

# Cours de gestion des risques d'assurances et de théorie de la ruine

Stéphane Loisel

ISFA, 2005-2006

# Table des matières

<b>I Modélisation de la charge sinistre : du modèle individuel au modèle collectif</b>	<b>5</b>
<b>1 Modèle individuel</b>	<b>7</b>
<b>2 Modèle collectif</b>	<b>9</b>
2.1 Modèle collectif . . . . .	9
2.2 Lois utilisées pour le nombre de sinistres . . . . .	9
2.3 Lois composées . . . . .	10
2.4 Rappels sur les transformées de Laplace, et les fonctions génératrices . . . . .	11
2.4.1 Définitions et premières propriétés . . . . .	11
2.4.2 Moments, fonctions génératrices, transformée de Laplace . . . . .	13
2.4.3 Injectivité et inversion . . . . .	16
2.4.4 Indépendance et caractérisation de l'indépendance . . . . .	17
2.4.5 Lois usuelles . . . . .	19
2.4.6 Lois composées . . . . .	19
2.5 Famille et algorithme de Panjer . . . . .	20
2.5.1 Etude et caractérisation des distributions vérifiant la relation de récurrence de Panjer . . . . .	20
2.5.2 Algorithme de Panjer . . . . .	21
2.5.3 Comment utiliser l'algorithme de Panjer pour des v.a. positives ou nulles générales ? . . . . .	22
2.6 Hétérogénéité dans le modèle collectif, lois mélanges . . . . .	22
2.6.1 Propriétés générales des lois mélange . . . . .	22
2.6.2 Lois Poisson-mélange . . . . .	24
2.6.3 Mélange de lois exponentielles . . . . .	27
2.6.4 Lois composées mélange . . . . .	28
<b>3 Approximation du modèle individuel par le modèle collectif</b>	<b>30</b>
<b>4 Compléments sur la charge sinistre</b>	<b>31</b>
4.1 Normal Power, Gamma de Bowers . . . . .	31
4.2 FFT . . . . .	31

<b>II Processus de Poisson</b>	<b>32</b>
<b>5 Rappels autour de la loi exponentielle</b>	<b>34</b>
5.1 Définition et premières propriétés . . . . .	34
5.2 Minimum de variables aléatoires exponentielles indépendantes . . . . .	35
5.3 Lois exponentielles multivariées . . . . .	37
5.4 Sommes de variables aléatoires exponentielles indépendantes . . . . .	38
<b>6 Processus de Poisson : définition et premières propriétés</b>	<b>41</b>
6.1 Processus de Poisson homogène . . . . .	41
6.2 Processus de Poisson non homogène . . . . .	48
<b>7 Processus de Poisson composé</b>	<b>50</b>
<b>8 Propriétés de Markov et martingales</b>	<b>52</b>
8.1 Propriétés de Markov . . . . .	52
8.2 Martingales . . . . .	52
<b>9 Thinning, superposition et conditionnement</b>	<b>53</b>
9.1 Thinning et superposition . . . . .	53
9.2 Conditionnement . . . . .	54
9.2.1 Cas d'un processus de Poisson homogène . . . . .	54
9.2.2 Cas d'un processus de Poisson inhomogène . . . . .	55
<b>III Théorie de la ruine</b>	<b>56</b>
9.3 Méthodes de martingales . . . . .	61
<b>10 Quatre différents concepts de ruine</b>	<b>64</b>
10.1 La ruine "vue par les praticiens" . . . . .	64
10.1.1 Ruine économique . . . . .	64
10.1.2 Ruine réglementaire . . . . .	64
10.2 La ruine "vue par les académiques" . . . . .	64
10.2.1 Ruine en temps continu . . . . .	64
10.2.2 Ruine à l'inventaire . . . . .	64
10.2.3 Lien entre ruine et ruine à l'inventaire . . . . .	64
<b>11 Processus de renouvellement</b>	<b>65</b>
11.0.4 Definition . . . . .	65
11.0.5 Some Elementary Results . . . . .	65
11.0.6 Asymptotic Results . . . . .	66
<b>12 Modèle de Cramer-Lundberg</b>	<b>68</b>
12.1 Classical risk process . . . . .	68
<b>13 Probabilité de ruine en temps infini</b>	<b>70</b>

<b>14 Probabilité de ruine en temps fini</b>	71
<b>15 Méthodes de martingales, temps de ruine</b>	72
<b>16 Mesures de risque étudiées en théorie de la ruine</b>	73
<b>17 Problèmes de réassurance et d'investissements optimaux, contrôle stochastique</b>	74
<b>18 Processus de Lévy, résultats connus sur les temps d'atteinte et les extrema</b>	75
<b>19 Versement de dividendes jusqu'à la ruine</b>	76
<b>20 Modèles fluides et lien avec les files d'attente</b>	77
<b>IV Mesures de risque</b>	78
<b>21 Typologie moderne des risques</b>	79
<b>22 Introduction</b>	80
<b>23 Mesures de risque cohérentes</b>	81
<b>24 VaR et autres mesures de risques</b>	82
<b>25 Mesures de risques agrégés</b>	83
<b>26 Mesures de risques dynamiques</b>	84
26.1 Risques d'assurance multiples . . . . .	84
26.2 Risque de marché . . . . .	84
26.3 Risque de crédit . . . . .	84
26.4 Risque opérationnel . . . . .	85
26.5 Risque de modèle et risque de paramètre . . . . .	87
26.6 Risque de dérive . . . . .	87
<b>V Dépendance stochastique</b>	88
<b>27 Introduction</b>	89
<b>28 Copulas ou coupleurs</b>	90
28.1 Définition des coupleurs et théorème de Sklar . . . . .	90
28.2 Copules à densité et densités conditionnelles . . . . .	90
28.3 Familles de copules usuelles . . . . .	90
28.4 Inférence statistique des copules . . . . .	90
28.5 Copules archimédiennes . . . . .	90

<b>29 Concepts et mesures de dépendance</b>	<b>91</b>
<b>30 Modèles à chocs communs</b>	<b>92</b>
<b>31 Modèles à environnement markovien, modèles à facteurs</b>	<b>93</b>
<b>32 Dépendance des extrêmes</b>	<b>94</b>
<b>VI Appendice, pense-bête</b>	<b>95</b>
<b>33 Lois usuelles</b>	<b>96</b>
33.1 Lois de probabilité usuelles . . . . .	96
<b>34 Types de convergence</b>	<b>98</b>
34.1 Convergence en Loi . . . . .	98
34.2 Convergence presque sûre et convergence en probabilité . . . . .	103
34.3 Convergence en moyenne . . . . .	111
<b>35 Théorèmes de convergence</b>	<b>115</b>
35.1 Loi des grands nombres . . . . .	115
35.2 Théorème central limite . . . . .	118
35.3 Convergence de la fonction de répartition empirique . . . . .	122
<b>36 Espérance conditionnelle</b>	<b>123</b>
36.1 Definition of Conditional Expectation . . . . .	123
36.1.1 General definition . . . . .	123
36.1.2 Couples of random variables with p.d.f. . . . .	125
36.2 Properties of Conditional Expectation . . . . .	126
36.2.1 Conditional expectation . . . . .	126
36.2.2 Conditional variance . . . . .	128
36.2.3 Compound distributions . . . . .	129
<b>37 Citations (à dispatcher)</b>	<b>130</b>
<b>VII Bibliographie</b>	<b>131</b>

## Première partie

# Modélisation de la charge sinistre : du modèle individuel au modèle collectif

Pour représenter et quantifier le montant cumulé de tous les sinistres à payer sur une période donnée, l'actuaire peut utiliser le modèle individuel (voir section [1]) ou le modèle collectif (voir section [2]). L'avantage du modèle individuel est qu'il permet de tenir compte de l'hétérogénéité du portefeuille. En effet, si tous les contrats ont les mêmes caractéristiques, alors le modèle individuel correspond exactement au modèle collectif, et fait intervenir les distributions composées (voir section [2.3]). Le modèle collectif peut également permettre de tenir compte de l'hétérogénéité du portefeuille en utilisant les lois mélanges et les lois mélanges composées (voir section [2.6]).

# Chapitre 1

## Modèle individuel

Le modèle individuel vise à représenter le montant total des sinistres à payer par la compagnie d'assurances sur une période donnée (typiquement un an) en sommant assuré par assuré les montants des sinistres subis par chaque individu sur cette période. Soit  $n$  le nombre d'assurés, aussi appelé l'effectif du portefeuille d'assurances, ou encore le nombre de polices. Soit  $I_i$  la variable aléatoire de Bernoulli qui vaut 1 si au moins un sinistre a touché la  $i$ -ème police et 0 sinon, et  $W_i$  la variable aléatoire à valeurs dans  $\mathbb{R}^+$  représentant le montant total de ces éventuels sinistres (qui peut être décrite soit directement, soit à l'aide du modèle collectif décrit dans la section 2). La **charge sinistre** totale  $S$  sur la période considérée est alors donnée par la formule

$$S = \sum_{i=1}^n I_i \cdot W_i.$$

En pratique, il est très difficile de mener les calculs dès que le nombre de polices est élevé, même sous des hypothèses restrictives. Le plus souvent, on supposera que les  $I_i$  sont i.i.d., avec  $P(I_1 = 1) = p$ , que les  $W_i$  sont i.i.d., et que les  $W_i$  sont indépendants des  $I_i$ . Dans ce cas, la fonction de répartition de  $S$  est donnée par la formule classique des convolutions :

$$F_S(x) = \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} F_W^{*k}(x),$$

où  $F_W^{*k}$  est la fonction de répartition de  $W_1 + \dots + W_k$  et vérifie la relation de récurrence

$$F_W^{*(k+1)}(x) = \int_0^x F_W^{*k}(x-y) dF_W(y).$$

Remarquons que

$$C_n^k p^k (1-p)^{n-k}$$

représente ici la probabilité que  $k$  contrats parmi  $n$  aient subi au moins un sinistre sur la période considérée. Dans ce cas,

$$E(S) = npE(W_1) \quad \text{et} \quad \text{Var}(S) = np^2\text{Var}(W_1) + np(1-p)[E(W_1)]^2.$$

Lorsque  $W$  suit un certain type de lois, comme les lois Gamma, il est possible d'utiliser les propriétés d'additivité de ces lois pour obtenir directement les  $F_W^{*k}$  pour  $k \geq 1$ . Par exemple, si  $W \sim \Gamma(\alpha, \lambda)$ , dont la densité est donnée par l'équation 5.1 alors  $F_W^{*k}$  est la fonction de répartition d'une loi Gamma de paramètres  $(k\alpha, \lambda)$ . Ce résultat se généralise si l'indépendance est vérifiée, même si les  $W_i \sim \Gamma(\alpha_i, \lambda)$ , avec des paramètres  $\alpha_i$  différents, mais le même paramètre  $\lambda$ .

Si les  $I_i$  sont indépendants des  $W_i$ , dans le cas général, on peut juste écrire :

$$F_S(x) = \sum_{i_1=0}^1 \dots \sum_{i_n=0}^1 P(I_1 = i_1, \dots, I_n = i_n) P(i_1 W_1 + \dots + i_n W_n \leq x), \quad (1.1)$$

ce qui correspond à une somme de  $2^n$  termes, impossible à utiliser. Dans l'exemple, si  $W_i \sim \Gamma(\alpha_i, \lambda)$ , avec des paramètres  $\alpha_i$  différents, mais le même paramètre  $\lambda$ , alors

$$P(i_1 W_1 + \dots + i_n W_n \leq x)$$

est la fonction de répartition d'une loi

$$\Gamma \left( \sum_{j=1}^n i_j \alpha_i, \lambda \right).$$

Si les  $W_i$  sont i.i.d., on peut simplifier (1.1) en

$$F_S(x) = \sum_{k=0}^n P(N = k) F_W^{*k}(x),$$

où  $N$  est la variable aléatoire correspondant au nombre de polices (parmi  $n$ ) ayant subi au moins un accident. Le modèle individuel étant souvent difficile à utiliser numériquement, on lui préfère souvent le modèle collectif, dans lequel le  $N$  n'est plus le nombre de polices ayant au moins un accident, mais le nombre total d'accidents, sans distinguer police par police. Avant de passer à l'étude du modèle collectif, mentionnons l'existence de l'algorithme de De Pril, qui permet théoriquement d'obtenir la loi de  $S$ , mais qui est en pratique très peu utilisé (voir Partrat et Besson (2004) page 129, et exercice en TD).

# Chapitre 2

## Modèle collectif

### 2.1 Modèle collectif

Le modèle collectif consiste à approcher le modèle individuel non plus en regardant si chaque police fait défaut ou pas, mais en comptabilisant un nombre aléatoires de montants de sinistres i.i.d.. On définit ainsi la charge sinistre totale sur une période  $T$  dans le modèle collectif par la variable aléatoire positive

$$S^{\text{coll}} = \sum_{i=1}^N W_i,$$

où  $N$  est une variable aléatoire à valeurs dans  $\mathbb{N}$  représentant le nombre de sinistres sur la période  $T$ , et pour  $i \geq 1$ ,  $W_i$  est une variable aléatoire à valeurs dans  $\mathbb{R}^+$  représentant le coût du  $i$ -ème sinistre, avec la convention selon laquelle la somme est nulle si  $N = 0$ . Les  $(W_i)_{i \geq 1}$  sont supposés indépendants et identiquement distribués, et indépendants de  $N$  (indépendance fréquences-coûts). Ces deux hypothèses sont contestables, le fait que les montants soient représentés par des variables identiquement distribuées n'est possible que si le facteur d'actualisation peut être négligé, et s'il n'y a pas de risque de dérive du coût des sinistres ; l'indépendance fréquences-coûts est valable si le portefeuille est homogène. Si le portefeuille ne l'est pas (par exemple géographiquement), alors on peut tenir compte de cette hétérogénéité par des lois mélanges et composées mélanges (voir section 2.6). Toutefois ces hypothèses certes restrictives facilitent énormément les calculs grâce aux résultats sur les lois composées (voir section 2.3).

### 2.2 Lois utilisées pour le nombre de sinistres

La loi de Poisson a le double avantage d'être compatible à une étude dynamique de la survenance des sinistres (voir le chapitre 6) et d'apparaître naturellement comme limite de lois binomiales (voir page 134 de Durrett (1999)).

Néanmoins, la surdispersion due à l'hétérogénéité du portefeuille peut conduire à lui préférer d'autres types de lois (voir le chapitre 2.6 de ce document et le chapitre *Modélisation de la fréquence des sinistres* du cours de Christian Partrat).

## 2.3 Lois composées

Soit

$$S = \sum_{i=1}^N W_i,$$

où les  $W_i$  sont des v.a.i.i.d. et indépendantes de  $N$ , et où  $S = 0$  si  $N = 0$ . Il y a trois types de résultats principaux à connaître sur les distributions composées : la formule (2.1) qui donne la fonction de répartition de  $S$ , les formules sur les moments, et la formule sur les transformées de Laplace et les fonctions génératrices (voir section 2.4.6 après un rappel sur les transformées de Laplace, de Fourier et sur les fonctions génératrices), qui permet d'ailleurs de retrouver les formules sur les moments par dérivations successives. Dans certains cas, on dispose de méthodes numériques pour accélérer les calculs (algorithme de Panjer (voir section 2.5), FFT (voir section 4.2) ou approximations du type Gamma ou Normal Power (voir section 4.1)).

Commençons par écrire la fonction de répartition de  $S$  en conditionnant par le nombre de sinistres  $N$  :

$$F_S(x) = \sum_{n \geq 0} P(N = n) F_W^{*n}(x), \quad (2.1)$$

où  $F_W^{*n}$  est la fonction de répartition de  $W_1 + \dots + W_n$  et vérifie la relation de récurrence

$$\text{pour } n \geq 0, \quad F_W^{*(n+1)}(x) = \int_0^x F_W^{*n}(x-y) dF_W(y).$$

On obtient également par conditionnement sur  $N$  les premiers moments de  $S$  :

$$E(S) = E(E[S | N]) = E(N.E(W_1)) = E(N).E(W_1),$$

et

$$\begin{aligned} \text{Var}(S) &= E(\text{Var}[S | N]) + \text{Var}(E[S | N]) \\ &= E(N.\text{Var}(W_1)) + \text{Var}(N.E(W_1)) \\ \text{Var}(S) &= E(N).\text{Var}(W_1) + [E(W_1)]^2.\text{Var}(N) \end{aligned} \quad (2.2)$$

Cas particulier important : loi Poisson-composée.

Si  $N \sim \text{Poi}(\lambda)$ , alors  $E(N) = \text{Var}(N) = \lambda$ , et la formule (2.2) se simplifie en

$$\text{Var}(S) = \lambda E(W_1^2).$$

D'une manière générale, la formule (2.2) décompose la variance de  $S$  en deux termes : le premier correspond à la variabilité des coûts des sinistres autour du coût moyen, le second correspond à la variabilité du nombre de sinistres autour de la moyenne. Elle fait appel à la notion de variance conditionnelle (voir définition VI.4 en appendice) et à la formule de décomposition de la variance (voir section 36.2 en appendice). Ces formules peuvent aussi se retrouver grâce aux fonctions génératrices (voir section 2.4.6). Nous devons d'abord revenir sur les définitions et les propriétés de ces objets.

## 2.4 Rappels sur les transformées de Laplace, et les fonctions génératrices

### 2.4.1 Définitions et premières propriétés

La fonction caractéristique d'une variable aléatoire est définie à partir de la transformée de Fourier de sa loi. Il convient de rappeler quelques propriétés de la transformée de Fourier, dont la définition peut varier d'une source à l'autre, sans changer l'essence des résultats qui s'y rapporte. Ces résultats sur la transformée de Fourier sont principalement

- son **injectivité**, qui va nous permettre de caractériser une loi par sa **fonction caractéristique**,
- le fait que la transformée de Fourier d'un produit de convolution est le produit des transformées de Fourier, qui implique que **la fonction caractéristique de la somme de deux v.a. indépendantes est le produit de leurs fonctions caractéristiques** (associé au point précédent, cela nous donne d'ailleurs un critère d'indépendance, et un moyen simple de démontrer la stabilité d'une famille de lois (par exemple Poisson ou normale) par l'addition),
- une **formule d'inversion** qui nous permet de retrouver la loi quand on connaît la transformée de Fourier, et qui peut fournir des méthodes numériques (nous reviendrons sur les méthodes FFT (Fast Fourier Transform) au chapitre 4.2),
- le fait que **les moments de la variable aléatoire s'obtiennent en fonction des dérivées successives de la fonction caractéristique**. Ces moments (voir section 2.4.2 pour un rappel sur ce sujet), ainsi que les *cumulants*, peuvent être obtenus directement en dérivant des fonctions construites à partir de la fonction caractéristique, ce qui nous amènera à les considérer (fonction génératrice des moments, des cumulants, des probabilités, transformée de Laplace). Ces objets ne seront pas définis sur le même domaine, mais dans les conditions habituelles vérifieront les principes d'indépendance (voir récapitulatif dans le théorème L.8) et de caractérisation. On choisira l'une ou l'autre selon le problème (v.a. à valeurs dans  $\mathbb{N}$  ou continues, les quantités recherchées sont les moments centrés ou factoriels, ...), mais les idées générales restent les mêmes.

Nous rappelons ici uniquement les résultats dont nous aurons besoin. Pour plus de détails sur les transformées de Fourier, et plus largement sur l'analyse réelle et complexe, consulter le livre de Rudin [1987]. Il est aussi utile de se remémorer le théorème de transfert et la proposition I.1 (particulièrement dans le sens  $1 \Rightarrow 4$ ), que nous utiliserons très bientôt (respectivement dès la définition I.2 et dès la proposition I.5).

### Théorème I.1 Théorème de transfert, ou de la loi image

*Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et  $X$  une variable aléatoire réelle. On note  $P_X$  la probabilité image de  $P$  par  $X$ , définie sur  $(\mathbb{R}, \mathcal{B})$  par  $P_X(B) = P(X^{-1}(B))$ ,  $\forall B \in \mathcal{B}$ . Soit  $g$  une variable aléatoire réelle définie sur  $(\mathbb{R}, \mathcal{B})$ .  $g(X)$  est intégrable par rapport à  $P$  si et seulement si  $g$  est intégrable par rapport à  $P_X$  et on a alors :*

$$E^P[g(X)] = E^{P_X}[g].$$

**Proposition I.1** *Soit  $X$  et  $Y$  deux variables aléatoires définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Les quatre propositions suivantes sont équivalentes :*

1.  *$X$  et  $Y$  sont indépendantes.*
2. *Pour toutes fonctions réelles  $f$  et  $g$  mesurables et positives,  $E^P[f(X)g(Y)] = E^P[f(X)]E^P[g(Y)]$ .*
3. *Pour toutes fonctions réelles  $f$  et  $g$  mesurables et bornées,  $E^P[f(X)g(Y)] = E^P[f(X)]E^P[g(Y)]$ .*
4. *Pour toutes fonctions complexes mesurables  $f$  et  $g$  telles que  $|f(X)|$  et  $|g(Y)|$  sont bornés,  $E^P[f(X)g(Y)] = E^P[f(X)]E^P[g(Y)]$ .*

### Definition I.1 Transformée de Fourier d'une mesure

*Soit  $P$  une mesure de probabilité définie sur  $(\mathbb{R}, \mathcal{B})$ . On appelle transformée de Fourier de la mesure  $P$ , la fonction  $\varphi$  définie sur  $\mathbb{R}$ , à valeur dans  $\mathbb{C}$ , définie par :  $\varphi(t) = \int e^{itx} dP(x)$ .*

#### Remarques :

- si on note  $f$ , la fonction réelle mesurable, à valeurs dans  $\mathbb{C}$ , qui à  $x$  associe  $e^{itx}$ , on peut réécrire  $\int e^{itx} dP(x)$  comme  $E^P(f)$  avec les notations probabilistes.
- $f$  étant continue, c'est une v.a. sur  $(\mathbb{R}, \mathcal{B})$ .
- On remarque que  $|e^{itx}| = 1$  et donc  $\int |e^{itx}| dP(x) = 1$ .  $x \mapsto e^{itx}$  est donc  $P$ -intégrable et la transformée de Fourier  $\varphi(t)$  est définie en tout  $t \in \mathbb{R}$ .

### Definition I.2 Fonction caractéristique

*Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et  $X$  une variable aléatoire réelle définie sur  $(\Omega, \mathcal{A})$ . On définit la fonction caractéristique de  $X$ , comme étant la transformée de Fourier de la mesure image de  $P$  par  $X$ ,  $P_X^X$ , et on note  $\varphi_X(t) = \int e^{itx} dP_X^X(x) = E^P(e^{itX})$ .*

**Proposition I.2** *Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et  $X$  une variable aléatoire réelle définie sur  $(\Omega, \mathcal{A})$ . La fonction caractéristique de  $X$ ,  $\varphi_X(t)$*

1. est définie sur  $\mathbb{R}$ ,
2. vérifie  $\varphi_X(0) = 1$ ,
3. et pour tout  $t \in \mathbb{R}$ ,  $|\varphi_X(t)| \leq 1$ ,
4. est uniformément continue :

$$\forall \epsilon > 0, \exists \delta > 0, \forall s, t \in \mathbb{R}, |t - s| < \delta \Rightarrow |\varphi_X(t) - \varphi_X(s)| < \epsilon.$$

5. De plus, pour tout  $t \in \mathbb{R}$ ,  $\varphi_X(-t) = \overline{\varphi_X(t)}$ ,
6. et pour tous  $a, b, t \in \mathbb{R}$ ,  $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$ .

