}$  from Theorem 6.2.1 for this particular setting. Then  $\gamma_{(2)}^{-+}$  is a  $2d_1 \times 2$  matrix whose  $(i, j)$ -th entry (with  $i \in E^- = E^\zeta \times E^Y$  and  $j \in E^+ = E^\zeta$ ) represents the probability that the associated fluid-flow process up-crosses 0 while being at state  $j$ , given that it started in state  $i$ . Hence

$$\gamma_{(2)}^{-+} = \begin{pmatrix} \gamma^{(1)} & \gamma^{(2)} \\ \mathbf{0} & \gamma^{(1)} \end{pmatrix},$$

where  $\gamma^{(1)} := \gamma_{(1)}^{-+}$  and  $\gamma^{(2)}$  is (currently) unknown. The explanation for said structure is completely analogous to (5.2.7).

To completely calculate  $\gamma_{(2)}^{-+}$ , it is necessary that we calculate the column vector  $\gamma^{(2)}$ . To do so, we shall use formula (6.2.2), which in this setting takes the form

$$\begin{aligned} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} &= \begin{pmatrix} \gamma^{(1)} & \gamma^{(2)} \\ \mathbf{0} & \gamma^{(1)} \end{pmatrix} \left( \begin{pmatrix} -k & k \\ 0 & -k \end{pmatrix} - \begin{pmatrix} p & 0 \\ 0 & p \end{pmatrix} + p \begin{pmatrix} \delta & \mathbf{0} \\ \mathbf{0} & \delta \end{pmatrix} \begin{pmatrix} \gamma^{(1)} & \gamma^{(2)} \\ \mathbf{0} & \gamma^{(1)} \end{pmatrix} \right) \\ &+ \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \gamma^{(1)} & \gamma^{(2)} \\ \mathbf{0} & \gamma^{(1)} \end{pmatrix} + \begin{pmatrix} \mathbf{d} & \mathbf{0} \\ \mathbf{0} & \mathbf{d} \end{pmatrix} \\ &= \begin{pmatrix} \gamma^{(1)} & \gamma^{(2)} \\ \mathbf{0} & \gamma^{(1)} \end{pmatrix} \begin{pmatrix} -k - p + p\delta\gamma^{(1)} & k + p\delta\gamma^{(2)} \\ 0 & -k - p + p\delta\gamma^{(1)} \end{pmatrix} \\ &+ \begin{pmatrix} \mathbf{D}\gamma^{(1)} & \mathbf{D}\gamma^{(2)} \\ \mathbf{0} & \mathbf{D}\gamma^{(1)} \end{pmatrix} + \begin{pmatrix} \mathbf{d} & \mathbf{0} \\ \mathbf{0} & \mathbf{d} \end{pmatrix}; \end{aligned}$$

next, computing only the upper-right block of the previous equality we get that

$$\begin{aligned}\mathbf{0} &= (k + p\delta\gamma^{(2)})\gamma^{(1)} + (-k - p + p\delta\gamma^{(1)})\gamma^{(2)} + \mathbf{D}\gamma^{(2)} \\ &= k\gamma^{(1)} + (\mathbf{D} + p\gamma^{(1)}\delta - z\mathbf{I}_{d_1})\gamma^{(2)},\end{aligned}$$

where  $z$  was defined in (6.2.5). This means that

$$\gamma^{(2)} = k(-\mathbf{D} - p\gamma^{(1)}\delta + z\mathbf{I}_{d_1})^{-1}\gamma^{(1)},$$

and so, the case when the parisian clock are  $Erl(2, k)$ -distributed is completely solved, since we have found a way for explicitly computing the matrix  $\gamma_{(2)}^{-+}$ .

Finally, let us examine what happens when  $\zeta_1 \sim Erl(n, k)$ . Analogous to (5.2.9) we have that  $\gamma_{(n)}^{-+}$  (defined as the matrix  $\gamma^{-+}$  from Theorem 6.2.1 for this particular setting) has the form

$$\gamma_{(n)}^{-+} = \begin{pmatrix} \gamma^{(1)} & \gamma^{(2)} & \gamma^{(3)} & \dots & \gamma^{(n)} \\ \mathbf{0} & \gamma^{(1)} & \gamma^{(2)} & \dots & \gamma^{(n-1)} \\ \mathbf{0} & \mathbf{0} & \gamma^{(1)} & \dots & \gamma^{(n-2)} \\ \vdots & \vdots & \vdots & \ddots & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \gamma^{(1)} \end{pmatrix}.$$

By plugging this matrix in (6.2.2) and computing the upper left block of that equality, we get that

$$\begin{aligned}\mathbf{0} &= \sum_{i=1}^{n-1} (k + p\delta\gamma^{(n-i+1)})\gamma^{(i)} + (-k - p + p\delta\gamma^{(1)})\gamma^{(n)} + \mathbf{D}\gamma^{(n)} \\ &= k\gamma^{(1)} + \sum_{i=2}^{n-1} (k + p\delta\gamma^{(n-i+1)})\gamma^{(i)} + (\mathbf{D} + p\gamma^{(1)}\delta - z\mathbf{I}_{d_1})\gamma^{(n)},\end{aligned}$$

so

$$\gamma^{(n)} = (\mathbf{D} + p\gamma^{(1)}\delta - z\mathbf{I}_{d_1})^{-1} \left( k\gamma^{(1)} + \sum_{i=2}^{n-1} (k + p\delta\gamma^{(n-i+1)})\gamma^{(i)} \right);$$

all this can be summarized in the next result.