### 2.4.2 Moments, fonctions génératrices, transformée de Laplace

#### Théorème I.2 Fonction caractéristique et moments

Soit  $X$  une variable aléatoire réelle définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ .

Si  $E[|X|^n] < \infty$ , la fonction caractéristique  $\varphi_X(t)$  est  $n$  fois dérivable et :

$$\varphi_X^{(k)}(0) = i^k E(X^k), \forall k \leq n.$$

La réciproque du théorème I.2 est vrai pour  $n$  pair uniquement. Pour  $n$  impair, si  $\varphi_X$  dérivable  $n$  fois, alors  $\varphi_X$  dérivable  $n - 1$  fois, et comme  $n - 1$  est pair, on peut utiliser la réciproque du théorème I.2 avec  $n - 1$ .

On a le même genre de résultat avec la transformée de Laplace, et avec d'autres concepts qu'on peut choisir d'utiliser en fonction du problème. L'outil clé de ce paragraphe est la dérivation sous le signe somme.

**Definition I.3 Transformée de Laplace d'une variable aléatoire** Soit  $X$  une variable aléatoire réelle définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On appelle transformée de Laplace de la loi de  $X$  (ou de la mesure de probabilité  $P^X$ ), la fonction réelle  $\mathcal{M}_X$ , définie par  $\mathcal{M}_X(t) = E^P[e^{-tx}] = \int e^{-tx} dP^X(x)$ .  $\mathcal{M}_X$  est définie pour  $t$  tel que l'espérance précédente est finie.

**Propriété I.1** Soit  $X$  une variable aléatoire réelle définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . La transformée de Laplace  $\mathcal{M}_X$  est définie sur un intervalle contenant 0.

De plus, si  $X$  est à valeurs dans  $\mathbb{R}^+$ , alors l'ensemble de définition de  $\mathcal{M}_X$  contient  $\mathbb{R}^+$ .

La transformée de Laplace est un outil indispensable en théorie de la ruine, on le verra par exemple dans l'étude du modèle de Cramer-Lundberg (voir section 12).

#### Definition I.4 Fonction génératrice des moments

Soit  $X$  une variable aléatoire réelle définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On appelle **fonction génératrice des moments** de la variable aléatoire  $X$ , la fonction  $M_X(t)$  définie par  $M_X(t) = E^P[e^{tX}]$ , pour tout  $t \in \mathbb{R}$  tel que  $e^{tX}$  est  $P$ -intégrable.

La fonction génératrice des moments est s'obtient de manière immédiate à partir de la transformée de Laplace

$$M(s) = L(-s),$$

et on peut transposer les propriétés déjà présentées pour la transformée de Laplace.

**Propriété I.2** *Si  $X$  est une variable aléatoire réelle positive, et si  $\mathcal{M}_X$  (ou  $L_X$ ) reste finie sur un intervalle ouvert contenant 0, alors  $\mathcal{M}_X$  est infiniment dérivable en zéro,  $X$  admet des moments à tous les ordres et*

$$E^P [X^r] = M_X^{(r)}(0) = (-1)^r L_X^{(r)}(0).$$

#### Definition I.5 Fonction génératrice des cumulants

Soit  $X$  une variable aléatoire réelle définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On appelle fonction génératrice des cumulants de la loi de  $X$ , la fonction  $\mathcal{K}_X$  définie par  $\mathcal{K}_X(t) = \log E^P [e^{tX}]$  pour tout  $t \in \mathbb{R}$ , tel que l'espérance est définie.

La fonction génératrice des cumulants est simplement le logarithme de la fonction génératrice des moments. Les “cumulants” interviennent dans le développement en série (quand celui ci existe) de  $\mathcal{K}_X$  (voir page 15).

#### Definition I.6 seconde fonction caractéristique

Soit  $X$  une variable aléatoire réelle définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On appelle seconde fonction caractéristique de la loi de  $X$ , la fonction par  $t \rightarrow \log \varphi_X(t)$  (il s'agit du logarithme complexe).

#### Definition I.7 fonction génératrice (des probabilités)

Soit  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . La fonction génératrice de  $X$  au point  $u$  est définie par

$$G_X(u) = E^P [u^X].$$

Remarquons que pour  $s \in \mathbb{R}$  tel que  $M_X(s)$  existe, alors  $G_X(e^s)$  existe et

$$M_X(s) = G_X(e^s).$$

**Propriété I.3** Soit  $N$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  et à valeurs dans  $\mathbb{N}$ . Alors

$$G_N(u) = \sum_{n=0}^{+\infty} P(N=n) u^n$$

est une série entière de rayon de convergence supérieur ou égal à 1, qui caractérise la loi de  $N$ , et telle que pour  $n \geq 0$ ,

$$P(N=n) = \frac{G_N^{(n)}(0)}{n!}.$$

Ces outils permettent d'obtenir différents renseignements sur la loi de la variable aléatoire en les dérivant. La propriété précédente permet d'obtenir des probabilités. Voyons donc quel type de moment on peut obtenir, et comment.

## Petits rappels sur les différents types de moments

On peut être amené à utiliser différents types de moments, les moments simples, centrés ou factoriels, voire les cumulants.

- Pour  $k \in \mathbb{N}$ , le moment simple d'ordre  $k$  d'une variable aléatoire  $X$ , noté  $m_k$ , est défini par

$$m_k = E(X)^k.$$

Remarquons que  $m_0 = 1$ , et que  $m_1$ , noté aussi  $m$ , est égal à  $E(X)$ . Le moment simple d'ordre  $k$ , lorsqu'il existe, est égal à la dérivée  $k$ -ème en zéro de la fonction génératrice des moments d'après la propriété I.2.

- Pour  $k \geq 1$ , le moment centré d'ordre  $k$  d'une variable aléatoire  $X$ , noté  $\mu_k$ , est défini par

$$\mu_k = E(X - m)^k.$$

Remarquons que  $\mu_1 = m - m = 0$ , et que  $\mu_2 = \text{Var}(X)$ . On peut les obtenir à partir des moments simples par la formule

$$\mu_k = \sum_{j=0}^k (-1)^{k+l} C_k^l m_l m^{k-l}.$$

- Les cumulants, notés  $\kappa_r$ , sont les coefficients du développement en série entière en zéro (qui existe dès que  $M_X$  est définie sur un voisinage de 0) de la fonction génératrice des cumulants (voir définition I.5), qui est définie par

$$\mathcal{K}_X(t) = \log(M_X(t)).$$

En cas d'existence, on a donc

$$\kappa_r = \mathcal{K}_X^{(r)}(0).$$

En particulier,

$$\kappa_1 = m, \quad \kappa_2 = \mu_2 = \text{Var}(X), \quad \kappa_3 = \mu_3, \quad \kappa_4 = \mu_4 - 3\mu_2^2. \quad (2.3)$$

- Pour une variable aléatoire  $N$  à valeurs dans  $\mathbb{N}$  (représentant par exemple le nombre de sinistres sur une période donnée), le moment factoriel d'ordre  $k$ , noté  $\mu_{(k)}$ , est défini par

$$\mu_{(k)} = E(X(X-1)\dots(X-k+1)).$$

Il est possible de récupérer les moments simples à partir des moments factoriels grâce aux nombres de Stirling de seconde espèce :

$$m_k = \sum_{j=1}^k S(k, j) \mu_{(j)},$$

où les nombres de Stirling de seconde espèce  $S(n, k)$  correspondent aux nombres de partitions de  $\{1, \dots, n\}$  en  $k$  sous-ensembles non vides, et peuvent être obtenus récursivement par :

$$\begin{aligned} S(n, k) &= S(n - 1, k - 1) + k \cdot S(n - 1, k) \\ S(n, 1) &= S(n, n) = 1. \end{aligned}$$

En particulier,

$$\begin{aligned} m &= \mu_{(1)}, \quad m_2 = \mu_{(1)} + \mu_{(2)}, \quad m_3 = \mu_{(1)} + 3\mu_{(2)} + \mu_{(3)}, \\ \text{et } m_3 &= \mu_{(1)} + 7\mu_{(2)} + 6\mu_{(3)} + \mu_{(4)}. \end{aligned}$$

Rappelons aussi pour  $n \geq 0$ ,  $P(N = n)$  est donné directement par la dérivée  $n$ -ème de  $G_N$  en 0 d'après la propriété I.3. De plus, lorsque le rayon de convergence de la série entière  $G_N(u)$  est strictement supérieur à 1, alors on peut obtenir les moments factoriels de  $N$  grâce au développement en série entière de  $G_N$  en 1 (et non plus en 0) :

$$\forall n \in \mathbb{N}, \quad \mu_{(n)} = G_N^{(n)}(1).$$

Pour calculer les moments simples d'une variable aléatoire entière, lorsqu'on connaît  $G_N$  et qu'elle a une expression qui s'y prête, il peut être plus efficace de dériver plusieurs fois  $G_N$  en 1 pour obtenir les  $n$  premiers moments factoriels de  $N$ , puis d'utiliser les formules avec les nombres de Stirling.

### 2.4.3 Injectivité et inversion

#### Théorème I.3 formule d'inversion avec fonction de répartition

Soit  $P$  une probabilité sur  $(\mathbb{R}, \mathcal{B})$  ayant pour transformée de Fourier,  $\varphi(t) = \int e^{itx} dP(x)$ . On note  $F(x) = P([-∞, x])$ , la fonction de répartition de  $P$ .

1. Soient  $a$  et  $b$  deux réels ( $a < b$ ). Alors,

$$\frac{F(b) + F(b^+)}{2} - \frac{F(a) + F(a^+)}{2} = \frac{1}{2\pi} \lim_{c \rightarrow \infty} \int_{-c}^c \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt.$$

2. Si  $F$  est continue en  $a$  et en  $b$ , alors,

$$F(b) - F(a) = \frac{1}{2\pi} \lim_{c \rightarrow \infty} \int_{-c}^c \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt.$$

#### Théorème I.4 injectivité de la transformée de Fourier d'une mesure de probabilité

Soit  $P$  et  $Q$  deux probabilités sur  $(\mathbb{R}, \mathcal{B})$ . Si  $\int e^{itx} dP(x) = \int e^{itx} dQ(x)$ , pour tout  $t \in \mathbb{R}$  (égalité entre les transformées de Fourier), alors  $P = Q$ .

**Corollaire I.1** Soit  $X$  et  $Y$ , deux variables aléatoires réelles définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . L'égalité entre les fonctions caractéristiques de  $X$  et  $Y$  implique que  $X$  et  $Y$  ont la même loi, c'est-à-dire que  $P^X = P^Y$ .

La fonction caractéristique  $\varphi_X$  caractérise donc  $X$ . Il en est de même pour la transformée de Laplace si elle est définie sur un intervalle contenant 0.

### Théorème I.5 formule d'inversion

Soit  $P$  une probabilité sur  $(\mathbb{R}, \mathcal{B})$  ayant pour transformée de Fourier,  $\varphi(t)$ . Si  $\varphi$  est intégrable par rapport à la mesure de Lebesgue sur  $\mathbb{R}$  (c'est-à-dire  $\int_{\mathbb{R}} |\varphi(t)| dt < \infty$ ), alors  $P$  possède une densité  $p$  par rapport à la mesure de Lebesgue, cette densité est continue et est donnée par :

$$p(x) = \frac{1}{2\pi} \int e^{-itx} \varphi(t) dt.$$

#### 2.4.4 Indépendance et caractérisation de l'indépendance

##### Definition I.8 cas multivarié

Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire défini sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . La fonction caractéristique de  $X$  est la fonction complexe, définie par  $\varphi_X(t) = E^P [e^{i\langle t, X \rangle}]$ , pour tout  $t = (t_1, \dots, t_n) \in \mathbb{R}^n$ , où  $\langle t, X \rangle = \sum_{i=1}^n t_i X_i$ .

##### Definition I.9 Transformée de Laplace d'un vecteur aléatoire

Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire défini sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On appelle transformée de Laplace de la loi de  $X$  (ou de la mesure de probabilité  $P^X$ ), la fonction  $\mathcal{M}_X(t) = E^P [e^{-\langle t, X \rangle}]$ , pour tout  $t = (t_1, \dots, t_n)$  tel que l'intégrale précédente est définie.

##### Definition I.10 fonction génératrice des moments d'un vecteur aléatoire

Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire défini sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On appelle **fonction génératrice des moments conjoints** du vecteur aléatoire  $X$ , la fonction  $\bar{\varphi}_X(t) = E^P [e^{t_1 X_1 + \dots + t_n X_n}]$ , pour tout  $(t_1, \dots, t_n) \in \mathbb{R}^n$ , tel que  $e^{t_1 X_1 + \dots + t_n X_n}$  est  $P$ -intégrable.

##### Definition I.11 fonction génératrice d'un vecteur aléatoire

Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire défini sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . La fonction génératrice de  $X$  au point  $(u_1, \dots, u_n)$  est définie par

$$G_X(u_1, \dots, u_n) = E^P [u_1^{X_1} u_2^{X_2} \dots u_n^{X_n}].$$

L'injectivité de la transformée de Fourier reste vraie dans le cas vectoriel, ce qui fournit le résultat suivant, vrai aussi pour la transformée de Laplace si elle est définie sur un voisinage de 0.

**Propriété I.4** Soit  $X$  et  $Y$  deux vecteurs aléatoires définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Si  $\varphi_X = \varphi_Y$ , alors  $P^X = P^Y$ .

### Propriété I.5 Somme de deux variables indépendantes

*Si  $X$  et  $Y$  sont deux variables aléatoires réelles définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  indépendantes, alors :*

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t).$$

La propriété I.5 nous sera très utile pour étudier les distributions composées, en fournissant entre autres l'argument essentiel à la preuve de la proposition I.3.

### Théorème I.6 fonctions caractéristiques et indépendance

*Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire défini sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Une condition nécessaire et suffisante pour que les composantes  $X_i$  soient indépendantes est que la fonction caractéristique de  $X$ ,  $\varphi_X(t)$  soit le produit des fonctions caractéristiques des  $X_i$ ,  $\varphi_{X_i}(t_i)$ , pour tout  $t = (t_1, \dots, t_n)$  de  $\mathbb{R}^n$ .*

### Théorème I.7 fonctions caractéristiques et indépendance

*Soit  $X$  et  $Y$  deux vecteurs aléatoires définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  à valeurs respectivement dans  $\mathbb{R}^n$  et  $\mathbb{R}^m$ .  $X$  et  $Y$  sont indépendants si et seulement si leur fonction caractéristique conjointe  $\varphi_{(X,Y)}(u, v)$  (où  $u \in \mathbb{R}^n$  et  $v \in \mathbb{R}^m$ ) est égale au produit des fonctions caractéristiques de  $X$  et de  $Y$ ,  $\varphi_X(u)\varphi_Y(v)$ ,  $\forall(u, v) \in \mathbb{R}^n \times \mathbb{R}^m$ .*

### Propriété I.6 lien entre le cas multivarié et le cas univarié

*Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^n$  défini sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Alors  $\varphi_X(t) = \varphi_{t'X}(1)$ .*

Terminons par un récapitulatif important :

**Théorème I.8** *Soit  $X$  et  $Y$  deux variables aléatoires définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  à valeurs dans  $\mathbb{R}$ . Lorsque ces quantités existent, si  $X$  et  $Y$  sont indépendantes, alors*

$$\begin{aligned}\varphi_{X+Y}(t) &= \varphi_X(t).\varphi_Y(t) \\ M_{X+Y}(t) &= M_X(t).M_Y(t) \\ L_{X+Y}(t) &= L_X(t).L_Y(t) \\ G_{X+Y}(t) &= G_X(t).G_Y(t) \\ C_{X+Y}(t) &= C_X(t) + C_Y(t).\end{aligned}$$

En particulier, la dernière égalité du théorème précédent nous permet en identifiant avec les formules (2.3) de retrouver que si  $X$  et  $Y$  sont indépendantes, alors  $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$ , mais surtout de montrer que

$$\mu_3(X+Y) = \mu_3(X) + \mu_3(Y),$$

et que cette formule ne se généralise pas pour  $k \geq 4$ .

### 2.4.5 Lois usuelles

Exemples de fonctions caractéristiques pour des lois usuelles.

Distribution	probabilités	fonction caractéristique $\varphi_X(t)$
Binomiale $B(n, p)$	$C_n^k p^k q^{n-k}$	$((1-p) + pe^{it})^n$
Poisson	$\frac{e^{-\lambda} \lambda^k}{k!}$	$\exp[\lambda(e^{it} - 1)]$
Binomiale négative	$C_{k-1}^{r-1} p^r q^{k-r}$ , $k \geq r$	

Distribution	densité	fonction caractéristique $\varphi_X(t)$
Uniforme	$\mathbb{I}_{[0,1]}(x)$	$\frac{e^{it}-1}{it}$
Normale centrée réduite	$(2\pi)^{-1/2} e^{-x^2/2}$	$e^{-t^2/2}$
Normale	$(2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/(2\sigma^2)}$	$e^{i\mu t - \sigma^2 t^2/2}$
Gamma	$\frac{\theta^p}{\Gamma(p)} e^{-\theta x} x^{p-1}$	$\left(\frac{1}{1-\theta it}\right)^p$
Cauchy	$\frac{1}{\pi(x^2+1)}$	$e^{- t }$
Laplace	$\frac{1}{2} e^{- x }$	$\frac{1}{1+t^2}$

### 2.4.6 Lois composées

**Proposition I.3** Pour

$$S = \sum_{i=1}^N Y_i,$$

avec les hypothèses d'indépendance,

$$M_S(s) = G_N(M_Y(s)). \quad (2.4)$$

On verra également des propriétés des fonctions génératrices des lois mélanges (voir section 2.6). On en déduit le résultat suivant sur les moments des lois composées :

**Proposition I.4** (A CORRIGER, faire cas général) Soit

$$S = \sum_{i=1}^N W_i,$$

Pour  $r, t > 0$ , si  $G_W$  est la fonction génératrice de  $W_1$ ,

$$h(r, t) = E(r^S) = e^{-\mu(t)(1-G_W(r))},$$

d'où

$$E(S) = \frac{d}{dr} [E(r^S)]_{r=1} = \lambda t G'_W(1) = \mu(t) E(W_1)$$

et

$$Var(S) = \frac{d^2}{dr^2} [E(r^S)]_{r=1} = \mu(t) E(W_1^2).$$

## 2.5 Famille et algorithme de Panjer

On démontre ici sous la forme d'un problème que les lois vérifiant la relation de Panjer (2.5) sont exactement les lois binomiales, de Poisson, et binomiales négatives, et que pour les distributions composées, i.e. de variables aléatoires du type

$$S = \sum_{i=1}^N Y_i,$$

lorsque  $N$  appartient à la famille de Panjer, et que les  $Y_i$  sont à valeurs dans  $d\mathbb{N}^*$ , où  $d > 0$ , on peut calculer directement avec l'algorithme de Panjer (2.6) les masses de probabilité de  $S$ .

### 2.5.1 Etude et caractérisation des distributions vérifiant la relation de récurrence de Panjer

On représente le nombre de sinistres rencontrés par une compagnie d'assurances sur une période par une variable aléatoire discrète  $N$ , à valeurs dans  $\mathbb{N}$ , et décrite par les  $p_k = P(N = k)$  pour  $k \in \mathbb{N}$ . Parmi ces distributions on s'intéressera ici à celles vérifiant la relation de récurrence de Panjer :

$$\exists a < 1, \quad b \in \mathbb{R}, \quad \forall k \in \mathbb{N}^*, \quad p_k = \left( a + \frac{b}{k} \right) p_{k-1} \quad (2.5)$$

Le but de cette partie est de caractériser les distributions vérifiant (2.5).

Rappelons que la loi de Poisson de paramètre  $\lambda$  est décrite par

$$\forall k \in \mathbb{N}, p_k = e^{-\lambda} \frac{\lambda^k}{k!},$$

et que la loi binomiale de paramètres  $n \in \mathbb{N}^*$  et  $p \in ]0, 1[$  est décrite par

$$\forall 0 \leq k \leq n, \quad p_k = \binom{n}{k} p^k (1-p)^{n-k}$$

et  $p_k = 0$  pour  $k > n$ .

1. Montrer que si  $N$  vérifie (2.5) et que  $a = 0$ , alors  $N$  suit une loi de Poisson dont on précisera le paramètre.
2. On définit la distribution binomiale négative de paramètres  $\alpha > 0$  et  $p \in ]0, 1[$  par
$$\forall k \in \mathbb{N}, p_k = \binom{\alpha + k - 1}{k} (1-p)^\alpha p^k$$
3. Calculer sa moyenne et sa variance en fonction de  $\alpha$  et de  $p$ .
4. Comment pouvez-vous interpréter cette distribution à partir d'une expérience réalisée avec succès avec probabilité  $p$  ?

5. Quelle distribution obtient-on lorsque  $\alpha = 1$  ?
6. Montrer que dans le cas  $\alpha = 1$  la distribution obtenue est sans mémoire, c'est-à-dire que  $P(N > n + m | N > n) = P(N > m)$  pour tous  $m, n \in \mathbb{N}$ .
7. Toujours dans le cas  $\alpha = 1$ , montrer que  $N$  peut être vu comme la partie entière d'une variable aléatoire absolument continue sans mémoire que l'on précisera.
8. Lorsque  $N$  vérifie la relation de récurrence (2.5) avec  $a \neq 0$ , montrer que pour tout  $k \in \mathbb{N}$ ,

$$p_k = p_0 \frac{a^k}{k!} \prod_{i=0}^{k-1} (\Delta + i)$$

pour un certain  $\Delta$  à préciser.

9. Montrer que pour tout  $k \in \mathbb{N}$ ,

$$p_k = \binom{\Delta + k - 1}{k} (1-p)^\Delta a^k (1-a)^\Delta$$

10. En déduire selon le signe de  $a$  la distribution de  $N$ .
11. Conclure que les distributions vérifiant la relation (2.5) sont exactement les distributions de Poisson, binomiales et binomiales négatives.
12. Parmi ces 3 types de distributions, lequel choisiriez-vous pour modéliser  $N$  si des observations vous montraient que la variance empirique de  $N$  est beaucoup plus grande que la moyenne empirique de  $N$  ?

### 2.5.2 Algorithme de Panjer

On s'intéresse maintenant à un montant composé

$$X = \sum_{k=1}^N U_k$$

où  $N$  et les  $U_k$ ,  $k \in \mathbb{N}$  sont des variables aléatoires à valeurs entières. De plus les  $U_k$  sont indépendantes, identiquement distribuées, de loi décrite par les  $q_k = P(U_i = k)$ , et indépendantes de  $N$ . La somme est nulle par convention si  $N = 0$ . Définissons de plus l'espérance d'une variable aléatoire  $Y$  à valeurs dans  $\mathbb{N}$  conditionnellement à un événement  $A$  par :

$$E(Y|A) = \sum_{k \in \mathbb{N}} k P(Y = k|A)$$

Le but de cette partie est de démontrer la validité de l'algorithme de Panjer qui permet d'obtenir récursivement la loi de  $X$ .

1. Montrer que pour  $j, k \in \mathbb{N}$  et  $n \in \mathbb{N}^*$ ,

$$E\left(U_1 \mid \sum_{i=1}^n U_i = j\right) = \frac{j}{n}$$

2. Définissons pour  $n, j \in \mathbb{N}$  les probabilités  $q_j^{*n} = P(U_1 + \dots + U_n = j)$ . Montrer que pour  $k, j \in \mathbb{N}$ ,

$$P(U_1 = k | \sum_{i=1}^n U_i = j) = \frac{q_k q_{j-k}^{*(n-1)}}{q_j^{*n}}$$

3. Supposons dans la suite que  $N$  vérifie la relation (2.5). Soit  $r_k = P(X = k)$  pour  $k \in \mathbb{N}$ . Calculer  $r_0$  en fonction des  $q_k$  et des  $p_k$ .
4. En utilisant les questions précédentes, démontrer la formule récursive de l'algorithme de Panjer :

$$\forall j \in \mathbb{N}^*, r_j = \sum_{k=1}^j \left( a + \frac{bk}{j} \right) q_k r_{j-k} \quad (2.6)$$

5. Exemple : pour  $n \in \mathbb{N}$  on définit la transformation stop-loss du cumul  $X$  par  $\pi_n = E[(X - n)_+]$ , où  $(X - n)_+ = \max(X - n, 0)$ . Obtenir  $\pi_n$  en fonction des  $r_k$ .
6. Démontrer une relation de récurrence entre les  $\pi_n$ . (On pourra faire intervenir dans cette relation de récurrence les  $F_X(n) = P(X \leq n)$ .)
7. Expliquer alors comment calculer les  $\pi_n$ .

### 2.5.3 Comment utiliser l'algorithme de Panjer pour des v.a. positives ou nulles générales ?

En pratique, deux types de problèmes risquent de se poser : la loi de la variable aléatoire représentant le coût d'un sinistre peut être absolument continue, ou peut avoir un atome en 0.

Il est facile d'éliminer les sauts d'amplitude nulle dans le modèle Poisson composé.

Pour des v.a. à densité, il faut discréteriser pour obtenir une v.a. à valeurs dans  $\delta\mathbb{N}^*$ , en respectant la moyenne, ou en respectant une règle de prudence qui dépend du problème considéré.

## 2.6 Hétérogénéité dans le modèle collectif, lois mélanges

L'introduction du mélange sert à prendre en compte l'hétérogénéité du portefeuille d'assurance, et a pour effet principal d'augmenter la variance du montant cumulé des sinistres.

### 2.6.1 Propriétés générales des lois mélange

Soit  $X$  une variable aléatoire suivant une certaine loi  $\mathcal{L}$  à un ou plusieurs paramètres, et soit  $\alpha$  le paramètre sur lequel va porter le mélange (dans

la modélisation, on suppose que l'hétérogénéité du portefeuille porte principalement sur ce paramètre). On dit que  $Y$  suit une loi  $\mathcal{L}$ -mélange de loi de mélange  $\Theta$  sur le paramètre  $\alpha$  si

$$Y \sim \mathcal{L}(\alpha\Theta),$$

où le paramètre  $\alpha\Theta$  est donc aléatoire. La variable aléatoire de mélange  $\Theta$  est en général une variable aléatoire de moyenne 1 (de manière à assurer que la moyenne reste préservée, i.e.  $E(X) = E(Y)$ ), et à valeurs dans un ensemble  $A \subset \mathbb{R}$  tel que pour  $\theta \in A$ ,  $\alpha\theta$  soit une valeur possible du paramètre de la loi  $\mathcal{L}$ . La restriction  $E(\Theta) = 1$  n'est pas automatique et dépend des auteurs et du problème.

Plus généralement, on peut adopter la définition suivante :

**Definition I.12** Soit  $\Theta$  une variable aléatoire (de fonction de répartition  $F_\Theta$ ) et  $A \subset \mathbb{R}$  tel que  $P(\Theta \in A) = 1$ , et  $(F(\cdot | \theta))_{\theta \in A}$  une collection de fonctions de répartitions pour  $A \subset \mathbb{R}$ . On dit que  $X$  suit un mélange de lois (avec  $\Theta$  comme loi de mélange) si pour tout  $x \in \mathbb{R}$ , pour tout  $\theta \in A$ ,

$$P(X \leq x | \Theta = \theta) = F(x | \theta).$$

Dans ce cas, pour tout  $x \in \mathbb{R}$ ,

$$F_X(x) = P(X \leq x) = E(E[1_{\{X \leq x\}} | \Theta]) = \int_{\theta \in A} F(x | \theta) dF_\Theta(\theta).$$

Prenons tout de suite deux exemples, pour éclaircir cette notion, les lois Poisson-mélange, qui seront étudiées en détail dans la sous-section suivante, et les lois binomiale-mélange. On dit que  $N$  suit une loi Poisson-mélange si

$$N \sim \text{Poi}(\lambda\Theta),$$

où  $\Theta$  est une variable aléatoire positive et de moyenne 1 (de manière à assurer que  $\lambda\Theta$  est toujours positif, et que  $E(N) = \lambda$ ).

On dit que  $N$  suit une loi binomiale-mélange si

$$N \sim \text{Bin}(n, p\Theta),$$

où  $\Theta$  est une variable aléatoire à valeurs dans  $[0, 1/p]$  et de moyenne 1 (de manière à assurer que  $p\Theta$  est toujours entre 0 et 1, et que  $E(N) = np$ ).

Une formulation plus correcte de ces deux exemples aurait consisté à dire que sachant que  $\Theta = \theta$ ,  $N$  suit une certaine loi de paramètres dépendant de la valeur de  $\theta$  (voir la définition I.14 dans le cas poissonien).

Remarquons qu'on a déjà vu un exemple (certes peu représentatif de l'usage habituel des lois mélanges) de loi mélange dans un chapitre précédent. Le modèle collectif, avec  $N \sim \text{Poi}(\lambda)$  et  $W \sim \text{Exp}(\mu)$ , fournit un exemple de loi Gamma-mélange avec une loi de Poisson comme loi de mélange : en effet si  $N = n$ , la charge sinistre globale

$$S \sim \Gamma(\mu, n),$$

et donc  $S$  suit une loi Gamma-mélange

$$S \sim \Gamma(\mu, N).$$

Dans ce cas, remarquons qu'on n'impose pas  $E(N) = 1$ . La moyenne et la variance de  $S$  qu'on obtenait dans le modèle collectif correspondent à un cas particulier du résultat général suivant, dont la démonstration est basée sur les notions d'espérance de variance conditionnelles, et sur la formule de décomposition de la variance.

**Proposition I.5** *Soit  $X$  une variable aléatoire suivant un mélange de lois (avec  $\Theta$  comme loi de mélange). Alors*

$$E(X) = E(E[X | \Theta]).$$

$$\text{Var}(X) = E(\text{Var}[X | \Theta]) + \text{Var}(E[X | \Theta]).$$

*Preuve :*

voir page [128] les rappels sur la variance conditionnelle.

□

## 2.6.2 Lois Poisson-mélange

Les lois Poisson-mélange forment une classe très importante et très utilisée des lois mélange. La loi de Poisson étant équidispersée, les lois Poisson-mélange (avec une loi de mélange non triviale) sont de fait surdispersées.

**Definition I.13 Lois Poisson-mélange**

*Soit  $\Theta$  une variable aléatoire (de fonction de répartition  $F_\Theta$ ) et  $A \subset ]0, +\infty[$  tel que  $P(\Theta \in A) = 1$ . On dit que  $N$  suit une loi Poisson-mélange de paramètres  $(\lambda, \Theta)$  (avec  $\Theta$  comme loi de mélange) si  $E(\Theta) = 1$  et si pour tout  $n \in \mathbb{N}$ , pour tout  $\theta \in A$ ,*

$$P(N = n | \Theta = \theta) = e^{-\lambda\theta} \frac{(\lambda\theta)^n}{n!}.$$

**Exercice I.1** (*Exemple simple en lien avec la théorie de la crédibilité (voir cours de Pierre Thérond) et les processus de Poisson (voir chapitre 6)) Modèle à bons et mauvais conducteurs :*

*On suppose qu'il y a exactement 2 sortes de conducteurs : les bons conducteurs qui ont un accident tous les 10 ans en moyenne et les mauvais conducteurs qui ont un accident tous les 5 ans en moyenne. Soit  $B$  l'événement “être un bon conducteur” et  $M$  l'événement “être un mauvais conducteur”. Supposons qu'il y a autant de bons que de mauvais conducteurs, si bien que  $P(B) = P(M) = \frac{1}{2}$ .*

- (a) *On modélise le temps en années jusqu'au prochain accident d'un bon (resp. mauvais) conducteur par une variable aléatoire exponentielle de paramètre  $\lambda_B$  (resp.  $\lambda_M$ ). Quelles sont les valeurs que l'on doit prendre pour  $\lambda_B$  et  $\lambda_M$  ?*

- (b) On suppose ou on démontre que le nombre de sinistres pendant un temps  $t$  suit une loi de Poisson de paramètre  $\lambda t$  si le temps entre deux sinistres suit une loi exponentielle de paramètre  $\lambda$ . De plus dans toute cette partie on suppose que le coût d'un sinistre est déterministe égal à  $c$ . A l'aide des probabilités totales calculer la prime pure qu'un assureur pourrait faire payer à un nouvel assuré pour un an, c'est-à-dire  $\pi = cE(N)$  où  $N$  est le nombre de sinistres subis par l'assuré en 1 an.
- (c) Sachant que l'assuré n'a pas eu d'accident la première année, calculer la nouvelle probabilité qu'il soit un bon conducteur :  $P(B|N=0)$ .
- (d) Calculer  $P(B|N=k)$  pour  $k = 1, 2, 3, 4$ .
- (e) Comparez ces probabilités à  $P(B)$ . Qu'observez-vous ? Comment pouvez-vous l'expliquer ?

Après cet exercice, voici les propriétés les plus importantes des lois Poisson mélange.

**Proposition I.6 Identification des lois Poisson-mélange**

Soit  $N_1$  et  $N_2$  deux variables aléatoires suivant des lois Poisson mélange de paramètres respectifs  $(\lambda, \Theta_1)$  et  $(\lambda, \Theta_2)$ .  $N_1$  et  $N_2$  ont même loi équivaut alors à  $\Theta_1$  et  $\Theta_2$  ont même loi.

*Preuve :*

La démonstration de cette proposition est immédiate d'après la proposition I.9 et le fait que la fonction génératrice suffit à caractériser une loi.

□

**Proposition I.7 Moyenne et variance des lois Poisson-mélange**

Soit  $N$  une variable aléatoire suivant une loi Poisson mélange de paramètres  $(\lambda, \Theta)$ . Alors

$$E(N) = \lambda \quad \text{et} \quad \text{Var}(N) = \lambda(1 + \lambda \text{Var}(\Theta)).$$

*Preuve :*

Démonstration directe avec la proposition générale I.5 et la moyenne et la variance d'une loi de Poisson.

□

Le mélange augmente donc la variance, qui, contrairement au modèle de Poisson (sans mélange), devient strictement plus grande que la moyenne, ce qui correspond au phénomène de **surdispersion**. La proposition précédente nous dit qu'après mélange les valeurs prises par la variable aléatoire sont en moyenne plus éloignées de la moyenne qu'avant le mélange. Dans le cas des lois de Poisson mélange, le théorème suivant nous donne encore plus d'information : les masses de probabilité sont augmentées par le mélange en-dehors d'un intervalle donné par deux valeurs  $t_1 < t_2$ , et diminuent pour les valeurs comprises entre  $t_1$  et  $t_2$ .

**Proposition I.8 Théorème des deux croisements de Shaked (1980)**

Soit  $N$  une variable aléatoire suivant une loi Poisson mélange de paramètres  $(\lambda, \Theta)$ . Alors il existe deux entiers  $0 \leq k_1 < k_2$  tels que

$$P(N = n) \geq e^{-\lambda} \frac{(\lambda)^n}{n!} \quad \text{pour } n \leq k_1 \text{ et pour } n > k_2$$

et

$$P(N = n) \leq e^{-\lambda} \frac{(\lambda)^n}{n!} \quad \text{pour } k_1 < n \leq k_2.$$

*Preuve :*

$$\begin{aligned} c(k) &= \frac{P(N = k)}{e^{-\lambda} \lambda^k / k!} - 1 \\ &= \int_0^{+\infty} e^{\lambda - \theta} \left( \frac{\theta}{\lambda} \right)^k f_\Theta(\theta) d\theta - 1 \end{aligned}$$

$c$  est convexe en  $k$  comme barycentre de fonctions convexes, car pour tout  $x > 0$ , la fonction

$$\alpha \rightarrow x^\alpha = e^{\alpha \ln(x)}$$

est convexe.  $c$  peut donc s'annuler et changer de signe au plus deux fois.  $c$  doit avoir au moins un changement de signe, sinon on aurait  $E(N) > \lambda$  ou  $E(N) < \lambda$ . Elle ne peut pas en avoir un seul car  $c(0) = e^\lambda - 1 > 0$  et  $\lim_{n \rightarrow +\infty} c(n) = +\infty$ .

□

**Proposition I.9 Mélange et fonctions génératrices**

Soit  $N$  une variable aléatoire suivant une loi Poisson mélange de paramètres  $(\lambda, \Theta)$ ,  $G_N(\cdot) = E(\cdot^N)$  la fonction génératrice de  $N$ , et  $L_\Theta(\cdot) = E(e^{-\Theta})$  la transformée de Laplace de  $\Theta$ . Alors pour  $x \geq 0$ ,

$$G_N(x) = L_\Theta(\lambda(1-x)).$$

*Preuve :*

Il suffit de se souvenir que pour  $x \geq 0$ , si  $N'$  suit une loi de Poisson de paramètre  $\lambda'$ ,

$$G_{N'}(x) = e^{\lambda'(x-1)}$$

pour obtenir en conditionnant par  $\Theta$  le résultat souhaité :

$$G_N(x) = \int_{\theta \in A} e^{\lambda \theta (x-1)} f_\Theta(\theta) d\theta = L_\Theta(\lambda(1-x)).$$

□

**Proposition I.10** Soit  $N$  une variable aléatoire suivant une loi Poisson mélange de paramètres  $\lambda$  et

$$\Theta \sim \Gamma(\alpha, \alpha).$$

Alors

$$N \sim \mathcal{P}(\lambda\Theta)$$

suit une loi binômiale négative :

$$N \sim \mathcal{BN}\left(\alpha, \frac{\alpha}{\alpha + \lambda}\right).$$

Preuve :

Exercice : calculer la fonction génératrice des probabilités de  $N$ , et reconnaître celle d'une loi binômiale négative.

□

**Exercice I.2** Soit  $N_1$  et  $N_2$  deux variables aléatoires indépendantes suivant des lois Poisson-mélange de paramètres respectifs  $(\lambda_1, \Theta_1)$  et  $(\lambda_2, \Theta_2)$ . Alors  $N_1 + N_2$  suit une loi Poisson-mélange de paramètres

$$\left(\lambda_1 + \lambda_2, \frac{1}{\lambda_1 + \lambda_2}(\lambda_1\Theta_1 + \lambda_2\Theta_2)\right).$$

### 2.6.3 Mélange de lois exponentielles

On peut aussi obtenir la loi de Pareto de deuxième espèce comme mélange de lois exponentielles :

**Proposition I.11** Soit

$$X \sim \text{Exp}(\Theta)$$

avec

$$\Theta \sim \Gamma(\alpha, t).$$

Alors pour  $x \geq 0$ ,

$$P(X > x) = \left(\frac{t}{x+t}\right)^\alpha,$$

i.e.  $X$  suit une loi de Pareto de deuxième espèce de paramètres  $t$  et  $\alpha$ .

Preuve :

Exercice : il suffit d'écrire la formule classique du mélange.

□

**Proposition I.12** La fonction de queue  $\bar{F}$  de  $X \sim \text{Exp}(\Theta)$  est donnée au point  $x \geq 0$  par

$$\bar{F}(x) = P(X > x) = \int_0^{+\infty} e^{-\theta x} f_\Theta(\theta) d\theta = E(e^{-x\Theta})$$

et correspond donc à la transformée de Laplace de la variable aléatoire positive  $\Theta$ .

### 2.6.4 Lois composées mélange

Les lois composées mélange sont juste des lois mélange pour lesquelles la loi sous-jacente est une loi composée quelle que soit la valeur du paramètre de mélange  $\theta$ . On a donc exactement les mêmes propriétés que précédemment. Voyons ce que cela donne pour les loi Poisson-composées mélange.

#### Definition I.14 Lois Poisson-composées mélange

*Soit  $\Theta$  une variable aléatoire (de fonction de répartition  $F_\Theta$ ) et  $A \subset ]0, +\infty[$  tel que  $P(\Theta \in A) = 1$ . Soit  $W$  une variable aléatoire. On dit que  $S$  suit une loi Poisson-composée mélange de paramètres  $(\lambda, W, \Theta)$  (avec  $\Theta$  comme loi de mélange) si  $E(\Theta) = 1$  et si pour tout  $x \in \mathbb{R}$ , pour tout  $\theta \in A$ ,*

$$P(S \leq x \mid \Theta = \theta) = P\left(\sum_{i=1}^{N_\theta} W_i \leq x\right),$$

*où  $N_\theta$  suit une loi de Poisson de paramètre  $\lambda\theta$ , et où les  $(W_i)_{i \geq 1}$  forment une suite de variables aléatoires indépendantes, identiquement distribuées, de même loi que  $W$ , et indépendantes de  $N_\theta$ , et avec la convention que la somme ci-dessus est nulle si  $N_\theta = 0$ .*

#### Proposition I.13 Moyenne et variance des lois Poisson-composées mélange

*Soit  $S$  une variable aléatoire suivant une loi Poisson-composée mélange de paramètres  $(\lambda, W, \Theta)$ , avec  $W$  et  $\Theta$  de carré intégrable. Alors*

$$E(S) = \lambda E(W) \quad \text{et} \quad \text{Var}(S) = \lambda E(W^2) + (\lambda E(W))^2 \text{Var}(\Theta).$$

*Preuve :*

Démonstration directe avec la proposition générale I.5 et la moyenne et la variance d'une loi Poisson-composée.

□

La variance d'une loi Poisson-composée mélange peut se réécrire sous la forme suivante

$$\text{Var}(S) = \lambda \text{Var}(W) + \lambda [E(W)]^2 + (\lambda E(W))^2 \text{Var}(\Theta).$$

Exercice : interpréter cette décomposition de la variance de la charge sinistre globale.

Remarque : on aurait pu introduire de l'hétérogénéité dans la fréquence et dans le coût des sinistres, en définissant une suite de variables aléatoires  $(W_\theta)_{\theta \in A}$ . Toutefois, dans la littérature, il n'y a en général pas de mélange sur les coûts dans ce qui est appelé loi Poisson-composée mélange.

#### Proposition I.14 Convergence en loi

*Soit  $(S_\lambda)_{\lambda > 0}$  une collection de variables aléatoires telle que pour  $\lambda > 0$ ,  $S_\lambda$*

suit une loi Poisson-composée mélange de paramètres  $(\lambda, W, \Theta)$ . Supposons de plus que  $W$  et  $\Theta$  sont de carré intégrable. Alors

$$\frac{S_\lambda}{\lambda} \rightarrow \Theta \quad \text{en loi}$$

quand  $\lambda \rightarrow +\infty$ .

*Preuve :*  
exercice.

□

Notons que contrairement au cas sans mélange, on n'a pas de convergence vers une loi normale (éventuellement en renormalisant) avec un argument du type théorème central limite, mais vers la loi de mélange elle-même.

## Chapitre 3

# Approximation du modèle individuel par le modèle collectif

En séances de TD seront abordées diverses majorations de l'erreur faite en approchant le modèle individuel par le modèle collectif. On peut se référer par exemple à Charpentier et Denuit (2004) page 285.

## Chapitre 4

# Compléments sur la charge sinistre

### 4.1 Normal Power, Gamma de Bowers

Voir exercice en séance de TD, et par exemple Partrat et Besson (2004) pages 542 et suivantes, Charpentier et Denuit (2004) page 210, ou le polycopié de Pierre Théron sur le modèle collectif.

### 4.2 FFT

Voir exercice en séance de TD et par exemple page 141 de Rolski et al. (1999).

## **Deuxième partie**

# **Processus de Poisson**

Ce chapitre est utile pour le cours de modèle de durées de Frédéric Planchet, l'étude du modèle à chocs communs (voir partie 30), les modèles structurels et à intensité en théorie du risque de crédit, et bien sûr le cours de théorie de la ruine (voir partie III).

## Chapitre 5

# Rappels autour de la loi exponentielle

### 5.1 Définition et premières propriétés

**Definition II.1** On dit qu'une variable aléatoire  $X$  suit la loi exponentielle de paramètre  $\lambda > 0$  si sa fonction de répartition est donnée pour  $x \in \mathbb{R}$  par

$$F_X(x) = (1 - e^{-\lambda x})1_{\mathbb{R}^+}(x).$$

On utilisera alors la notation  $X \sim \text{Exp}(\lambda)$ .

**Proposition II.1** Soit  $X \sim \text{Exp}(\lambda)$ . Alors,  $X$  admet pour densité  $f_X$  donné pour  $x \in \mathbb{R}$  par

$$f_X(x) = \lambda e^{-\lambda x}1_{\mathbb{R}^+}(x).$$

De plus,  $X$  est de carré intégrable et

$$E(X) = \frac{1}{\lambda} \quad \text{et} \quad E(X^2) = \frac{1}{\lambda^2}.$$

*Preuve :*

On obtient les deux premiers moments par une simple intégration par parties.

□

La proposition suivante sera très utilisée en théorie de la ruine, car elle permettra de dire qu'un processus de Poisson composé perd sa mémoire à un instant bien choisi, ou d'utiliser des méthodes de martingales lorsque les coûts de sinistres suivent une loi exponentielle (voir partie III).

**Proposition II.2** La loi exponentielle possède la propriété de perte de mémoire : pour tous  $s, t \geq 0$

$$P(X > t + s \mid X > s) = P(X > t).$$

On peut démontrer que la loi exponentielle est la seule loi continue sur  $\mathbb{R}^+$  à avoir cette propriété.

**Théorème II.1** *La seule loi continue sur  $\mathbb{R}^+$  à vérifier la propriété de perte de mémoire est la loi exponentielle. La seule loi à valeurs dans  $\mathbb{N}^*$  à vérifier la propriété de perte de mémoire est la loi géométrique.*

*Preuve :*

Exercice, voir [Rolski et al.] (1999).