**Theorem 6.2.3.** *Let  $R$  be the PH-Cramér-Lundberg described in Subsection 4.2.1 and  $\{\zeta_i\}$  be Erl( $n, k$ )-distributed. Then the parisian probability of ruin for  $R$  is given by the formula (5.2.1), where*

$$\gamma_{(n)}^{-+} = \begin{pmatrix} \gamma^{(1)} & \gamma^{(2)} & \gamma^{(3)} & \dots & \gamma^{(n)} \\ \mathbf{0} & \gamma^{(1)} & \gamma^{(2)} & \dots & \gamma^{(n-1)} \\ \mathbf{0} & \mathbf{0} & \gamma^{(1)} & \dots & \gamma^{(n-2)} \\ \vdots & \vdots & \vdots & \ddots & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \gamma^{(1)} \end{pmatrix},$$

$\gamma^{(1)} = \gamma_{(1)}^{-+}$  which is given by (6.2.7) and  $\gamma^{(j)}$  can be computed recursively by the formula

$$\gamma^{(j)} = (\mathbf{D} + p\gamma^{(1)}\boldsymbol{\delta} - z\mathbf{I}_{d_1})^{-1} \left( k\gamma^{(1)} + \sum_{i=2}^{j-1} (k + p\boldsymbol{\delta}\gamma^{(j-i+1)})\gamma^{(i)} \right),$$

where  $z$  is the unique positive number that meets the condition  $\phi_{R-u}(z) = k$ .

# Conclusions and future work.

In addition to the review made to the existing literature, we have successfully given a method for computing the parisian probability of ruin for the *PH*-Crámer-Lundberg and *PH*-Sparre-Andersen process, and even found an efficient way for computing it for the *PH*-Crámer-Lundberg process with Erlang distributed clocks, so that erlangization can be carried on quite easily.

At the moment two extensions are being considered:

- The parisian ruin probability for MArP-driven<sup>5</sup> risk processes. Apparently, this extension can be carried without problem to that model without much effort, although computational effort increases considerably.
- The parisian ruin probability for spectrally negative Lévy processes. This one is far more analytical than the previous one, since the aid of functional calculus is needed.

A paper considering the previous models and the ones of Chapter 6 is in the works.

---

<sup>5</sup>MArP are de initials for Markovian Arrival Process, not to be confused with Markov Additive Process.

# Bibliography

- [1] David Aldous and Larry Shepp. The least variable phase type distribution is erlang. *Stochastic Models*, 3(3):467–473, 1987.
- [2] Charalambos D. Aliprantis and Kim C. Border. *Infinite dimensional analysis*. Springer, Berlin, third edition, 2006. A hitchhiker’s guide.
- [3] Søren Asmussen. Stationary distributions for fluid flow models with or without Brownian noise. *Comm. Statist. Stochastic Models*, 11(1):21–49, 1995.
- [4] Søren Asmussen, Florin Avram, and Miguel Usabel. Erlangian approximations for finite-horizon ruin probabilities. *Astin Bull.*, 32(2):267–281, 2002.
- [5] L. Breuer and D. Baum. *An introduction to queueing theory and matrix-analytic methods*. Springer, Dordrecht, 2005.
- [6] Angelos Dassios and Shanle Wu. Parisian ruin with exponential claims. 2008.
- [7] Winfried Grassmann. Transient solutions in Markovian queues; an algorithm for finding them and determining their waiting-time distributions. *European J. Oper. Res.*, 1(6):396–402, 1977.
- [8] J. Hoffmann-Jørgensen. Probability in Banach Spaces. In P.-L. Hennequin, editor, *Ecole d’Eté de Probabilités de Saint-Flour VI-1976*, volume 598 of *Lecture Notes in Mathematics*, pages 1–186. Springer Berlin Heidelberg, 1977.
- [9] R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit. *Modern Actuarial Risk Theory: Using R*. SpringerLink : Bücher. Springer, 2008.

- [10] Andreas E. Kyprianou. *Fluctuations of Lévy processes with applications*. Universitext. Springer, Heidelberg, second edition, 2014. Introductory lectures.
- [11] Alfred Müller and Ludger Rüschendorf. On the optimal stopping values induced by general dependence structures. *J. Appl. Probab.*, 38(3):672–684, 09 2001.
- [12] James R. Munkres. *Topology: a first course*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975.
- [13] Colm Art O’Cinneide. Phase-type distributions and majorization. *The Annals of Applied Probability*, pages 219–227, 1991.
- [14] K.R. Parthasarathy. *Probability Measures on Metric Spaces*. AMS Chelsea Publishing Series. Academic Press, 1972.
- [15] Oscar Peralta. Modelos de reserva con reclamaciones tipo fase y matriz-exponencial. Bachelor thesis, Facultad de Ciencias UNAM, México DF, 2013.
- [16] Walter Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.
- [17] Walter Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition, 1991.
- [18] Volker Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, pages 423–439, 1965.
- [19] Flemming Topsøe. Compactness in spaces of measures. *Studia Math.*, 36:195–222. (errata insert), 1970.
- [20] Flemming Topsøe. *Topology and measure*. Lecture notes in mathematics. Springer-Verlag, 1970.
- [21] Stephen Willard. *General topology*. Dover Publications, Inc., Mineola, NY, 2004. Reprint of the 1970 original [Addison-Wesley, Reading, MA; MR0264581].