□

## 5.2 Minimum de variables aléatoires exponentielles indépendantes

Soit  $n \geq 1$  et  $X_1, \dots, X_n$  des variables aléatoires indépendantes telles que pour  $1 \leq i \leq n$ ,

$$X_i \sim \text{Exp}(\lambda_i)$$

avec  $\lambda_i > 0$ . En risque de défaut, cette situation pourrait intéresser un investisseur pouvant acheter des obligations de  $n$  différentes compagnies. Dans les modèles à intensité (voir chapitre 26.3 sur le risque de crédit), on modélise le temps au bout duquel la  $i$ -ème compagnie fait défaut par le premier instant de saut d'un processus de Poisson non homogène (qui sera défini chapitre 6). Les variables aléatoires  $X_i$  ne sont dans ces modèles en général pas indépendantes. Toutefois, afin d'interpréter les résultats suivants, on peut s'intéresser au cas particulier de  $n$  processus de Poisson homogènes indépendants, ce qui nous donne (comme nous le verrons dans la partie 6)  $n$  premiers instants de sauts indépendants et suivant des lois exponentielles. Il existe des produits financiers du type *first-to-default swap* dont le payoff dépend de la survenance ou non d'un événement défavorable avant une date  $t$  sur un des  $n$  actifs. Un investisseur serait entre autres intéressé par connaître la loi du premier temps de défaut, c'est-à-dire de

$$T = \min(X_1, \dots, X_n),$$

et par savoir lequel des actifs risque de faire défaut en premier. On peut répondre très facilement à ces deux questions pour des variables aléatoires indépendantes et de lois exponentielles.

**Proposition II.3** *Soit  $n \geq 1$  et  $X_1, \dots, X_n$  des variables aléatoires indépendantes telles que pour  $1 \leq i \leq n$ ,*

$$X_i \sim \text{Exp}(\lambda_i).$$

*Alors*

$$T = \min_{i=1}^n (X_i) \sim \text{Exp} \left( \sum_{i=1}^n \lambda_i \right).$$

*Preuve :*

$$P(T > t) = P(X_1 > t, \dots, X_n > t) = \prod_{i=1}^n P(X_i > t) = e^{-(\sum_{i=1}^n \lambda_i)t},$$

ce qui montre que

$$T \sim \text{Exp} \left( \sum_{i=1}^n \lambda_i \right).$$

□

Quel actif fait défaut en premier ?

Considérons le cas  $n = 2$ .

$$\begin{aligned} P(X_1 < X_2) &= \int_0^{+\infty} f_{X_1}(s) P(X_2 > t + s \mid X_1 = s) \\ &= \int_0^{+\infty} \lambda_1 e^{-\lambda_1 s} e^{-\lambda_2 s} ds \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2}, \end{aligned}$$

car

$$\int_0^{+\infty} (\lambda_1 + \lambda_2) e^{-(\lambda_1 + \lambda_2)s} ds = 1.$$

On montre de la même manière la proposition suivante :

**Proposition II.4** Pour  $n \geq 1$ , et pour  $1 \leq i \leq n$ , la probabilité que  $X_i$  soit le plus petit est donnée par

$$P(\min(X_1, \dots, X_n) = X_i) = \frac{\lambda_i}{\lambda_1 + \dots + \lambda_n}.$$

Soit  $I$  la variable aléatoire à valeurs dans  $[1, n]$ , définie par

$$\{I = i\} = \{\min(X_1, \dots, X_n) = X_i\}.$$

Supposons que  $n = 2$  et que  $\lambda_1 = 100000$  et que  $\lambda_2 = 0.00001$  (ce qui correspond à  $EX_1 = 1/100000$  et  $EX_2 = 100000$ ). On pourrait penser que si  $\min(X_1, X_2)$  était égal à 200000, la probabilité que  $I = 2$  serait plus élevée que si  $\min(X_1, X_2)$  était égal à 0.00002. En fait, ce raisonnement est faux, et au contraire, on a le résultat suivant :

**Proposition II.5** Les variables aléatoires  $I$  et

$$T = \min(X_1, \dots, X_n)$$

sont indépendantes.

*Preuve :*

Calculons la densité jointe de  $I$  et  $T$  (où la densité  $I$  correspond à la densité discrète) : pour  $1 \leq i \leq n$  et  $t > 0$ ,

$$\begin{aligned} f_{(I,T)}(i, t) &= f_{X_i}(t)P(\forall j \neq i, X_j > t) \\ &= \lambda_i e^{-\lambda_i t} \cdot \prod_{j \neq i} e^{-\lambda_j t} \\ &= \frac{\lambda_i}{\lambda_1 + \dots + \lambda_n} \cdot (\lambda_1 + \dots + \lambda_n) e^{-(\lambda_1 + \dots + \lambda_n)t} \\ &= f_I(i) \cdot f_T(t) \end{aligned}$$

d'après les propositions II.3 et II.4

□

### 5.3 Lois exponentielles multivariées

En risque de crédit, les instants de défaut de différentes entreprises sont a priori loin d'être indépendants. On s'intéresse donc naturellement aux lois multivariées dont les marginales sont exponentielles. Une fois les lois marginales fixées, on obtient la loi jointe par l'intermédiaire d'une **copule**. Cette notion sera abordée en détail dans les chapitres 28.1 à 28.5. Les lois dont les marginales sont des lois exponentielles sont souvent appelées lois de Marshall-Olkin, ce qui peut aussi concerter uniquement une partie de ces lois. On donnera ici trois exemples de familles de lois dont les lois marginales sont des lois exponentielles, et qui fournissent autant d'exercices de maniement sur les lois exponentielles et sur les lois jointes :

- les lois bivariées de Marshall-Olkin : ce sont celles des couples

$$X = (X_1, X_2) = (\min(Y_1, Y), \min(Y_2, Y)),$$

où  $Y_1$ ,  $Y_2$  et  $Y$  sont des variables aléatoires indépendantes de lois exponentielles de paramètres respectifs  $\lambda_1$ ,  $\lambda_2$  et  $\lambda$ . D'après la proposition II.3  $X_i \sim \text{Exp}(\lambda_i + \lambda)$  pour  $i = 1, 2$ , et de plus

$$\min(X_1, X_2) \sim \text{Exp}(\lambda_1 + \lambda_2 + \lambda)$$

puisque

$$\min(X_1, X_2) = \min(X_1, X_2, X).$$

Sa fonction de queue de distribution bivariée est donnée par

$$\bar{F}_X(x_1, x_2) = P(X_1 > x_1, X_2 > x_2) = \exp(-(\lambda_1 x_1 + \lambda_2 x_2 + \lambda \min(x_1, x_2)))$$

pour  $x_1, x_2 > 0$ . En particulier, pour  $h > 0$ , et pour  $x_1, x_2 > 0$ ,

$$P(X_1 > x_1 + h, X_2 > x_2 + h \mid X_1 > x_1, X_2 > x_2) = P(X_1 > h, X_2 > h),$$

ce qui correspond à une propriété de perte de mémoire pour une loi bivariée. La notion de perte de mémoire n'est ici plus unique, nous en verrons une autre définition possible avec la famille de lois suivante. Ce cadre peut être étendu aux lois multivariées en général, et il est possible d'obtenir explicitement la forme de la fonction copule associée (voir chapitre 28.1).

- les lois exponentielles multivariées de Gumbel : une autre façon de définir la perte de mémoire est de le faire uniquement sur les accroissements d'une des marginales : pour  $h > 0$ , et pour  $x_1, x_2 > 0$ ,

$$P(X_1 > x_1 + h \mid X_1 > x_1, X_2 > x_2) = P(X_1 > h \mid X_2 > x_2),$$

et

$$P(X_2 > x_2 + h \mid X_1 > x_1, X_2 > x_2) = P(X_2 > h \mid X_1 > x_1).$$

Ceci est vérifié pour les lois bivariées de Gumbel, définies par la fonction de répartition jointe

$$F_X(x_1, x_2) = P(X_1 > x_1, X_2 > x_2) = 1 - e^{-x_1} - e^{-x_2} + e^{-x_1-x_2-\theta x_1 x_2}.$$

L'unique paramètre est  $\theta \in [0, 1]$  et aura un impact à la fois sur les marginales et sur la dépendance.

- les lois exponentielles de Basu-Block, sont données par

$$\begin{aligned} \bar{F}(x_1, x_2) &= \exp(-\lambda_1 x_1 - \lambda_2 x_2 - \lambda \max(x_1, x_2)) \\ &+ \frac{\lambda}{\lambda_1 + \lambda_2} [\exp(-\lambda_1 x_1 - \lambda_2 x_2 - \lambda \max(x_1, x_2)) - \exp(-(\lambda_1 + \lambda_2 + \lambda \max(x_1, x_2)))] \end{aligned}$$

pour  $x_1, x_2 > 0$ . Comme les lois de Marshall-Olkin bivariées, elles sont définies par trois paramètres  $\lambda_1, \lambda_2$  et  $\lambda > 0$ , et vérifient la propriété suivante :

$$\min(X_1, X_2) \sim \text{Exp}(\lambda_1 + \lambda_2 + \lambda)$$

(ce qui est aussi vérifié par les lois de Marshall-Olkin).

## 5.4 Sommes de variables aléatoires exponentielles indépendantes

Nous verrons dans le chapitre 6 que la loi du temps entre deux sauts d'un processus de Poisson homogène d'intensité  $\lambda > 0$  est la loi exponentielle de paramètre  $\lambda$ . En théorie de la ruine, chaque saut d'un processus de Poisson correspond à un sinistre (accident de voiture, incendie, ...) que la compagnie d'assurances va devoir indemniser. Si on modélise par  $X_1$  l'instant de survenance du premier sinistre, puis par  $X_2$  le temps écoulé entre le premier et le deuxième sinistre, et plus généralement par  $X_i$  le

temps écoulé entre le  $i$ -ème et le  $(i+1)$ -ième sinistre, et si l'on suppose que les  $(X_i)_{i \geq 1}$  sont des variables aléatoires indépendantes, identiquement distribuées et de loi exponentielle de paramètre  $\lambda$ , quelle est la loi de la date de survenance du  $n$ -ème sinistre ?

**Proposition II.6** Soit  $n \geq 1$  et  $X_1, \dots, X_n$  des variables aléatoires indépendantes, identiquement distribuées et de loi exponentielle de paramètre  $\lambda > 0$ . Alors

$$S_n = X_1 + \dots + X_n$$

suit une loi Gamma de paramètres  $(n, \lambda)$ , de densité

$$f_{S_n}(x) = \frac{x^{n-1} \lambda^n e^{-\lambda x}}{(n-1)!} 1_{\mathbb{R}^+}(x).$$

Preuve :

La démonstration se fait par récurrence sur  $n$ .

□

**Proposition II.7** Sous les hypothèses de la proposition précédente,

$$E(S_n) = \frac{n}{\lambda} \text{ et } \text{Var}(S_n) = \frac{n}{\lambda^2}.$$

Preuve :

Immédiat d'après la proposition II.1, car les  $X_i$  sont i.i.d..

□

On parle également de loi d'Erlang de paramètres  $(n, \lambda)$ . Les lois d'Erlang correspondent à la sous-famille des lois Gamma  $(\alpha, \lambda)$  pour lesquelles  $\alpha \in \mathbb{N}^*$ . Rappelons que dans le cas général, une variable aléatoire  $X$  suit une loi Gamma de paramètres  $\alpha$  et  $\lambda$  strictement positifs si elle admet pour densité :

$$f_X(x) = \frac{x^{\alpha-1} \lambda^\alpha e^{-\lambda x}}{\Gamma(\alpha)} 1_{\mathbb{R}^+}(x), \quad (5.1)$$

où, pour  $\alpha > 0$ ,

$$\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt.$$

On peut définir des lois Gamma multivariées à partir de variables aléatoires exponentielles indépendantes. La famille des lois de Cherian est définie par

$$(X_1, X_2) = (Y_1 + Y, Y_2 + Y),$$

où  $Y_1$ ,  $Y_2$  et  $Y$  sont des variables aléatoires indépendantes exponentielles de paramètres respectifs  $\lambda_1$ ,  $\lambda_2$  et  $\lambda$ . Les lois marginales sont des lois Gamma( $2, \lambda$ ) si  $\lambda_1 = \lambda_2 = \lambda$ . Dans le cas général, les lois marginales ne sont pas des lois Gamma, mais des lois *phase-type*, qui correspondent

au temps d'atteinte d'un état absorbant par un processus de Markov à nombre d'états finis. La densité jointe est alors donnée sur  $\mathbb{R}^+ \times \mathbb{R}^+$  par

$$f_X(x_1, x_2) = \frac{e^{-(x_1+x_2)}}{\Gamma(\lambda_1)\Gamma(\lambda_2)\Gamma(\lambda)} \int_0^{\min(x_1, x_2)} (x_1 - t)^{\lambda_1 - 1} (x_2 - t)^{\lambda_2 - 1} t^{\lambda - 1} dt.$$

Les sommes de variables aléatoires exponentielles indépendantes vérifient également des propriétés conditionnelles liées aux statistiques d'ordre, ce que nous verrons au chapitre suivant (voir en particulier le théorème II.3).

## Chapitre 6

# Processus de Poisson : définition et premières propriétés

### 6.1 Processus de Poisson homogène

Il y a plusieurs manières équivalentes de définir un processus de Poisson homogène. Nous adopterons celle qui permet de construire facilement les trajectoires du processus, et qui se généralise aux processus dits de renouvellement (voir chapitre [11]). Néanmoins, la seconde, qui sera énoncée dans le théorème [II.2], se généralisera plus facilement aux processus de Poisson inhomogènes (voir chapitre [6.2]) et aux processus de Poisson en dimensions supérieures.

#### **Definition II.2 Processus de Poisson homogène**

*Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables aléatoires indépendantes, identiquement distribuées, et de loi exponentielle de paramètre  $\lambda > 0$ . Pour  $n \geq 1$ , notons  $S_n = X_1 + X_2 + \dots + X_n$ , et  $S_0 = 0$ . Le processus  $(N(t))_{t \geq 0}$ , défini par*

$$N(t) = \sum_{n \geq 1} \mathbf{1}_{\{S_n \leq t\}} = \text{nombre de } S_n \text{ entre } 0 \text{ et } t = \sup\{n, S_n \leq t\}$$

*est appelé processus de Poisson (homogène) d'intensité  $\lambda$ .*

**Remarque :** Rappelons que, d'après la proposition [II.6],  $S_n$  suit une loi Gamma de paramètres  $n$  et  $\lambda$ ,  $\Gamma(n, \lambda)$ , de densité donnée par

$$f_{S_n}(s) = e^{-\lambda s} \lambda^n \frac{s^{n-1}}{(n-1)!} \mathbf{1}_{\mathbb{R}^+}(s) ds.$$

Par des intégrations par parties successives, on peut montrer que la fonction de répartition de  $S_n$  est donnée pour  $x \in \mathbb{R}$  par

$$F_{S_n}(x) = 1 - \sum_{k=0}^{n-1} e^{-\lambda x} \frac{(\lambda x)^k}{k!} 1_{\mathbb{R}^+}(x). \quad (6.1)$$

Remarquons également que

$$\{N(t) = n\} = \{S_n \leq t < S_{n+1}\} \quad (6.2)$$

et que

$$\{N(t) \geq n\} = \{S_n \leq t\}. \quad (6.3)$$

Si l'on reprend l'exemple de la page 38, dans lequel  $X_i$  représente le temps entre le  $i$ -ème et le  $(i+1)$ -ème sinistre,  $S_n$  représente la date d'occurrence du  $n$ -ème sinistre, et  $N(t)$  représente le nombre de sinistres survenus avant la date  $t$ . L'équation (6.2) dit que le nombre de sinistres survenus avant  $t$  est égal à  $n$  si et seulement si le  $n$ -ème sinistre a eu lieu avant  $t$ , et le  $n+1$ -ème après  $t$ . L'équation (6.3) dit simplement que le nombre de sinistres survenus avant  $t$  est supérieur ou égal à  $n$  si et seulement si le  $n$ -ème sinistre a eu lieu avant  $t$ . En utilisant les équations (6.1) et (6.3), on comprend maintenant pourquoi le processus est dit de Poisson et non exponentiel. En effet, pour  $n \geq 0$  et  $t > 0$ ,

$$\begin{aligned} P(N(t) = n) &= P(N(t) \geq n) - P(N(t) \geq n+1) \\ &= P(S_n \leq t) - P(S_{n+1} \leq t) \\ &= \left[ 1 - \sum_{k=0}^{n-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \right] \\ &\quad - \left[ 1 - \sum_{k=0}^n e^{-\lambda t} \frac{(\lambda t)^k}{k!} \right] \\ &= e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \end{aligned}$$

ce qui montre le lemme suivant :

**Lemme II.1** Soit  $N(t)$  un processus de Poisson homogène de paramètre  $\lambda$ . Pour tout  $t > 0$ ,  $N(t)$  suit une loi de Poisson de paramètre  $\lambda t$ .

Rappelons quelques propriétés importantes de la loi de Poisson qui vont nous servir (voir le chapitre 2 pour les démonstrations) :

**Lemme II.2** Soit  $X$  une variable aléatoire suivant une loi de Poisson de paramètre  $\lambda > 0$ .

– Sa fonction génératrice est donnée pour  $t > 0$  par

$$G_X(t) = E(t^X) = e^{-\lambda(1-t)}.$$

- Ses deux premiers moments valent

$$E(X) = \text{Var}(X) = \lambda.$$

Cette propriété, dite d'**équidispersion** (moyenne = variance) est très importante, car elle est relativement facile à vérifier empiriquement, et permettra de confirmer ou de mettre en doute la pertinence du modèle Poisson-composé (voir chapitres 2 et 7).

- Si  $Y$  suit une loi de Poisson de paramètre  $\mu > 0$  et est indépendante de  $X$ , alors  $X + Y$  suit une loi de Poisson de paramètre  $\lambda + \mu$ .

### Definition II.3 Processus à accroissements stationnaires, à accroissements indépendants

- On dit qu'un processus  $(X(t))_{t \geq 0}$  est à accroissements indépendants si pour tout  $n \geq 2$ , pour tous  $0 \leq t_1 < t_2 < \dots < t_n$ , les variables aléatoires

$$X(t_1), X(t_2) - X(t_1), \dots, X(t_n) - X(t_{n-1})$$

sont indépendantes.

- On dit qu'un processus  $(X(t))_{t \geq 0}$  est à accroissements stationnaires si pour tout  $n \geq 2$ , pour tous  $0 \leq t_1 < t_2 < \dots < t_n$  et pour  $h \geq 0$ , la loi de

$$(X(t_2 + h) - X(t_1 + h), \dots, X(t_n + h) - X(t_{n-1} + h))$$

ne dépend pas de  $h$ .

Remarquons que si  $(X(t))_{t \geq 0}$  est à accroissements indépendants, alors  $X(t)$  est à accroissements stationnaires si et seulement si pour tout  $s > 0$ , la loi de  $X(t+s) - X(t)$  ne dépend pas de  $t$ , et est donc la même que celle de  $X(s) - X(0)$ .

Nous pouvons maintenant énoncer le théorème II.2 qui contient une deuxième manière de définir un processus de Poisson et qui montre au passage qu'elle est bien équivalente à la définition II.2 que nous avions choisie.

### Théorème II.2 Propriétés caractéristiques d'un processus de Poisson (1)

(a)  $N(0) = 0$  presque sûrement.

(b)  $\forall t > s$ ,  $N(t) - N(s)$  suit une loi de Poisson de paramètre  $\lambda(t-s)$  (en particulier,  $(N(t))_{t \geq 0}$  est un processus à accroissements stationnaires).

(c)  $(N(t))_{t \geq 0}$  est à accroissements indépendants.

Réiproquement, tout processus  $(N(t))_{t \geq 0}$  qui vérifie les points 1 à 3 est un processus de Poisson homogène d'intensité  $\lambda$ .

Preuve :

Exercice (voir [Durrett (1999) page 132 et [Rolski et al. (1999)]).

□

Ce théorème explique la dénomination *processus de Poisson*. Néanmoins, il ne s'agit pas des seules propriétés caractéristiques du processus de Poisson homogène. Les deux autres caractéristiques principales sont la répartition uniforme des sauts une fois leur nombre connu, et le fait que lorsque  $\Delta t \rightarrow 0$ , la probabilité qu'il y ait un saut dans l'intervalle de temps  $[0, \Delta t]$  est équivalente à  $\lambda \Delta t$  et la probabilité qu'il n'y ait pas de saut à  $1 - \lambda \Delta t$ . Avant de résumer tout cela dans le théorème **II.3**, rappelons la définition et les propriétés des statistiques d'ordre, ou lois de Dirichlet, dont nous allons avoir besoin.

**Definition II.4** Soit  $n \geq 1$  et  $(U_1, \dots, U_n)$   $n$  variables aléatoires indépendantes uniformément distribuées sur un intervalle fini  $[a, b]$ . Notons  $(V_1, \dots, V_n)$  leur réarrangement dans l'ordre croissant. La loi de  $(V_1, \dots, V_n)$  est appelée la **statistique d'ordre  $n$**  ou **loi de Dirichlet d'ordre  $n$  sur  $[a, b]$** , et notée  $D_n([a, b])$ .

**Propriété II.1** Soit

$$(V_1, \dots, V_n) \sim D_n([a, b]).$$

(a)  $(V_1, \dots, V_n)$  admet pour densité sur  $\mathbb{R}^n$   $f_{(V_1, \dots, V_n)}$  donnée par :

$$f_{(V_1, \dots, V_n)}(t_1, \dots, t_n) = \frac{n!}{(b-a)^n} 1_{\{a \leq t_1 < \dots < t_n \leq b\}}.$$

(b) La densité de  $V_n$  est donnée par :

$$f_{V_n}(t) = \frac{n(t-a)^{n-1}}{(b-a)^n} 1_{[a,b]}(t).$$

- (c) Pour tout  $c \in [a, b]$ , la loi conditionnelle de  $(V_1, \dots, V_{n-1})$  sachant que  $V_n = c$  est  $D_{n-1}([a, c])$ .
- (d) Pour tout  $c \in [a, b]$ , la loi conditionnelle de  $(V_1, \dots, V_{n-1})$  sachant que  $V_{n-1} \leq c \leq V_n$  est  $D_{n-1}([a, c])$ .
- (e) Pour tout  $c \in [a, b]$ , la loi conditionnelle de  $(V_2, \dots, V_n)$  sachant que  $V_1 = c$  est  $D_{n-1}([c, b])$ .
- (f) Pour tout  $c \in [a, b]$ , la loi conditionnelle de  $(V_2, \dots, V_n)$  sachant que  $V_1 \leq c \leq V_2$  est  $D_{n-1}([c, b])$ .
- (g) Pour  $1 \leq k \leq p \leq n$ , avec  $p \geq k+2$ ,  $(V_k, \dots, V_p)$  et  $(V_1, \dots, V_k, V_p, \dots, V_n)$  sont conditionnellement indépendants sachant  $(V_k, V_p)$ . De plus, pour  $a \leq c < d \leq b$ , la loi de  $(V_{k+1}, \dots, V_{p-1})$  sachant  $V_k = c$  et  $V_p = d$  est donnée par  $D_{p-k-1}([c, d])$ .

Preuve :

Exercice d'application du cours de probas 2.

□

On peut maintenant énoncer le théorème annoncé, dont les points 3 et 4 montrent que les lois de Dirichlet apparaissent également lorsqu'on conditionne des sommes de variables aléatoires i.i.d. de loi exponentielle.

### Théorème II.3 Propriétés caractéristiques d'un processus de Poisson (2)

*Soit  $N(t)$  un processus à valeurs dans  $\mathbb{N}$ . Les assertions suivantes sont équivalentes :*

(a)  $(N(t))_{t \geq 0}$  est un processus de Poisson homogène d'intensité  $\lambda$ .

(b) –  $(N(t))_{t \geq 0}$  est à accroissements indépendants et stationnaires,  
– et pour tout  $t \geq 0$ ,

$$N(t) \sim Poi(\lambda t).$$

(c) – Pour tout  $t \geq 0$ ,

$$N(t) \sim Poi(\lambda t),$$

– et pour  $n \geq 1$ , sachant que  $N(t) = n$ , le vecteur aléatoire

$$(S_1, \dots, S_n)$$

*des instants de sauts a pour loi  $D_n([0, t])$ , la statistique d'ordre  $n$  sur l'intervalle  $[0, t]$ .*

(d) –  $(N(t))_{t \geq 0}$  est à accroissements indépendants,

–  $E(N(1)) = \lambda$ ,

– et pour  $n \geq 1$ , sachant que  $N(t) = n$ , le vecteur aléatoire

$$(S_1, \dots, S_n)$$

*des instants de sauts a pour loi  $D_n([0, t])$ , la statistique d'ordre  $n$  sur l'intervalle  $[0, t]$ .*

(e) –  $(N(t))_{t \geq 0}$  est à accroissements indépendants et stationnaires,

– et lorsque  $h \downarrow 0$ ,

$$P(N(h) = 0) = 1 - \lambda h + o(h) \quad \text{et} \quad P(N(h) = 1) = \lambda h + o(h). \quad (6.4)$$

*Preuve :*

Exercice (voir [Durrett (1999)] page 132 et [Rolski et al. (1999)]).

□

La dernière propriété des assertions 3 et 4 du théorème II.3 est très importante, et sera expliquée en détail, démontrée et généralisée au chapitre 9. Elle nous permettra entre autres d'écrire l'espérance d'une fonction intégrable  $f$  des instants de sauts  $S_1, \dots, S_{N(T)}$  d'un processus de Poisson homogène  $(N(t))_{t \geq 0}$  jusqu'à un temps fini  $T$  sachant que  $N(T) = n$  sous la forme

$$E(f(S_1, \dots, S_{N(T)})) = \int_0^T \int_0^{t_{n-1}} \dots \int_0^{t_2} f(t_1, \dots, t_n) dt_1 \dots dt_n. \quad (6.5)$$

En particulier, pour un processus de Poisson homogène  $(N(t))_{t \geq 0}$ , sachant que  $N(t) = 1$ , l'instant de saut  $S_1$  est uniformément distribué sur  $[0, t]$ .

La propriété (6.4) sera très utile pour les démonstrations heuristiques de résultats de théorie de la ruine (voir chapitre III), par exemple pour les équations intégro-différentielles (12.1) et (12.2).

La proposition suivante correspond à ce qu'on appelle le *paradoxe de l'inspection*.

**Proposition II.8** (a) Pour tous  $t > 0$ ,  $x \geq 0$ ,  $0 \leq y \leq t$ ,

$$\mathbb{P}(S_{N_{t+1}} - t \geq x, t - S_{N_t} \geq y) = \frac{1}{E(X_1)} \int_{x+y}^{\infty} \mathbb{P}(X_1 \geq u) du.$$

(b) Les variables aléatoires  $S_{N_{t+1}} - t$  et  $t - S_{N_t}$  sont indépendantes.

*Preuve :*

Montrons que 1  $\Rightarrow$  2 : d'après le point 1,  $X_1 \sim \exp(\lambda)$  a pour moyenne  $E(X_1) = \frac{1}{\lambda}$  et densité  $f_{X_1} = \lambda e^{-\lambda u} \mathbf{1}_{\mathbb{R}^+}(u)$ . En remplaçant dans la formule, on obtient :

$$\mathbb{P}(S_{N_{t+1}} - t \geq x, t - S_{N_t} \geq y) = \lambda \int_{x+y}^{\infty} e^{-\lambda u} du = e^{-\lambda(x+y)}, \quad (6.6)$$

ce qui implique que  $S_{N_{t+1}} - t$  et  $t - S_{N_t}$  sont des variables aléatoires indépendantes. Remarquons au passage que

$$E(S_{N_{t+1}} - S_{N_t}) = E(S_{N_{t+1}} - t) + E(t - S_{N_t}) > \frac{1}{\lambda},$$

alors que  $S_{N_{t+1}} - S_{N_t} = X_{N_t}$ .  
Il ne reste plus qu'à montrer la première affirmation.

$$\begin{aligned}
& \mathbb{P}(S_{N_{t+1}} - t \geq x, t - S_{N_t} \geq y) \\
&= \sum_{n \geq 0} \mathbb{P}(N_t = n, S_n + X_{n+1} - t \geq x, t - S_n \geq y) \\
&= \sum_{n \geq 0} \mathbb{P}(S_n < t < S_n + X_{n+1}, S_n + X_{n+1} - t \geq x, t - S_n \geq y) \\
&= \sum_{n \geq 1} \int_0^\infty \int_0^\infty \mathbf{1}_{s < t < s+u} \mathbf{1}_{s+u-t \geq x} \mathbf{1}_{t-s \geq y} \lambda e^{-\lambda u} \lambda^n \frac{s^{n-1} e^{-\lambda s}}{(n-1)!} ds du \\
&\quad + \mathbb{P}(N_t = 0, X_1 \geq t + x, t \geq y) \\
&= \sum_{n \geq 1} \int_0^\infty \mathbf{1}_{s < t} \mathbf{1}_{t-s \geq y} \left( \int_0^\infty \mathbf{1}_{u \geq t-s+x} \lambda e^{-\lambda u} du \right) \lambda^n \frac{s^{n-1} e^{-\lambda s}}{(n-1)!} ds + e^{-\lambda(t+x)} \\
&= \sum_{n \geq 1} \int_0^\infty \mathbf{1}_{s \leq t-y} e^{-\lambda(t-s+x)} \lambda^n \frac{s^{n-1} e^{-\lambda s}}{(n-1)!} ds + e^{-\lambda(t+x)} \\
&= \int_0^{t-y} e^{-\lambda(t+x)} \lambda e^{\lambda s} ds + e^{-\lambda(t+x)} \\
&= e^{-\lambda(t+x)} (-1 + e^{\lambda(t-y)}) + e^{-\lambda(t+x)} \\
&= e^{-\lambda(y+x)} - e^{-\lambda(t+x)} + e^{-\lambda(t+x)} \\
&= e^{-\lambda(y+x)}
\end{aligned}$$

□

In particular from point 1 with  $y = 0$  we get that the time between  $t$  and the next time of accident is distributed as  $\text{Exp}(\lambda)$ . With  $x = 0$  we have that

$$E(t - S_{N_t}) = \frac{1 - e^{-\lambda t}}{\lambda} \rightarrow \frac{1}{\lambda}$$

as  $t \rightarrow +\infty$ . You could then think : as for  $t > 0$ , we have  $E(S_{N_{t+1}} - t) = \frac{1}{\lambda}$  and  $E(t - S_{N_t}) = \frac{1 - e^{-\lambda t}}{\lambda}$ , we should have

$$EX_{N_{t+1}} = E(t - S_{N_t}) + E(S_{N_{t+1}} - t) = \frac{1 - e^{-\lambda t}}{\lambda} + \frac{1}{\lambda}$$

which would be a contradiction with  $EX_{N_{t+1}} = \frac{1}{\lambda}$ . The last assumption is actually not true, because the real experiment that you carry out is to choose a time  $t$  at random, and to look at the times of the previous and next accident. But when you select  $t$  at random, you have a greater probability to choose  $t$  into a large interval between two accidents than in a small interval between two accidents which occur one just after the other. This paradox is known as the **inspection paradox**.

In driving insurance, one simple model for the number  $N_t$  of accidents up to time  $t$  is to take

$$N_t \equiv \mathcal{P}(\lambda t)$$

It could seem natural to choose exponentially-distributed inter-occurrence times due to the memoryless property of the exponential distribution, and in this case, you get a Poisson distribution for  $N_t$ . Maybe the true reason is that computations are much easier for Poisson processes!

For an actuary, given the experience data that he gets, it is quite easy to verify whether the Poisson assumption is realistic or not : indeed for  $N_t \equiv \mathcal{P}(\lambda)$ , we have

$$EN_t = Var(N_t) = \lambda t \text{ (equidispersion property)}$$

With the historical data, a mere computation of the empirical mean and variance of the number of accidents during a given period may help you determine if you are using a realistic model or not. In case  $Var(N_t) >> EN_t$  (over-dispersion), you may use for example a negative binomial distribution instead of the Poisson distribution for the number of claims during a given period. If  $Var(N_t) << EN_t$  (under-dispersion), you may use a binomial distribution for example.

## 6.2 Processus de Poisson non homogène

Si l'on cherche à représenter le nombre d'accidents survenant avant chaque date  $t \geq 0$  par un processus de Poisson homogène  $(N(t))_{t \geq 0}$ , on pourrait nous opposer le fait qu'il y a plus de chance que des accidents surviennent le jour que la nuit, la circulation étant beaucoup plus intense le jour. Une manière de prendre en compte ce phénomène est de ne plus imposer que l'intensité  $\lambda$  soit constante, mais de lui laisser la possibilité de varier au cours du temps, de façon à avoir pour  $t \geq 0$  lorsque  $h \downarrow 0$  :

$$P(N(t+h) - N(t)) = \lambda(t)h + o(h)$$

. On obtient alors un processus de Poisson non homogène :

**Definition II.5** *Un processus stochastique  $(N(t))_{t \geq 0}$  est un processus de Poisson (inhomogène, ou général) de fonction d'intensité  $\lambda(t)$  si*

- (a)  $N(0) = 0$  presque sûrement,
- (b)  $(N(t))_{t \geq 0}$  est un processus à accroissements indépendants,
- (c) et  $\forall t > s$ ,  $N(t) - N(s)$  suit une loi de Poisson de paramètre

$$\int_s^t \lambda(u)du.$$

Le processus  $N(t)$  n'est plus à accroissements stationnaires, et les temps inter-sauts  $X_i$  ne sont plus indépendants, et ne suivent plus une loi exponentielle si  $\lambda(\cdot)$  n'est pas constante. En effet,

$$P(N(t) = 0) = P(X_1 > t) = e^{-\int_0^t \lambda(u)du},$$

donc l'instant du premier saut  $X_1$  a pour densité en  $t$

$$f_{X_1}(t) = \lambda(t)e^{-\mu(t)},$$

où  $\mu$  est appelée la fonction d'intensité cumulée de  $N(t)$ , et est définie pour  $t \geq 0$  par

$$\mu(t) = \int_0^t \lambda(u)du.$$

Remarquons que  $\mu$  est une fonction positive, nulle en zéro, et croissante sur  $\mathbb{R}^+$ .

La densité jointe des  $n$  premiers instants de sauts  $S_1, \dots, S_n$  est donnée pour  $0 \leq t_1 \leq \dots \leq t_n$  par

$$f_{(S_1, \dots, S_n)}(t_1, \dots, t_n) = \lambda(t_1)e^{-\mu(t_1)} \cdot \lambda(t_2)e^{-(\mu(t_2)-\mu(t_1))} \dots \lambda(t_n)e^{-(\mu(t_n)-\mu(t_{n-1}))} = e^{-\mu(t_n)} \cdot \prod_{i=1}^n \lambda(t_i). \quad (6.7)$$

En particulier, la densité jointe des deux premiers temps inter-sauts est donnée pour  $s, t \geq 0$  par

$$f_{(X_1, X_2)}(s, t) = f_{(S_1, S_2)}(s, s+t) = \lambda(s)e^{-\mu(s)} \cdot \lambda(s+t)e^{-(\mu(s+t)-\mu(s))},$$

ce qui montre que  $X_1$  et  $X_2$  ne sont pas indépendants dès que  $\lambda(\cdot)$  n'est pas constante.

## Chapitre 7

# Processus de Poisson composé

Le processus de Poisson peut nous permettre de modéliser les dates de survenance des sinistres. Le processus de Poisson composé nous permet d'associer à chaque sinistre son coût.

**Definition II.6** Soit  $(N(t))_{t \geq 0}$  un processus de Poisson de fonction d'intensité  $\lambda$  et  $W_n$  une suite de variables aléatoires indépendantes et identiquement distribuées, et indépendantes du processus  $(N(t))_{t \geq 0}$ . Le processus  $(S(t))_{t \geq 0}$ , à valeurs dans  $\mathbb{R}$  et défini par

$$S(t) = W_1 + \cdots + W_{N(t)},$$

avec la convention  $S(t) = 0$  si  $N(t) = 0$ , est appelé processus de Poisson composé de caractéristiques  $(\lambda, W)$ .

En assurance non-vie,  $W_i$  représentera le coût du  $i$ -ème sinistre (et sera donc à valeurs dans  $\mathbb{R}^+$ ),  $N(t)$  le nombre de sinistres jusqu'au temps  $t$ , et  $S(t)$  le montant cumulé de tous les sinistres survenus avant la date  $t$ .  $(S(t))_{t \geq 0}$  sera alors un processus croissant et à valeurs dans  $\mathbb{R}^+$ . D'après les propriétés des lois composées et des fonctions génératrices (voir chapitres 2.3 et 2.4.6), rappelons la proposition suivante :

**Proposition II.9** Soit  $(S(t))_{t \geq 0}$  un processus de Poisson composé de caractéristiques  $(\lambda, W_1)$ , et de fonction d'intensité cumulée  $\mu$ . Pour  $r, t > 0$ , si  $G_W$  est la fonction génératrice de  $W_1$ ,

$$h(r, t) = E \left( r^{S(t)} \right) = e^{-\mu(t)(1-G_W(r))},$$

d'où

$$E(S(t)) = \frac{d}{dr} \left[ E \left( r^{S(t)} \right) \right]_{|r=1} = \lambda t G'_W(1) = \mu(t) E(W_1)$$

et

$$Var(S(t)) = \frac{d^2}{dr^2} \left[ E(r^{S(t)}) \right]_{|r=1} = \mu(t)E(W_1^2).$$

*Preuve :*

Exercice de révision.

□

## Chapitre 8

# Propriétés de Markov et martingales

### 8.1 Propriétés de Markov

Tout processus de Poisson homogène est à accroissements indépendants et stationnaires, et vérifie donc la propriété de Markov faible et la propriété de Markov forte. En particulier, si  $(N(t))_{t \geq 0}$  est un processus de Poisson homogène d'intensité  $\lambda$ , alors pour tout  $s \geq 0$ ,  $(N(t+s) - N(s))_{t \geq 0}$  est un processus de Poisson d'intensité  $\lambda$  indépendant de  $(N(u))_{u \leq s}$ . De plus, si  $\tau$  est un temps d'arrêt, alors  $(N(t+\tau) - N(\tau))_{t \geq 0}$  est un processus de Poisson d'intensité  $\lambda$  indépendant de  $(N(u))_{u \leq \tau}$ .

### 8.2 Martingales

**Théorème II.4** Soit  $(N(t))_{t \geq 0}$  un processus de Poisson homogène d'intensité  $\lambda$ . Alors

$$N(t) - \lambda t,$$

et

$$e^{\alpha N(t) - (e^\alpha - 1)\lambda t}$$

sont des martingales par rapport à la filtration naturelle de  $(N(t))_{t \geq 0}$ .

Ce résultat se généralise sans peine aux processus de Poisson inhomogènes et aux processus de Poisson composés.

## Chapitre 9

# Thinning, superposition et conditionnement

### 9.1 Thinning et superposition

**Théorème II.5** Soit  $(S(t))_{t \geq 0}$  un processus de Poisson composé défini à partir d'un processus de Poisson homogène de paramètre  $\lambda$  et d'une suite de v.a. i.i.d.  $(W_i)_{i \geq 1}$ . Soit  $k \geq 1$ , et  $A_1, \dots, A_k$  une partition de  $\mathbb{R}$ . Pour  $1 \leq j \leq k$ , soit  $(N_j(t))_{t \geq 0}$  le processus de comptage défini pour  $t \geq 0$  par

$$N_j(t) = \sum_{i=1}^{N(t)} 1_{\{W_i \in A_j\}},$$

et  $N_j(t) = 0$  si  $N(t) = 0$ .

Alors les  $(N_j(t))_{t \geq 0}$  sont des processus de Poisson homogènes indépendants de paramètres respectifs  $\lambda \cdot P(W_1 \in A_j)$ .

En particulier, en assurance non-vie, en prenant  $A_1 = \{0\}$ ,  $A_2 = ]0, +\infty[$ , et  $A_3 = ]-\infty, 0[$ , comme  $P(W_1 \in A_3) = 0$ , si le processus décrivant le nombre de sinistres (nuls et non nuls) jusqu'au temps  $t$  est un processus de Poisson homogène d'intensité  $\lambda$ , alors celui décrivant le nombre de sinistres non nuls jusqu'au temps  $t$  est un processus de Poisson de paramètre  $\lambda(1 - P(W_1 = 0))$ . Séparer un processus de Poisson d'une telle manière se dit **thinning**. Le contraire, l'addition de processus de Poisson indépendants s'appelle **superposition**.

**Théorème II.6** Soit  $k \geq 2$ , et  $(N_j(t))_{t \geq 0}$ ,  $1 \leq j \leq k$  des processus de Poisson homogènes d'intensité  $\lambda_1, \dots, \lambda_k$ . Alors  $(N(t))_{t \geq 0}$  défini pour  $t \geq 0$  par

$$N(t) = N_1(t) + \dots + N_k(t)$$

est un processus de Poisson d'intensité  $\lambda_1 + \dots + \lambda_k$ .

Ce théorème se généralise sans peine à des processus de Poisson inhomogènes.

## 9.2 Conditionnement

Dans cette section, on s'intéresse à la position des instants de sauts sachant qu'il y en a  $n$  entre 0 et  $T$ . On commence par le cas le plus facile, celui d'un processus homogène, et on généralise ensuite les résultats pour un processus de Poisson inhomogène.

### 9.2.1 Cas d'un processus de Poisson homogène

Soit  $U_1, \dots, U_n, \dots$  des variables aléatoires indépendantes et uniformément distribuées sur un intervalle de temps fini et fixé  $[0, T]$ . Soit  $S_1, \dots, S_n, \dots$  les instants de sauts d'un processus de Poisson homogène d'intensité  $\lambda$ .

**Théorème II.7** *Conditionnellement à  $\{N(T) = n\}$ , l'ensemble des instants de sauts  $\{S_1, \dots, S_n\}$  a la même loi que  $\{U_1, \dots, U_n\}$ .*

En d'autres termes, le vecteur aléatoire  $(S_1, \dots, S_n)$  a la même loi que  $n$ -ème statistique d'ordre sur  $[0, T]$ , i.e. sa densité est donnée pour  $0 \leq t_1 \leq \dots \leq t_n \leq T$  par

$$f_{(S_1, \dots, S_n)}(t_1, \dots, t_n) = \frac{1}{T^n/n!} = \frac{n!}{T^n}.$$

Rappelons que si  $U_1, \dots, U_n$  sont des variables indépendantes uniformément distribuées sur  $[0, T]$ , les  $(V_i)_{1 \leq i \leq n}$  définis à partir des  $U_i$  en les rangeant par ordre croissant forment la  $n$ -ème statistique d'ordre sur  $[0, T]$ . Les  $U_i$  ont pour densité jointe

$$f_{(U_1, \dots, U_n)}(t_1, \dots, t_n) = \frac{1}{T^n} 1_{\{\forall i, 0 \leq t_i \leq T\}},$$

alors que les  $V_i$  ont pour densité

$$f_{(V_1, \dots, V_n)}(t_1, \dots, t_n) = \frac{n!}{T^n} 1_{\{0 \leq t_1 \leq \dots \leq t_n \leq T\}}.$$

Le théorème II.7 permet de démontrer directement le résultat suivant, qui pourrait aussi être obtenu à partir de l'indépendance de  $N(s)$  et de  $N(t) - N(s)$  pour  $s < t$ .

**Théorème II.8** *Soit  $(N(t))_{t \geq 0}$  un processus de Poisson homogène. Pour  $s < t$  et pour tous  $0 \leq m \leq n$ , la loi de  $N(s)$  sachant que  $N(t) = n$  est une loi binomiale de paramètres  $(n, s/t)$  :*

$$P(N(s) = m | N(t) = n) = C_n^m \left(\frac{s}{t}\right)^m \left(1 - \frac{s}{t}\right)^{n-m}.$$

Remarquons que les lois conditionnelles obtenues dans les théorèmes [II.7](#) et [II.8](#) ne dépendent pas de  $\lambda$ . L'homogénéité est synonyme de symétrie, qui est brisée dès que l'intensité n'est plus constante. Les généralisations des deux théorèmes précédents au cas inhomogène feront donc apparaître cette dissymétrie en faisant intervenir la fonction d'intensité et la fonction d'intensité cumulée.

### 9.2.2 Cas d'un processus de Poisson inhomogène

Lorsque  $\lambda(t)$  n'est plus constant, sachant qu'il y a  $n$  sauts, la probabilité qu'un saut soit au voisinage d'un point  $u \in [0, T]$  est d'autant plus élevée que l'intensité  $\lambda(u)$  à ce point est élevée. En fait, on a le même résultat que précédemment en tenant compte proportionnellement de la fonction d'intensité.

**Théorème II.9** Soit  $(N(t))_{t \geq 0}$  un processus de Poisson inhomogène de fonction d'intensité  $\lambda(t)$ , et de fonction d'intensité cumulée  $\mu(t) = \int_0^t \lambda(u)du$ . Pour une date  $T > 0$  fixée, soit  $h_T$  la fonction de  $\mathbb{R}$  dans  $\mathbb{R}$  définie pour  $x \in \mathbb{R}$  par

$$h_T(x) = \frac{\lambda(x)}{\mu(T)} \cdot 1_{[0,T]}(x),$$

et soit  $U_1, \dots, U_n$  des variables aléatoires indépendantes de densité  $h_T$ . Alors, sachant que  $N(t) = n$ , l'ensemble des instants de sauts  $\{S_1, \dots, S_n\}$  a la même loi que l'ensemble des  $\{U_1, \dots, U_n\}$ .

On obtient aussi l'analogie du théorème [II.8](#) :

**Théorème II.10** Soit  $(N(t))_{t \geq 0}$  un processus de Poisson inhomogène de fonction d'intensité  $\lambda(t)$ , et de fonction d'intensité cumulée  $\mu(t) = \int_0^t \lambda(u)du$ . Pour  $s < t$  et pour tous  $0 \leq m \leq n$ , la loi de  $N(s)$  sachant que  $N(t) = n$  est une loi binomiale de paramètres  $(n, \mu(s)/\mu(t))$  :

$$P(N(s) = m \mid N(t) = n) = C_n^m \left( \frac{\mu(s)}{\mu(t)} \right)^m \left( 1 - \frac{\mu(s)}{\mu(t)} \right)^{n-m}.$$

## Troisième partie

### Théorie de la ruine

Soit  $R(t) = u + X(t)$  un processus de risque classique :

–  $(X(t))_{t \geq 0}$  est donc défini pour tout  $t \geq 0$  par

$$X(t) = ct - \sum_{i=1}^{N(t)} W_i,$$

- $(N(t))_{t \geq 0}$  est un processus de Poisson homogène d'intensité  $\lambda$ ,
- les  $W_i$  sont des v.a.i.i.d. positives de moyenne  $\mu$  et indépendantes de  $(N(t))_{t \geq 0}$ ,
- la somme est nulle si  $N(t) = 0$ ,
- le chargement de sécurité relatif

$$\rho = \frac{c - \lambda\mu}{\lambda\mu} > 0.$$

Définissons la probabilité de ruine

$$\psi(u) = P(\exists t > 0, \quad u + X(t) < 0)$$

et la probabilité de non-ruine (ou de survie)

$$\varphi(u) = 1 - \psi(u) = P(\forall t > 0, \quad u + X(t) \geq 0).$$

### Théorème III.1 (Equation intégralo-différentielle)

Pour tout  $u \geq 0$ ,

$$\varphi'_d(u) = \frac{\lambda}{c} \left( \varphi(u) - \int_0^u \varphi(u-y)dF_W(y) \right). \quad (9.1)$$

*Preuve :*

Pour  $h > 0$ ,

$$\varphi(u) = E(\varphi(u + X(T_1 \wedge h))).$$

L'idée est de distinguer les cas  $T_1 \leq h$  (dans ce cas le processus repart après un temps aléatoire  $T_1$  de la position aléatoire  $u + cT_1 - W_1$ ) et  $T_1 > h$  (dans ce cas, le processus repart après un temps  $h$  de la position  $u+ch$ ). La perte de mémoire de la loi exponentielle nous garantit la perte de mémoire du processus de Poisson  $(N(t))_{t \geq 0}$ , et donc du processus  $(X(t))_{t \geq 0}$  qui est *sans mémoire*.

$$\varphi(u) = \varphi(u + ch)P(T_1 > h) + \int_0^h f_{T_1}(t) \int_0^{+\infty} \varphi(u + ct - y)dF_{W_1}(y)dt.$$

En inversant les signes, en rajoutant  $\varphi(u + ch)$  à droite et à gauche, et en divisant par  $h$ , on obtient

$$\frac{\varphi(u + ch) - \varphi(u)}{h} = \varphi(u + ch) \left( \frac{1 - e^{-\lambda h}}{h} \right) - \frac{1}{h} \int_0^h \lambda e^{-\lambda t} \int_0^{+\infty} \varphi(u + ct - y)dF_{W_1}(y)dt.$$

En passant à la limite lorsque  $h \downarrow 0$ , en utilisant la continuité à droite de  $\varphi$  et le fait que

$$\frac{1}{h} \int_0^h g(t)dt \rightarrow g(0)$$

quand  $h \rightarrow 0$  pour toute fonction  $g$  continue à droite, on obtient

$$c\varphi'_d(u) = \lambda \left( \varphi(u) - \int_0^{+\infty} \varphi(u-y)dF_W(y) \right),$$

ce qui fournit le résultat demandé en observant que  $\varphi(x) = 0$  pour  $x < 0$ .

□

On obtient de même pour tout  $u \geq 0$ ,

$$\varphi'_g(u) = \frac{\lambda}{c} \left( \varphi(u) - \int_0^{u-} \varphi(u-y)dF_W(y) \right). \quad (9.2)$$

Dans le cas où  $F_W$  est continue, les dérivées à droite et à gauche sont les mêmes pour tout  $u$  et correspondent donc à  $\varphi'(u)$ .

Plaçons-nous dans le cas où  $W_1$  admet une densité  $f_{W_1}$ .

**Proposition III.1 (Equation intégrale)**

Pour tout  $u \geq 0$ ,

$$\varphi(u) = \varphi(0) + \frac{\lambda}{c} \int_0^u \varphi(u-y)(1-F_W(y))dy. \quad (9.3)$$

*Preuve :*

D'après (9.2), pour  $u > 0$ ,

$$\varphi'(u) = \frac{\lambda}{c} (\varphi(u) - (\varphi * f_W)(u)).$$

En prenant la transformée de Laplace, on obtient pour  $s > 0$

$$sL_\varphi(s) - \varphi(0) = \frac{\lambda}{c} L_\varphi(s) - \frac{\lambda}{c} L_\varphi(s).sL_{f_W}(s). \quad (9.4)$$

En effet, rappelons que

$$L_{\varphi'}(s) = \int_0^{+\infty} \varphi'(u).e^{-su}du = [\varphi(u).e^{-su}]_0^{+\infty} + \int_0^{+\infty} \varphi(u).se^{-su}du = -\varphi(0) + sL_\varphi(s)$$

en utilisant une intégration par parties classique. De la même manière,  $L_{\varphi * f_W} = L_\varphi.L_{f_W}$ , et donc comme  $F_W(0) = 0$ ,

$$L_{f_W}(s) = L_{F'_W}(s) = F_W(0) + sL_{F_W}(s) = sL_{F_W}(s).$$

En divisant par  $s$  l'équation (9.4), on obtient

$$L_\varphi(s) = \frac{\varphi(0)}{s} + \frac{\lambda}{c} \left( L_\varphi(s) \left( \frac{1}{s} - L_{F_W}(s) \right) \right).$$

Or, pour toute constante  $C > 0$ ,

$$L_C(s) = \frac{C}{s},$$

donc

$$\frac{1}{s} - L_{F_W}(s) = L_{1-F_W}(s)$$

et on obtient pour tout  $s > 0$

$$L_\varphi(s) = L_{\varphi(0) + \frac{\lambda}{c} \varphi * (1 - F_W)}(s),$$

ce qui donne le résultat recherché d'après l'injectivité de la transformée de Laplace.

□

### Proposition III.2

$$\varphi(+\infty) = \lim_{u \rightarrow +\infty} \varphi(u) = 1.$$

*Preuve :*

$\varphi$  est une fonction croissante. Pour  $n \geq 1$ ,

$$\varphi(n) = E(1_{\{\inf X(t) < -n\}}).$$

Or  $I_n = 1_{\{\inf X(t) > -n\}}$  est une suite croissante de v.a., qui tend vers 1 presque sûrement. En effet, comme  $\rho > 0$ , d'après la loi des grands nombres,  $X(t) \rightarrow +\infty$  quand  $t \rightarrow +\infty$ , et donc le processus est positif à partir d'un certain temps (aléatoire)  $T$  fini presque sûrement, et l'infimum pris sur le compact  $[0, T]$  est donc fini. D'après le théorème de convergence monotone,

$$\lim E(I_n) = E(\lim I_n) = E(1) = 1.$$

□

En passant à la limite dans l'équation (9.3), on obtient

$$\varphi(0) = 1 - \frac{\lambda\mu}{c},$$

ce qui constitue un résultat très robuste qui ne dépend de  $F_W$  que par la moyenne  $\mu$  de  $W$ .

Dans le cas où  $W \sim Exp(1/\mu)$ , d'après l'équation intégro-différentielle (9.2), pour tout  $u \geq 0$ ,

$$\varphi'(u) = \frac{\lambda}{c} \left( \varphi(u) - \int_0^u \varphi(u-y) e^{-y/\mu} dy \right) = \frac{\lambda}{c} \left( \varphi(u) - \int_0^u \varphi(y) e^{-(u-y)/\mu} dy \right). \quad (9.5)$$

En dérivant par rapport à  $u$ , on obtient

$$\varphi''(u) = \frac{\lambda}{c} \varphi'(u) - \frac{\lambda}{c} \frac{(-1)}{\mu} \int_0^u \varphi(y) e^{-(u-y)/\mu} dy - \frac{\lambda}{c\mu} \varphi(u). \quad (9.6)$$

(Par exemple factoriser le  $e^{-u/\mu}$  à l'extérieur de l'intégrale puis dériver le produit). Remarquons que d'après (9.2), le second terme de droite dans (9.6) est égal à

$$-\frac{\lambda}{c} \frac{(-1)}{\mu} \int_0^u \varphi(y) e^{-(u-y)/\mu} dy = \frac{1}{\mu} \left( \frac{\lambda}{c} \varphi(u) - \varphi'(u) \right).$$

L'équation (9.6) se simplifie donc en

$$\varphi''(u) = \frac{\lambda}{c} \varphi'(u) + \frac{1}{\mu} \left( \frac{\lambda}{c} \varphi(u) - \varphi'(u) \right) - \frac{\lambda}{c\mu} \varphi(u),$$

et donc

$$\varphi''(u) = \left( \frac{\lambda}{c} - \frac{1}{\mu} \right) \varphi'(u).$$

Soit

$$R = \frac{\lambda}{c} - \frac{1}{\mu} = \frac{\rho}{\mu(1+\rho)}.$$

En intégrant deux fois, on obtient

$$\varphi(u) = C_1 + C_2 e^{-Ru}.$$

$C_1 = \varphi(+\infty) = 1$  et

$$\varphi(0) = C_1 + C_2 = 1 - \frac{1}{1+\rho},$$

d'où pour  $u \geq 0$ ,

$$\varphi(u) = 1 - \frac{1}{1+\rho} e^{-Ru} \quad (9.7)$$

et

$$\psi(u) = \frac{1}{1+\rho} e^{-Ru}. \quad (9.8)$$

### 9.3 Méthodes de martingales

Appliquons le théorème d'arrêt optimal de Doob à la martingale  $(M(t))_{t \geq 0}$  (exercice : montrer qu'il s'agit bien d'une martingale par rapport à la filtration naturelle de  $(M(t))_{t \geq 0}$ ) définie par

$$M(t) = \frac{e^{-r(u+X(t))}}{E[e^{-rX(t)}]},$$

où  $r \geq 0$  est tel que

$$E[e^{-rX(t)}] < +\infty,$$

et au temps d'arrêt  $T_u \wedge t_0$ , où

$$T_u = \inf\{t \geq 0, u + X(t) < 0\}$$

est la variable aléatoire défective représentant l'instant de ruine, et  $t_0$  est un réel positif fixé. On obtient en conditionnant par  $T_u \leq t_0$  ou  $T_u > t_0$

$$e^{-ru} = E[M(0)] = E[M(T_u \wedge t_0)] \quad (9.9)$$

$$e^{-ru} = E[M(T_u \wedge t_0) | T_u \leq t_0] P(T_u \leq t_0) + E[M(T_u \wedge t_0) | T_u > t_0] P(T_u > t_0) \quad (9.10)$$

$$e^{-ru} \geq E[M(T_u) | T_u \leq t_0] P(T_u \leq t_0).$$

Rappelons que pour utiliser le théorème d'arrêt de Doob (voir cours de 2ème année), il faut (le plus souvent) avoir un temps d'arrêt, qui doit être fini presque sûrement. Or,

$$P(T_u = +\infty) = \varphi(u) > 0$$

dès que  $\rho > 0$ . La solution classique consiste à appliquer le théorème d'arrêt au temps d'arrêt  $T_u \wedge t_0$  (qui est toujours fini car inférieur à  $t_0 < \infty$ ), et à passer à la limite ( $t_0 \rightarrow +\infty$ ) pour obtenir le résultat avec  $T_u$ .

On a donc d'après [9.10]

$$P(T_u \leq t_0) \leq \frac{e^{-ru}}{E[M(T_u) | T_u \leq t_0]}. \quad (9.11)$$

Or

$$\frac{1}{E[M(T_u) | T_u \leq t_0]} = \frac{1}{E\left[\frac{e^{-r(u+X(T_u))}}{e^{g(r)T_u}} | T_u \leq t_0\right]} \leq \frac{1}{E\left[\frac{1}{e^{g(r)T_u}} | T_u \leq t_0\right]},$$

où  $g(r)$  est défini par

$$E[e^{-rX(t)}] = e^{g(r)t}$$

et est égal à

$$g(r) = \lambda(E[e^{rW}] - 1) - rc, \quad (9.12)$$

car sachant que  $T_u$  est fini,  $u + X(T_u) < 0$ . Donc l'inéquation (9.11) devient

$$P(T_u \leq t_0) \leq \frac{e^{-ru}}{E[e^{-g(r)T_u} \mid T_u \leq t_0]}. \quad (9.13)$$

On a conditionné par rapport à l'événement  $\{T_u \leq t_0\}$ , et donc

$$E[e^{-g(r)T_u} \mid T_u \leq t_0] \geq \inf_{0 \leq t \leq t_0} e^{-tg(r)},$$

et

$$\frac{1}{E[e^{-g(r)T_u} \mid T_u \leq t_0]} \leq \frac{1}{\inf_{0 \leq t \leq t_0} e^{-tg(r)}} = \sup_{0 \leq t \leq t_0} e^{tg(r)}.$$

Cela permet de réécrire (9.13) en

$$P(T_u \leq t_0) \leq e^{-ru} \sup_{0 \leq t \leq t_0} e^{tg(r)}.$$

En passant à la limite quand  $t_0 \rightarrow +\infty$ , on obtient

$$\psi(u) = P(T_u < +\infty) \leq e^{-ru} \sup_{t \geq 0} e^{tg(r)}.$$

Pour que la borne soit finie et présente un intérêt, il faut que  $r$  soit le plus grand possible tout en ayant  $g(r) \leq 0$ . Il faut donc prendre

$$R = \sup\{r \geq 0, g(r) \leq 0\}.$$

Dans le cas où  $W \sim \text{Exp}(1/\mu)$ , on retrouve la valeur de l'exposant de Cramer-Lundberg

$$R = \frac{\rho}{\mu(1+\rho)},$$

qui est ici le nombre strictement positif qui vérifie  $g(R) = 0$ . On obtient alors dans le cas général l'inégalité de Cramer-Lundberg

$$\psi(u) \leq e^{-Ru}. \quad (9.14)$$

Le cas où  $W \sim \text{Exp}(1/\mu)$  nous montre que la valeur de  $R$  est optimale. Il existe dans de nombreux cas des inégalités doubles qui permettent d'en-cadrer la probabilité de ruine par deux fonctions de type  $Cer^u$ .

On a appliqué le théorème de Doob à la martingale exponentielle pour pouvoir utiliser le théorème de convergence dominée :  $u + X(t) \rightarrow +\infty$  presque sûrement quand  $t \rightarrow +\infty$ , et donc  $e^{-r(u+X(t))}$  va tendre vers 0 presque sûrement.

L'inéquation (9.11) a été obtenue en passant de (9.9) à (9.10) en minorant brutalement par 0 le terme

$$E[M(T_u \wedge t_0) \mid T_u > t_0] P(T_u > t_0).$$

En prenant  $r = R$ , on peut montrer que ce terme tend vers 0 lorsque  $t_0 \rightarrow +\infty$ . Rappelons que dans ce cas  $g(R) = 0$  et que  $M(t)$  s'écrit simplement

$$M(t) = e^{-R(u+X(t))}.$$

Ceci nous permet d'écrire

$$0 \leq E[M(T_u \wedge t_0) \mid T_u > t_0] P(T_u > t_0) = E[e^{-R(u+X(t_0))} \cdot 1_{\{T_u > t_0\}}],$$

et comme

$$\{T_u > t_0\} \subset \{u + X(t_0) \geq 0\},$$

on obtient l'inégalité

$$0 \leq E[M(T_u \wedge t_0) \mid T_u > t_0] P(T_u > t_0) \leq E[e^{-R(u+X(t_0))} \cdot 1_{\{u+X(t_0) \geq 0\}}] \leq 1. \quad (9.15)$$

Or quand  $t \rightarrow +\infty$ ,

$$u + X(t) \rightarrow +\infty \quad \text{p.s.},$$

ce qui implique que lorsque  $t_0 \rightarrow +\infty$ ,

$$1_{\{u+X(t_0) \geq 0\}} \rightarrow 0 \quad \text{p.s..}$$

D'après le théorème de convergence dominée, on peut passer à la limite dans l'inéquation (9.15), ce qui donne à partir de (9.9) l'égalité suivante :

$$\psi(u) = \frac{e^{-Ru}}{E[e^{-R(u+X(T_u))} \mid T_u < +\infty]}. \quad (9.16)$$

**Exercice III.1** Dans le cas où  $W \sim \text{Exp}(1/\mu)$ , utiliser la perte de mémoire de la loi exponentielle et l'équation (9.16) pour retrouver la formule exacte de la probabilité de ruine dans le cas exponentiel (9.8).

Cet exercice sera corrigé en séance de TD. Il est vivement conseillé de le faire auparavant. Conseil : utiliser le conditionnement par rapport à la tribu qui correspond à l'information disponible sur le processus  $(X(t))_{t \geq 0}$  jusqu'à "juste avant la ruine" (c'est-à-dire obtenue à des temps strictement inférieurs à  $T_u$ ) :

$$\mathcal{F}_{T_u^-} = \sigma(A \cap \{t < T_u\}, \quad A \in \mathcal{F}_t, t > 0),$$

où  $(\mathcal{F}_t)_{t \geq 0}$  est la filtration naturelle du processus  $(X(t))_{t \geq 0}$ .

## Sixième partie

# Appendice, pense-bête

# Chapitre 33

## Lois usuelles

### 33.1 Lois de probabilité usuelles

Distribution	Probabilités $p_k$	Paramètres
Uniforme discrète	$1/N, k = 1, 2, \dots, N$	$N = 1, 2, \dots$
Bernoulli	$p_1 = p, p_0 = q$	$0 \leq p \leq 1, q = 1 - p$
Binômiale	$C_n^k p^k q^{n-k}, k = 0, 1, \dots, n$	$0 \leq p \leq 1, q = 1 - p, n = 1, 2, \dots$
Poisson	$e^{-\lambda} \frac{\lambda^k}{k!}, k = 0, 1, \dots$	$\lambda > 0$
Geométrique	$q^{k-1} p, k = 0, 1, \dots$	$0 \leq p \leq 1, q = 1 - p$
Binômiale négative	$C_{k-1}^{r-1} p^r q^{k-r}, k = r, r+1, \dots$	$0 \leq p \leq 1, q = 1 - p, r = 1, 2, \dots$

Distribution	Densité	Paramètres
Uniforme sur $[a, b]$	$1/(b-a)$ , $a \leq x \leq b$	$a, b \in \mathbb{R}$ ; $a < b$
Normale ou Gaussienne	$(2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/(2\sigma^2)}$ , $x \in \mathbb{R}$	$\mu \in \mathbb{R}$ , $\sigma > 0$
Log-normale	$\frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right)$ , $x > 0$	$\mu \in \mathbb{R}$ , $\sigma > 0$
Rayleigh	$xe^{-x^2/2}$ , $x \geq 0$	
Gamma	$\frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}$ , $x \geq 0$	$\alpha > 0$ , $\beta > 0$
Beta	$\frac{x^{r-1}(1-x)^{s-1}}{B(r,s)}$ , $0 \leq x \leq 1$	$r > 0$ , $s > 0$
Exponentielle ( $\Gamma$ , $\alpha = 1$ , $\beta = 1/\lambda$ )	$\lambda e^{-\lambda x}$ , $x \geq 0$	$\lambda > 0$
Laplace	$\frac{1}{2} \lambda e^{-\lambda x }$ , $x \in \mathbb{R}$	$\lambda > 0$
Chi-deux, $\chi^2$ ( $\Gamma$ , $\alpha = n/2$ , $\beta = 2$ )	$\frac{2^{-n/2} x^{n/2-1} e^{-x/2}}{\Gamma(n/2)}$ , $x \geq 0$	$n = 1, 2, \dots$
Student, $t$	$\frac{\Gamma(\frac{1}{2}(n+1))}{(n\pi)^{1/2}\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$ , $x \in \mathbb{R}$	$n = 1, 2, \dots$
$F$	$\frac{(m/n)^{m/2}}{B(m/2, n/2)} \frac{x^{m/2-1}}{(1+mx/n)^{m+n}/2}$	$m, n = 1, 2, \dots$
Cauchy	$\frac{\theta}{\pi(x^2+\theta^2)}$ , $x \in \mathbb{R}$	$\theta > 0$
Logistique	$\frac{e^{-(x-\alpha)/\beta}}{(1+e^{-(x-\alpha)/\beta})^2}$	$\alpha \in \mathbb{R}$ , $\beta > 0$
Weibull	$\alpha\theta x^{\alpha-1} e^{-\theta x^\alpha}$ , $x > 0$	$0 < \alpha < 1$ , $\theta > 0$
Gumbel	$\exp(x - e^x)$ , $x \in \mathbb{R}$	
Pareto	$\alpha k^\alpha x^{-(\alpha+1)}$ , $x \geq k$	$k > 0$ , $a > 0$ , $x \geq k$

## Chapitre 34

# Types de convergence

### 34.1 Convergence en Loi

Nous considérons ici une analyse classique de la convergence de mesures de probabilité appelée convergence en loi, sachant que d'autres approches utilisant des distances entre lois de probabilité (on peut par exemple construire une distance entre lois de probabilité absolument continues par rapport à une mesure dominante, comme la distance associée à la norme  $L^1$  des densités de probabilité par rapport à cette mesure dominante).

#### **Definition VI.1 Convergence complète d'une suite de fonctions de répartition**

*Soit  $F_n$  une suite de fonctions de répartition sur  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  et  $F$  une fonction de répartition sur  $\mathbb{R}^k$ . On dit que  $F_n$  converge complètement vers  $F$  et on note  $F_n \xrightarrow{c} F$  si  $F_n(x)$  converge vers  $F(x)$  en tout point  $x \in \mathbb{R}^k$  de continuité de  $F$ .*

#### **Remarques :**

- On a déjà étudié l'ensemble de continuité d'une fonction de répartition dans  $\mathbb{R}^n$ .
- Considérons une suite de fonctions de répartition  $F_n$  convergeant vers une fonction  $F$  en tout point de continuité de  $F$ .  $F$  n'est pas forcément une fonction de répartition. Par exemple, considérons la suite  $F_n$  définie par  $F_n(x) = 0$ , si  $x \leq -n$ ,  $F_n(x) = \frac{1}{2}$ , si  $x \in ]-n, n]$  et  $F_n(x) = 1$  si  $x > n$ .  $\lim F_n(x) = \frac{1}{2}$  pour tout  $x$  et  $F$  n'est pas une fonction de répartition.

#### **Definition VI.2 Convergence étroite d'une suite de mesures bornées sur $\mathbb{R}^k$**

*Soit  $P, P_1, P_2, \dots$  une suite de mesures positives bornées sur  $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ . On dit que  $P_n$  converge étroitement vers  $P$  et on note  $P_n \xrightarrow{e} P$ , si pour*

toute fonction réelle  $f$  définie sur  $\mathbb{R}^k$  **continue bornée**, on a :

$$\int f(x)dP_n(x) \longrightarrow \int f(x)dP(x).$$

– **Remarques :**

- Traduction anglaise de convergence étroite : *weak convergence*.
- La définition donnée pour des mesures positives bornées (ou finies, c'est-à-dire  $P_n(\mathbb{R}^k) < \infty$ ) s'applique en particulier à des mesures de probabilité.
- La notion de convergence étroite s'étend à des espaces métriques quelconques munis de leur tribu borélienne.
- On peut réécrire de manière équivalente  $\int f(x)dP_n(x) \longrightarrow \int f(x)dP(x)$  comme  $\int f(x)dF_n(x) \longrightarrow \int f(x)dF(x)$  où  $F, F_1, F_2, \dots$  sont les fonctions de répartition associées aux mesures  $P, P_1, P_2, \dots$

**Théorème VI.1 Helly-Bray** (*convergence complète  $\Rightarrow$  convergence étroite*)

Si  $F_n$  est une suite de fonctions de répartition sur  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  convergeant complètement vers une fonction de répartition sur  $\mathbb{R}^k$ ,  $F$ . Alors pour toute fonction  $f$  définie sur  $\mathbb{R}^k$ , continue bornée :

$$\int f(x)dF_n(x) \longrightarrow \int f(x)dF(x)$$

quand  $n \rightarrow \infty$ .

- **Remarque** : Notons  $P_n$  la mesure de probabilité sur  $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$  de fonction de répartition  $F_n$  et  $P$  la mesure de probabilité associée à  $F$ .
- La convergence complète des  $F_n$  implique la convergence étroite des mesures de probabilité  $P_n$ . Il existe une réciproque au théorème précédent.

**Théorème VI.2 équivalence entre convergence complète et convergence étroite**

Soit  $F, F_1, F_2, \dots$  les fonctions de répartition sur  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  de probabilités  $P, P_1, P_2, \dots$  sur  $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ . Alors :

$$P_n \xrightarrow{e} P \Leftrightarrow F_n \xrightarrow{c} F.$$

La démonstration des deux théorèmes précédents est admise. Dans le cas de mesures de probabilité admettant des densités, il existe un critère simple pour vérifier la convergence étroite.

**Propriété VI.1 convergence étroite de mesures absolument continues**

Soit  $(P_n)$  une suite de mesures de probabilités sur  $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$  admettant une densité  $f_n$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}^k$ . Si  $f_n$  converge presque partout (pour la mesure de Lebesgue sur  $\mathbb{R}^k$ ) vers une fonction de densité  $f$ , alors  $P_n$  converge étroitement vers  $P$  la mesure de probabilité sur  $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$  de densité  $f$ .

Sauf dans le cas de lois absolument continues par rapport à la mesure de Lebesgue, le concept de convergence complète est un peu délicat à manier dans  $\mathbb{R}^k$  (car il faut étudier l'ensemble de continuité de la loi limite). On peut néanmoins procéder avec l'approche suivante :

**Definition VI.3 Convergence étroite d'une suite de mesures bornées sur  $\mathbb{R}^k$**

*Soit  $P, P_1, P_2, \dots$  une suite de mesures de probabilité sur  $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ . On dit que  $P_n$  converge étroitement vers  $P$  et on note  $P_n \xrightarrow{e} P$ , si pour tout  $A \in \mathcal{B}(\mathbb{R}^k)$  tel que  $P(\partial A) = 0$ , on a  $P_n(A) \rightarrow P(A)$  quand  $n \rightarrow \infty$ .*

**Propriété VI.2** *Les deux définitions de convergence étroite sont équivalentes. Elles sont en outre équivalentes à :*

- (a)  $\liminf P_n(A) \geq P(A)$  pour tout  $A$  ouvert.
- (b)  $\limsup P_n(A) \leq P(A)$  pour tout  $A$  fermé.

– **Remarques :**

- $\partial A$  est la frontière de  $A$  définie par  $\partial A = \overline{A} \cap \overline{A^c}$  où  $\overline{A}$  est la fermeture de  $A$ .
- Dans cette définition, on ne demande pas la convergence pour tout ensemble mesurable, mais simplement pour tout ensemble mesurable dont la frontière n'est pas chargée pour la probabilité limite  $P$ .

Pour terminer considérons une dernière présentation de la convergence parfois utilisée :

**Definition VI.4 Convergence vague d'une suite de mesures bornées sur  $\mathbb{R}^k$**

*Soit  $P, P_1, P_2, \dots$  une suite de mesures positives bornées sur  $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ . On dit que  $P_n$  converge vaguement vers  $P$  et on note  $P_n \xrightarrow{v} P$ , si pour toute fonction réelle  $f$  définie sur  $\mathbb{R}^k$  continue à support compact, on a :*

$$\int f(x)dP_n(x) \rightarrow \int f(x)dP(x).$$

– **Remarques :**

- On montre qu'il est équivalent de dire :
  - (a)  $P_n$  converge étroitement vers  $P$ .
  - (b)  $P_n$  converge vaguement vers  $P$  et en outre  $P_n(\mathbb{R}^k) \rightarrow P(\mathbb{R}^k)$ .
- Dans le cas où  $P, P_1, P_2, \dots$  sont des mesures de probabilité, la condition  $P_n(\mathbb{R}^k) \rightarrow P(\mathbb{R}^k)$  est automatiquement vérifiée puisque  $P_n(\mathbb{R}^k) = P(\mathbb{R}^k) = 1$ .
- On peut alors se limiter à la condition  $\int f(x)dP_n(x) \rightarrow \int f(x)dP(x)$  pour toute fonction  $f$  continue à support compact.
- C'est parfois cette dernière définition qui est fournie pour la convergence étroite de mesures de probabilité.

Il existe une autre approche très simple pour étudier la convergence étroite, qui utilise les fonctions caractéristiques et ne fait pas appel à l'étude de la continuité de la fonction de répartition jointe.

**Propriété VI.3** *Si  $F, F_1, F_2, \dots$  sont des fonctions de répartition sur  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  avec des transformées de Fourier associées  $\varphi, \varphi_1, \varphi_2, \dots$  et si  $F_n \xrightarrow{c} F$  alors  $\varphi_n(t) \rightarrow \varphi(t)$  pour tout  $t \in \mathbb{R}^k$ .*

**Démonstration :** voir l'exercice **(VI.1)**.

**Exercice VI.1** *Montrer que si  $F, F_1, F_2, \dots$  sont des fonctions de répartition sur  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  de transformées de Fourier associées  $\varphi, \varphi_1, \varphi_2, \dots$  et si  $F_n \xrightarrow{c} F$  alors  $\varphi_n(t) \rightarrow \varphi(t)$  pour tout  $t \in \mathbb{R}^k$ .*

**Corrigé :**

- C'est une conséquence directe du théorème de Helly-Bray.
- Pour tout  $t \in \mathbb{R}^k$  donné, les fonctions qui à  $x \in \mathbb{R}^k$  associent  $\mathcal{R}(e^{i\langle t,x \rangle})$  et  $\mathcal{I}(e^{i\langle t,x \rangle})$  sont continues bornées.
- $\int e^{i\langle t,x \rangle} dF_n(x) \rightarrow \int e^{i\langle t,x \rangle} dF(x)$ , c'est-à-dire  $\varphi_n(t) \rightarrow \varphi(t)$ .

**Théorème VI.3 de continuité (dit de Lévy)**

*Soit  $(\varphi_n)$ ,  $n \in \mathbb{N}$ , une suite de transformées de Fourier, de mesures de probabilité  $P_n$  sur  $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ ,  $k \in \mathbb{N}$ . Si  $\varphi_n$  converge en tout point de  $\mathbb{R}^k$  vers une fonction  $\varphi$  continue au point  $(0, 0, \dots, 0)$ , alors  $\varphi$  est la transformée de Fourier d'une mesure de probabilité  $P$  sur  $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$  et  $P_n$  converge étroitement vers  $P$ .*

**Remarques :**

- Ce théorème (admis) établit une réciproque de la propriété précédente.
- Attention : toute fonction n'est pas la transformée de Fourier d'une mesure (il faut par exemple qu'il y ait continuité uniforme). Il existe plusieurs propriétés permettant de caractériser des situations où une fonction donnée est une fonction caractéristique.
- Si  $\varphi_n(t)$  converge vers une fonction caractéristique  $\varphi(t)$  d'une probabilité  $P$ , alors le théorème s'applique car  $\varphi$  est continue en zéro et  $P_n$  converge vers la probabilité  $P$ .

**Corollaire VI.1 continuité de la convolution**

*Soit  $(P_n)$  et  $(Q_n)$  deux suites de mesures de probabilité sur  $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$  convergeant étroitement, respectivement vers  $P$  et  $Q$ . Alors  $(P_n * Q_n)$  converge étroitement vers  $P * Q$ .*

**Démonstration :** La démonstration est immédiate en utilisant le fait que la transformée de Fourier de  $P_n * Q_n$  est le produit des transformées de Fourier et le théorème de Lévy.

**Definition VI.5 convergence en loi**

*Soit  $X_n$  une suite de vecteurs aléatoires définis sur un espace probabilisé à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$ . On dit que  $X_n$  converge en loi vers le vecteur aléatoire  $X$  (à valeurs dans  $\mathbb{R}^k$ ) et on note  $X_n \xrightarrow{\mathcal{L}} X$  si :*

$$P^{X_n} \xrightarrow{e} P^X.$$

**Remarque :**

- Si  $F_n, F$  sont les fonctions de répartition de  $X_n, X$ , la convergence en loi de  $X_n$  vers  $X$  est équivalente à la convergence de  $F_n(x)$  vers  $F(x)$  en tout point de continuité de  $F$ .

**Corollaire VI.2 Cramer-Wold**

*Soit  $X_n$  une suite de vecteurs aléatoires définis sur un espace probabilisé à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$ .  $X_n$  converge en loi vers le vecteur aléatoire  $X$  (à valeurs dans  $\mathbb{R}^k$ ) si et seulement si pour tout  $\lambda = (\lambda_1, \dots, \lambda_k) \in \mathbb{R}^k$ ,  $\lambda'X_n$  converge en loi vers  $\lambda'X$ .*

**Démonstration :** voir l'exercice [VI.2](#)

**Exercice VI.2 Cramer-Wold**

*Soit  $X_n$  une suite de vecteurs aléatoires définis sur un espace probabilisé à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$ . Montrer que  $X_n$  converge en loi vers le vecteur aléatoire  $X$  (à valeurs dans  $\mathbb{R}^k$ ) si et seulement si pour tout  $\lambda = (\lambda_1, \dots, \lambda_k) \in \mathbb{R}^k$ ,  $\lambda'X_n$  converge en loi vers  $\lambda'X$ .*

**Corrigé :** on utilise les fonctions caractéristiques :

- Supposons que  $X_n$  converge en loi vers  $X$ , c'est-à-dire que  $P^{X_n}$  converge étroitement vers  $P^X$ .
- D'après le théorème de continuité, il est équivalent de dire que  $\int e^{it'x} dP^{X_n}(x)$  converge vers  $\int e^{it'x} dP^X(x)$  pour tout  $t \in \mathbb{R}^k$ , c'est-à-dire d'après le théorème de transfert que  $E[e^{it'X_n}]$  converge vers  $E[e^{it'X}]$  ou encore avec les notations habituelles que  $\varphi_{X_n}(t)$  converge vers  $\varphi_X(t)$ .
- Soit  $u \in \mathbb{R}$ .  $\varphi_{\lambda'X_n}(u) = E[e^{iu\lambda'X_n}] = \varphi_{X_n}(u\lambda) \rightarrow \varphi_X(u\lambda) = \varphi_{\lambda'X}(u)$ .
- D'après le théorème de continuité,  $\lambda'X_n$  converge en loi vers  $\lambda'X$ .
- Réciproquement, supposons que  $\lambda'X_n$  converge en loi vers  $\lambda'X$  pour tout  $\lambda \in \mathbb{R}^k$ .
- Dans ce cas,  $\varphi_{\lambda'X_n}(1) \rightarrow \varphi_{\lambda'X}(1)$  en appliquant le théorème de continuité.
- En utilisant  $\varphi_{X_n}(\lambda) = \varphi_{\lambda'X_n}(1)$  et  $\varphi_{X_n}(\lambda) = \varphi_{\lambda'X_n}(1)$ , on obtient la convergence en loi de  $X_n$  vers  $X$ .

**Remarque :**

- En prenant  $\lambda_i = 1$  et  $\lambda_j = 0$ ,  $j \neq i$ , on obtient que la convergence en loi du vecteur aléatoire  $X_n$  vers le vecteur aléatoire  $X$  implique la convergence en loi des composantes de  $X_n$ ,  $X_n^i$  vers les composantes de  $X$ ,  $X^i$ .
- En revanche, la réciproque est fausse.
- Considérons par exemple  $X \sim N(0, 1)$  et les suites  $X_n = X$ ,  $Y_n = (-1)^n X$ .  $X_n$  et  $Y_n$  convergent en loi vers  $X$ .
- La suite  $(X_n, Y_n)$  ne converge pas en loi (ni la fonction de répartition jointe, ni la fonction caractéristique jointe ne convergent) et par conséquent  $(X_n, Y_n)$  ne converge pas en loi vers  $(X, X)$ .

**Corollaire VI.3** Soit  $(X_n)$  une suite de vecteurs aléatoires convergeant en loi vers  $X$ , tous ces vecteurs aléatoires étant définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  et à valeurs dans  $\mathbb{R}^k$ . Soit  $h$  une fonction continue de  $\mathbb{R}^k$  dans  $\mathbb{R}^m$ . Alors  $h(X_n)$  converge en loi vers  $h(X)$

**Démonstration :** voir exercice (VI.3).

**Exercice VI.3** Soit  $(X_n)$  une suite de vecteurs aléatoires convergeant en loi vers  $X$ , tous ces vecteurs aléatoires étant définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  et à valeurs dans  $\mathbb{R}^k$ . Soit  $h$  une fonction continue de  $\mathbb{R}^k$  dans  $\mathbb{R}^m$ . Montrer que  $h(X_n)$  converge en loi vers  $h(X)$

**Corrigé :**

- Soit  $t \in \mathbb{R}^m$  donné. La fonction  $x \rightarrow e^{it'h(x)}$  est une fonction continue bornée.
- D'après le théorème de Helly-Bray  $\int e^{it'h(x)} dP^{X_n}(x) \rightarrow \int e^{it'h(x)} dP^X(x)$ , c'est-à-dire  $\varphi_{h(X_n)}(t) \rightarrow \varphi_{h(X)}(t)$ , ce qui montre la propriété d'après le théorème de P. Levy.

**Propriété VI.4 équivalence entre convergence en loi et convergence simple des fonctions caractéristiques**

Soit  $X$  et  $X_n$ ,  $n \in \mathbb{N}$  des vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , de fonctions caractéristiques respectivement  $\varphi(t)$  et  $\varphi_n(t)$ ,  $t \in \mathbb{R}^k$ . Alors,

$$X_n \xrightarrow{\mathcal{L}} X \iff \lim_{n \rightarrow \infty} \varphi_{X_n}(t) = \varphi_X(t), \forall t \in \mathbb{R}^k.$$

**Propriété VI.5 un exemple rare où convergence en loi entraîne convergence presque sûre**

Soit  $X$  une variable aléatoire réelle définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  et  $f_0, f_1, \dots$  une suite de fonctions réelles croissantes. Si  $f_n(X)$  converge en loi vers  $f_0(X)$  alors  $f_n(X)$  converge  $P$ -presque sûrement vers  $f(X)$ .

Pour conclure, mentionnons que l'on peut s'intéresser à la métrisabilité de la convergence en loi, c'est-à-dire à la possibilité de construire une distance entre mesures de probabilité telle que la convergence étroite de mesures se caractérise par une distance entre les mesures et la mesure limite tendant vers 0.

## 34.2 Convergence presque sûre et convergence en probabilité

**Definition VI.1 convergence presque sûre** Soit  $(X_n = (X_n^1, \dots, X_n^k))$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On définit l'ensemble de convergence des  $X_n$  comme  $\{\vec{X_n}\} = \{\omega \in \Omega, \exists \lim_{n \rightarrow \infty} X_n^j(\omega), \forall j, 1 \leq j \leq k\}$ . On dit que  $X_n$  converge presque sûrement (ou presque-partout) si  $P(\{\vec{X_n}\}) = 1$ .

**Definition VI.2 convergence presque sûre** Soit  $(X_n)$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On dit que  $X_n$  converge presque sûrement (ou presque-partout) vers le vecteur aléatoire  $X$  lorsque  $n \rightarrow \infty$  et on note  $X_n \xrightarrow{p.s} X$ , si  $P(\{\omega \in \Omega, \lim X_n(\omega) = X(\omega)\}) = 1$ .

**Remarques :**

- En d'autres termes  $X_n$  converge simplement vers  $X$  sur un ensemble de probabilité égale à un.
- De manière équivalente, l'ensemble des  $\omega$  tels que  $X_n(\omega)$  ne converge pas vers  $X(\omega)$  est négligeable.
- La cohérence entre les deux définitions est assurée par la propriété suivante.

**Propriété VI.1** Soit  $(X_n = (X_n^1, \dots, X_n^k))$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  convergeant presque sûrement. Alors, il existe un vecteur aléatoire  $X$ , unique au sens de l'égalité  $P$ -presque sûre, tel que  $X_n \xrightarrow{p.s} X$ .

**Démonstration :** voir exercice VI.4.

**Exercice VI.4** Soit  $(X_n = (X_n^1, \dots, X_n^k))$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  convergeant presque sûrement. Montrer qu'il existe un vecteur aléatoire  $X$ , unique au sens de l'égalité  $P$ -presque sûre, tel que  $X_n \xrightarrow{p.s} X$ .

**Corrigé :**

- Pour chaque  $j = 1, \dots, k$ , considérons  $X^j = \inf_{n \geq 1} \sup_{m \geq n} X_m^j = \limsup X_n^j$ .
- $X^j$  est défini de manière sur  $\Omega$  et est une variable aléatoire.  $X = (X^1, \dots, X^k)$  est un vecteur aléatoire.
- Pour tout  $\omega \in \{\overrightarrow{X_n}\}$ ,  $X_n^j(\omega)$  converge et par conséquent  $X^j(\omega) = \limsup X_n^j(\omega) = \lim X_n^j(\omega)$ .
- On en déduit que pour tout  $j = 1, \dots, k$ ,  $X_n^j$  converge vers  $X^j$  sur un ensemble de probabilité égale à 1, ce qui montre la propriété.

**Notation :** Dans la suite, quand  $x = (x^1, \dots, x^k) \in \mathbb{R}^k$ , on note par  $|x|$ , la norme euclidienne de  $x$ , à savoir  $\left(\sum_{j=1}^k x_j^2\right)^{1/2}$ .

**Théorème VI.1 condition nécessaire et suffisante de convergence presque sûre**

Soient  $X$  une variable aléatoire réelle et  $(X_n)$  une suite de variables aléatoires réelles définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Une condition nécessaire et suffisante pour que  $X_n \xrightarrow{p.s} X$  est que :

$$P\left(\left\{\omega \in \Omega, \sup_{m \geq n} |X_m(\omega) - X(\omega)| > \varepsilon\right\}\right) \rightarrow 0, \quad \text{quand } n \rightarrow \infty,$$

pour tout  $\varepsilon > 0$ .

### Démonstration :

- Notons  $A_m^\varepsilon = \{\omega, |X_m(\omega) - X(\omega)| > \varepsilon\}$  et  $A^\varepsilon = \bigcap_{n \geq 1} \bigcup_{m \geq n} A_m^\varepsilon$  (en langage ensembliste, il s'agit de  $\limsup A_n^\varepsilon$ ).
- On remarque que  $\{\omega \in \Omega, \sup_{m \geq n} |X_m(\omega) - X(\omega)| > \varepsilon\} = \bigcup_{m \geq n} A_m^\varepsilon$  et la condition du théorème s'exprime de manière équivalente comme  $\lim_{n \rightarrow \infty} P\left(\bigcup_{m \geq n} A_m^\varepsilon\right) = 0$  pour tout  $\varepsilon > 0$ .
- On remarque que pour tout  $\varepsilon > 0$ , il existe  $k \in \mathbb{N}$  tel que  $\frac{1}{k} < \varepsilon$ . On alors  $A_m^\varepsilon \subset A_m^{1/k}$  et donc  $\bigcup_{m \geq n} A_m^\varepsilon \subset \bigcup_{m \geq n} A_m^{1/k}$ . En utilisant la croissance de  $P$ ,
- $P\left(\bigcup_{m \geq n} A_m^\varepsilon\right) \leq P\left(\bigcup_{m \geq n} A_m^{1/k}\right)$  et donc  $\lim_{n \rightarrow \infty} P\left(\bigcup_{m \geq n} A_m^\varepsilon\right) \leq \lim_{n \rightarrow \infty} P\left(\bigcup_{m \geq n} A_m^{1/k}\right)$ .
- Il suffit donc de montrer la propriété pour des  $\varepsilon$  de la forme  $\frac{1}{k}$  avec  $k \in \mathbb{N}$ .
- Examinons les ensembles de convergence,  $\{\vec{X_n}\}$ , et de non convergence,  $\{\vec{X_n}\}^c$  de  $X_n$ ,
- $\omega \in \{\vec{X_n}\} \Leftrightarrow \forall k \in \mathbb{N}^*, \exists n \in \mathbb{N}^*$  tel que  $\forall m \geq n, |X_m(\omega) - X(\omega)| \leq \frac{1}{k}$ .
- En utilisant les propriétés des quantificateurs,
- $\omega \in \{\vec{X_n}\}^c \Leftrightarrow \exists k \in \mathbb{N}^*, \forall n \in \mathbb{N}^*$  tel que  $\exists m \geq n, |X_m(\omega) - X(\omega)| > \frac{1}{k}$ .
- En utilisant les définitions précédentes, on voit que :
- $\omega \in \{\vec{X_n}\}^c \Leftrightarrow \exists k \in \mathbb{N}^*$  tel que  $\omega \in \bigcap_{n \geq 1} \bigcup_{m \geq n} A_m^{1/k}$ , c'est-à-dire tel que  $\omega \in A^{1/k}$ .
- On a ainsi  $\{\vec{X_n}\}^c = \bigcup_{k \geq 1} A^{1/k}$ .
- La convergence presque sûre des  $X_n$  est équivalente à :
- $P(\{\vec{X_n}\}^c) = P\left(\bigcup_{k \geq 1} A^{1/k}\right) = 0$ .
- $P\left(\bigcup_{k \geq 1} A^{1/k}\right) = 0 \Leftrightarrow P(A^{1/k}) = 0, \forall k \geq 1$ .
- Comme  $A^{1/k} = \bigcap_{n \geq 1} \bigcup_{m \geq n} A_m^{1/k}, P(A^{1/k}) = 0 \Leftrightarrow \lim_{n \rightarrow \infty} P\left(\bigcup_{m \geq n} A_m^{1/k}\right) = 0$  par continuité décroissante de  $P$ .
- Comme  $\bigcup_{m \geq n} A_m^{1/k} = \{\omega \in \Omega, \sup_{m \geq n} |X_m(\omega) - X(\omega)| > \frac{1}{k}\}$ , on obtient la propriété annoncée.

### Definition VI.3 convergence en probabilité

Soit  $(X_n)$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k, k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On dit que  $X_n$  converge en probabilité quand  $n \rightarrow \infty$  vers le vecteur aléatoire  $X$  (et on note  $X_n \xrightarrow{P} X$ ) si pour tout  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0.$$

### Théorème VI.2 convergence p.s implique convergence en probabilité

Soit  $(X_n)$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k, k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  convergeant presque sûrement vers un vecteur aléatoire  $X$ . Alors  $X_n$  converge en probabilité vers  $X$ .

**Démonstration :** voir exercice VI.5).

**Exercice VI.5 convergence p.s implique convergence en probabilité**

Soit  $(X_n)$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  convergeant presque sûrement vers un vecteur aléatoire  $X$ . Montrer qu'alors  $X_n$  converge en probabilité vers  $X$ .

**Corrigé :**

- D'après le théorème précédent caractérisant la convergence presque sûre, on a :
- $\{\omega \in \Omega, |X_n(\omega) - X(\omega)| > \varepsilon\} \subset \{\omega \in \Omega, \sup_{m \geq n} |X_m(\omega) - X(\omega)| > \varepsilon\}$ .
- D'où :
- $P(\{\omega \in \Omega, |X_n(\omega) - X(\omega)| > \varepsilon\}) \leq P(\{\omega \in \Omega, \sup_{m \geq n} |X_m(\omega) - X(\omega)| > \varepsilon\})$ .
- par passage à la limite en  $n$ , et en utilisant les notations simplifiées,
- $\lim_{n \rightarrow \infty} P(\{|X_n - X| > \varepsilon\}) \leq \lim_{n \rightarrow \infty} P(\{\sup_{m \geq n} |X_m - X| > \varepsilon\}) = 0$ . ■

**Definition VI.4 suite de cauchy pour la convergence en probabilité**

Soit  $(X_n)$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On dit que  $X_n$  est une suite de Cauchy pour la convergence en probabilité si et seulement si, pour tout  $\varepsilon > 0$  :

$$\lim_{n,m \rightarrow \infty} P(|X_n - X_m| > \varepsilon) = 0.$$

On vérifie immédiatement qu'une suite de variables aléatoires convergeant en probabilité est une suite de Cauchy pour la convergence en probabilité. On montrera ultérieurement la réciproque.

**Théorème VI.3** Soit  $(X_n)$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  de Cauchy pour la convergence en probabilité. Alors, on peut en extraire une sous suite  $(X_{n_k})$  convergeant presque sûrement.

**Démonstration :**

- La définition de  $X_n$  suite de Cauchy en probabilité peut se réécrire :
- $\forall \varepsilon > 0, \forall h > 0, \exists n_h, n \geq n_h, m \geq n_h \Rightarrow P(|X_n - X_m| > \varepsilon) < h$ .
- Prenons  $\varepsilon = h = \frac{1}{k^2}$  pour  $k \in \mathbb{N}^*$ .
- $\exists n_k, \forall n \geq n_k, \forall m \geq n_k \Rightarrow P(|X_n - X_m| > \frac{1}{k^2}) < \frac{1}{k^2}$ .
- On peut prendre la suite  $n_k$  croissante : si  $n_{k+1} < n_k$ , on peut remplacer  $n_{k+1}$  par  $n_k + 1$ .
- On a en particulier  $P(|X_{n_{k+1}} - X_{n_k}| > \frac{1}{k^2}) < \frac{1}{k^2}$ .
- $\sum_{k \geq 1} P(|X_{n_{k+1}} - X_{n_k}| > \frac{1}{k^2}) < \sum_{k \geq 1} \frac{1}{k^2} < \infty$ . On peut donc appliquer le lemme de Borel-Cantelli :
- $\limsup \{|X_{n_{k+1}} - X_{n_k}| > \frac{1}{k^2}\}$  est négligeable.
- $\limsup \{|X_{n_{k+1}} - X_{n_k}| > \frac{1}{k^2}\} = \bigcap_{n \geq 1} \bigcup_{k \geq n} \{|X_{n_{k+1}} - X_{n_k}| > \frac{1}{k^2}\}$ .
- $\omega \in \bigcap_{n \geq 1} \bigcup_{k \geq n} \{|X_{n_{k+1}} - X_{n_k}| > \frac{1}{k^2}\}$

- $\Leftrightarrow \forall n \geq 1, \exists k \geq n, |X_{n_{k+1}}(\omega) - X_{n_k}(\omega)| > \frac{1}{k^2}$
- $\omega \in \left( \bigcap_{n \geq 1} \bigcup_{k \geq n} \{ |X_{n_{k+1}} - X_{n_k}| > \frac{1}{k^2} \} \right)^c$
- $\Leftrightarrow \exists n \geq 1, \forall k \geq n, |X_{n_{k+1}}(\omega) - X_{n_k}(\omega)| \leq \frac{1}{k^2}$
- On en conclut que si  $\omega \in (\limsup \{ |X_{n_{k+1}} - X_{n_k}| > \frac{1}{k^2} \})^c$  (qui est de probabilité 1), alors
- $\sum_{k \geq 1} |X_{n_{k+1}}(\omega) - X_{n_k}(\omega)| \leq \sum_{k \geq 1} \frac{1}{k^2} < \infty$ .
- La série de terme général  $X_{n_{k+1}}(\omega) - X_{n_k}(\omega)$  est absolument convergente, donc convergente.
- Pour tout  $\omega \in (\limsup \{ |X_{n_{k+1}} - X_{n_k}| > \frac{1}{k^2} \})^c$ ,  $\sum_{k=1}^p X_{n_{k+1}}(\omega) - X_{n_k}(\omega) = X_{n_{p+1}}(\omega) - X_{n_1}(\omega)$  converge quand  $p \rightarrow \infty$ .
- Ceci montre que l'ensemble de convergence de  $(X_{n_p})$  est de probabilité 1, c'est-à-dire que la suite  $(X_{n_p})$  converge presque sûrement.

**Lemme VI.1** Soit  $X$  et  $Y$  deux vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Alors, pour tout  $\varepsilon > 0$ ,

$$P(|X + Y| \geq \varepsilon) \leq P\left(\left\{|X| \geq \frac{\varepsilon}{2}\right\}\right) + P\left(\left\{|Y| \geq \frac{\varepsilon}{2}\right\}\right).$$

**Démonstration :**

- Comme  $|X(\omega) + Y(\omega)| \leq |X(\omega)| + |Y(\omega)|$ ,
- $|X(\omega) + Y(\omega)| \geq \varepsilon \Rightarrow |X(\omega)| + |Y(\omega)| \geq \varepsilon$ .
- Si  $|X(\omega)| + |Y(\omega)| \geq \varepsilon$ , alors  $|X(\omega)| \geq \frac{\varepsilon}{2}$  ou  $|Y(\omega)| \geq \frac{\varepsilon}{2}$ ,
- c'est-à-dire  $\omega \in \{|X| \geq \frac{\varepsilon}{2}\}$  ou  $\omega \in \{|Y| \geq \frac{\varepsilon}{2}\}$ , ou encore
- $\{|X + Y| \geq \varepsilon\} \subset \{|X| \geq \frac{\varepsilon}{2}\} \cup \{|Y| \geq \frac{\varepsilon}{2}\}$ , ce qui donne le résultat du lemme par croissance de  $P$ .

**Propriété VI.2** Soit  $X_1, X_2, \dots$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$ , définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  convergeant en probabilité vers le vecteur aléatoire  $X$ . Alors  $X$  est unique au sens de l'égalité presque sûre.

**Démonstration :** voir exercice [VI.6].

**Exercice VI.6** Soit  $X_1, X_2, \dots$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$ , définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  convergeant en probabilité vers le vecteur aléatoire  $X$ . Montrer en utilisant le lemme [VI.1] que  $X$  est unique au sens de l'égalité presque sûre.

**Corrigé :**

- Supposons que  $X_n \xrightarrow{P} X$  et  $X_n \xrightarrow{P} Y$ .
- D'après le lemme précédent,
- $0 \leq P(|X - Y| \geq \varepsilon) \leq P\left(\left\{|X - X_n| \geq \frac{\varepsilon}{2}\right\}\right) + P\left(\left\{|X_n - Y| \geq \frac{\varepsilon}{2}\right\}\right)$ .
- De par la convergence en probabilité, le terme de droite tend vers 0 quand  $n \rightarrow \infty$  et donc pour tout  $\varepsilon > 0$ ,  $P(|X - Y| \geq \varepsilon) = 0$ .
- Par la continuité en zéro de  $P$ , on en déduit  $P(|X - Y| > 0) = 0$ ,
- soit  $P(X = Y) = 1$ , ce qui montre l'égalité presque sûre de  $X$  et  $Y$ .

**Théorème VI.4** Soit  $(X_n)$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  de Cauchy pour la convergence en probabilité. Alors, il existe un vecteur aléatoire  $X$  tel que  $X_n$  converge en probabilité vers  $X$ .

**Démonstration :** voir exercice (??).

**Exercice VI.7** Soit  $(X_n)$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  de Cauchy pour la convergence en probabilité. Montrer en utilisant le théorème [VI.3] et le lemme [VI.1] qu'il existe un vecteur aléatoire  $X$  tel que  $X_n$  converge en probabilité vers  $X$ .

**Corrigé :**

- Considérons  $X_{n_p}$  la sous-suite presque sûrement convergente mise en évidence au théorème [VI.3] et notons  $X$  sa limite.
- $X_{n_p}$  converge aussi en probabilité vers  $X$  d'après le théorème [VI.2].
- Pour tout  $\varepsilon > 0$ , d'après le lemme [VI.1], on a :
- $0 \leq P(\{|X_n - X| \geq \varepsilon\}) \leq P(\{|X_n - X_{n_p}| \geq \frac{\varepsilon}{2}\}) + P(\{|X_{n_p} - X| \geq \frac{\varepsilon}{2}\})$ .
- Quand  $p$  tend vers l'infini,  $n_p$  tend vers l'infini (c'est une suite croissante).
- Quand  $p$  et  $n$  tendent vers l'infini,  $P(\{|X_n - X_{n_p}| \geq \frac{\varepsilon}{2}\})$  tend vers zéro car  $X_n$  est de Cauchy en probabilité.
- De même  $P(\{|X_{n_p} - X| \geq \frac{\varepsilon}{2}\})$  tend vers 0 quand  $n \rightarrow \infty$  et donc la convergence en probabilité de  $X_n$  vers  $X$ .

**Remarques :**

- Par ailleurs, comme toute suite convergeant en probabilité est de cauchy en probabilité et en vertu du théorème [VI.3] si  $X_n$  converge en probabilité vers  $X$ , il existe une sous-suite  $X_{n_p}$  convergeant presque sûrement. Notons  $Z$  cette limite.
- De par le théorème [VI.2],  $X_{n_p}$  converge aussi en probabilité vers  $Z$ .
- Par ailleurs, on remarque que  $X_{n_p}$  converge en probabilité vers  $X$ .
- De par l'unicité de la limite pour la convergence en probabilité, on en conclut  $Z = X$ , c'est-à-dire que :
- Si  $X_n$  converge en probabilité vers  $X$ , il existe une sous-suite  $X_{n_p}$  convergeant presque sûrement vers  $X$ .
- Le théorème suivant généralise cette propriété.

**Théorème VI.5** Soit  $X$ ,  $(X_n)$  respectivement un vecteur aléatoire et une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$ , définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Une condition nécessaire et suffisante pour que  $X_n \xrightarrow{P} X$  est que toute sous-suite de  $X_n$  contienne une sous-suite qui converge presque-sûrement vers  $X$ .

**Démonstration :**

- Condition nécessaire :
- Supposons  $X_n \xrightarrow{P} X$ .

- En écrivant la définition de la convergence en loi, on remarque que toute sous-suite  $X_{n_p}$  ( $n_p$  croissante en  $p$ ) converge en loi vers  $X$ .
- De par la remarque précédente, il existe une sous-suite de  $X_{n_p}$  qui converge presque sûrement vers  $X$ , ce qui montre que la condition donnée est nécessaire.
- Condition suffisante :
- Raisonnons par l'absurde et supposons que toute sous-suite de  $X_n$  contient une sous-suite qui converge presque sûrement vers  $X$  mais que  $X_n$  ne converge pas en probabilité vers  $X$ .
- La condition de convergence en probabilité s'écrit  $\forall \varepsilon > 0, \forall h > 0, \exists k \in \mathbb{N}, \forall n \geq k, P(\{|X_n - X| > \varepsilon\}) < h$ .
- La non convergence s'écrit  $\exists \varepsilon > 0, \exists h > 0, \forall k, \exists n \geq k, P(\{|X_n - X| > \varepsilon\}) \geq h$ .
- Notons  $n_k$  ce  $n$ . On peut prendre  $n_k$  croissant en  $k$  sans perte de généralité.
- On a donc  $\forall k \in \mathbb{N}, P(\{|X_{n_k} - X| > \varepsilon\}) \geq h$  et ni  $X_{n_k}$  ni aucune sous-suite extraite de  $X_{n_k}$  ne converge en probabilité vers  $X$ .
- Ainsi, aucune sous-suite de  $X_{n_k}$  ne converge presque sûrement car si elle convergeait presque sûrement, elle convergerait aussi en probabilité.
- Ceci est en contradiction avec le point de départ.

### **Théorème VI.6 invariance de la convergence en probabilité par transformation continue**

*Soit  $X, (X_n)$  respectivement un vecteur aléatoire et une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$ , définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Soit  $f$  une fonction mesurable de  $\mathbb{R}^k$  dans  $\mathbb{R}^l$ ,  $l \in \mathbb{N}$ ,  $f$  étant en outre continue sur un borélien de  $\mathbb{R}^k$ ,  $B$ , tel que  $P(X \in B) = 1$ . Si  $X_n \xrightarrow{P} X$ , alors  $f(X_n) \xrightarrow{P} f(X)$*

#### **Démonstration :**

- La démonstration est une conséquence du théorème précédent.
- Considérons  $f(X_{n_j})$  une sous-suite de  $f(X_n)$ .
- Comme  $X_{n_j}$  converge en probabilité vers  $X$  (vérifier la définition), il existe une sous suite de  $X_{n_j}$ , soit  $X_{r_j}$  (c'est aussi une sous suite de  $X_n$ ) qui converge presque sûrement vers  $X$ .
- Notons  $A = \{\overrightarrow{X_{r_j}}\} \cap \{X \in B\}$ .
- $A$  appartient, intersection de deux événements appartient à  $\mathcal{A}$ . Ces deux événements étant de probabilité 1,  $A$  est également de probabilité 1.
- Pour tout  $\omega \in A$ ,  $X_{r_j}(\omega) \rightarrow X(\omega)$ .
- Comme en outre  $f$  est continue sur  $A$ , on en déduit que  $f(X_{r_j}(\omega)) \rightarrow f(X(\omega))$ , pour tout  $\omega \in A$ , c'est-à-dire que :
- $f(X_{r_j})$  converge presque sûrement vers  $f(X)$ .
- D'après le théorème précédent  $f(X_n)$  converge en probabilité vers  $f(X)$ .

**Propriété VI.3** Soient  $X, X_1, X_2, \dots$  et  $Y, Y_1, Y_2, \dots$  deux suites de vecteurs aléatoires à valeurs respectivement dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$ , et  $\mathbb{R}^l$ ,  $l \in \mathbb{N}$ ,

définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Si  $X_n \xrightarrow{P} X$  et  $Y_n \xrightarrow{P} Y$ , alors  $(X_n, Y_n) \xrightarrow{P} (X, Y)$ .

**Démonstration :** voir exercice [VI.8].

**Exercice VI.8** Soient  $X, X_1, X_2, \dots$  et  $Y, Y_1, Y_2, \dots$  deux suites de vecteurs aléatoires à valeurs respectivement dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$ , et  $\mathbb{R}^l$ ,  $l \in \mathbb{N}$ , définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Montrer en utilisant le lemme [VI.1] que si  $X_n \xrightarrow{P} X$  et  $Y_n \xrightarrow{P} Y$ , alors  $(X_n, Y_n) \xrightarrow{P} (X, Y)$ .

**Corrigé :**

- Rappelons tout d'abord que  $(X_n, Y_n)$  et  $(X, Y)$  sont bien des vecteurs aléatoires à valeurs dans  $\mathbb{R}^{k+l}$ , définis sur  $(\Omega, \mathcal{A}, P)$
- Avec les notations précédentes, l'inégalité triangulaire donne :
- $|(X_n, Y_n) - (X, Y)| \leq |(X_n, Y_n) - (X, Y_n)| + |(X, Y_n) - (X, Y)|$ .
- Par ailleurs,  $|(X_n, Y_n) - (X, Y_n)| = |X_n - X|$  et  $|(X, Y_n) - (X, Y)| = |Y_n - Y|$ .
- Avec le même raisonnement que dans le lemme précédent [VI.1], on montre que :
- $P(|(X_n, Y_n) - (X, Y)| \geq \varepsilon) \leq P(|X_n - X| \geq \frac{\varepsilon}{2}) + P(|Y_n - Y| \geq \frac{\varepsilon}{2})$ .
- Comme les deux quantités de droite convergent vers 0 quand  $n \rightarrow \infty$ , on en conclut la convergence en probabilité de  $(X_n, Y_n)$  vers  $(X, Y)$ .

**Corollaire VI.1** Soient  $X, X_1, X_2, \dots$  et  $Y, Y_1, Y_2, \dots$  deux suites de variables aléatoires réelles définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Si  $X_n \xrightarrow{P} X$  et  $Y_n \xrightarrow{P} Y$  et si  $f$  est une fonction réelle définie sur  $\mathbb{R}^2$ , continue sur un borélien  $B$  de  $\mathbb{R}^2$ , tel que  $P((X, Y) \in B) = 1$ , alors :

$$f(X_n, Y_n) \xrightarrow{P} f(X, Y).$$

**Démonstration :** Ceci est une conséquence immédiate de la propriété [VI.3] et du théorème [VI.6].

**Corollaire VI.2** Soient  $X, X_1, X_2, \dots$  et  $Y, Y_1, Y_2, \dots$  deux suites de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Si  $X_n \xrightarrow{P} X$  et  $Y_n \xrightarrow{P} Y$ , alors :

- (a)  $aX_n + bY_n \xrightarrow{P} aX + bY$ ,  $\forall a, b \in \mathbb{R}$
- (b)  $\frac{1}{X_n} \xrightarrow{P} \frac{1}{X}$  si  $P(X_n \neq 0) = P(X \neq 0) = 1$ ,  $\forall n$ .
- (c)  $X_n Y_n \xrightarrow{P} XY$

**Théorème VI.7 convergence en probabilité  $\Rightarrow$  convergence en loi**  
Soit  $X, X_1, X_2, \dots$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ .

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{\mathcal{L}} X.$$

**Démonstration :**

- $|\varphi_{X_n}(t) - \varphi_X(t)| = |E(e^{i\langle t, X_n \rangle} - e^{i\langle t, X \rangle})| \leq E(|e^{i\langle t, X_n \rangle} - e^{i\langle t, X \rangle}|)$  (voir cours sur l'intégration de fonctions à valeurs complexes).
- L'espérance précédente s'écrit comme :
- $\int_{|X_n - X| > \eta} |e^{i\langle t, X_n \rangle} - e^{i\langle t, X \rangle}| dP + \int_{|X_n - X| \leq \eta} |e^{i\langle t, X_n \rangle} - e^{i\langle t, X \rangle}| dP.$
- Comme  $|e^{i\langle t, X_n \rangle} - e^{i\langle t, X \rangle}| \leq 2$ , le premier terme est majoré par :
- $2P(\{|X_n - X| > \eta\}).$
- $|e^{i\langle t, X_n \rangle} - e^{i\langle t, X \rangle}|$  représente la longueur de la corde qui relie les points  $e^{i\langle t, X_n \rangle}$  et  $e^{i\langle t, X \rangle}$ , ces deux points étant sur le cercle unité.
- La longueur de la corde étant inférieure à la longueur de l'arc joignant les deux points, on en déduit :
- $|e^{i\langle t, X_n \rangle} - e^{i\langle t, X \rangle}| \leq |\langle t, X_n \rangle - \langle t, X \rangle|$ . Par Cauchy-Schwarz,
- $|e^{i\langle t, X_n \rangle} - e^{i\langle t, X \rangle}| \leq |\langle t, X_n \rangle - \langle t, X \rangle| \leq |t| \times |X_n - X|$ .
- Le deuxième terme est donc majoré par  $|t| \eta$ .
- Pour tout  $\varepsilon > 0$  ( $t$  étant donné), On peut choisir  $\eta$  tel que  $|t| \eta < \frac{\varepsilon}{2}$ .
- $\eta$  étant maintenant donné, il existe  $n$  tel que  $\forall m \geq n$ ,  $P(\{|X_n - X| > \eta\}) < \frac{\varepsilon}{4}$ , de par la convergence en probabilité de  $X_n$  vers  $X$ .
- Ceci montre que  $\forall m \geq n$ ,  $|\varphi_{X_n}(t) - \varphi_X(t)| < \varepsilon$ .
- $\varphi_{X_n}$  converge simplement vers  $\varphi_X$ , ce qui montre la convergence en loi de  $X_n$  vers  $X$ .

### 34.3 Convergence en moyenne

**Definition VI.1 Convergence en moyenne d'ordre  $p$** 

Soit  $X, X_1, X_2, \dots$  et  $Y, Y_1, Y_2, \dots$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On dit que  $X_n$  converge en moyenne d'ordre  $p$  (ou converge dans  $L_p$ ) vers  $X$  si  $X^p$  est intégrable ( $E(|X^p|) < \infty$ ) et si

$$\lim_{n \rightarrow \infty} E(|X_n - X|^p) = 0.$$

**Remarques :**

- On rappelle que l'application  $X \rightarrow \|X\|_p = (E[|X|^p])^{1/p}$  définit une norme sur l'ensemble des variables aléatoires intégrables à l'ordre  $p$ .
- On peut vérifier que l'on conserve une norme dans le cas de vecteurs aléatoires, en notant comme précédemment  $|X|$  la norme euclidienne de  $X$ .
- La définition peut se réécrire  $\|X_n - X\|_p \rightarrow 0$  quand  $n \rightarrow \infty$ , c'est-à-dire que l'on considère la convergence des vecteurs aléatoires pour la norme  $\|\cdot\|_p$ .
- Comme  $\|X_n\|_p \leq \|X\|_p + \|X_n - X\|_p$  et que  $\|X_n - X\|_p$ ,  $X_n$  a automatiquement des moments finis d'ordre  $p$  à partir d'un certain rang.

- Dans le cas où  $p = 2$ , on parle de convergence en moyenne quadratique.
- Dans le cas où  $p = 1$ , on parle de convergence en moyenne.
- Comme  $P$  est une probabilité, si  $r < p$ , nous avons déjà vu que  $\|X\|_r \leq \|X\|_p$ .
- Par conséquent, si  $\|X_n - X\|_p \rightarrow 0$ , alors  $\|X_n - X\|_r \rightarrow 0$ . La convergence en moyenne à l'ordre  $p$  entraîne la convergence en moyenne à tous les ordres inférieurs à  $p$ .
- En particulier, la convergence en moyenne quadratique (ou dans  $L_2$ ) entraîne la convergence en moyenne (ou dans  $L_1$ ).

**Propriété VI.1 convergence en moyenne implique convergence en probabilité**

Soit  $X, X_1, X_2, \dots$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . La convergence en moyenne de  $X_n$  vers  $X$  (notée  $X_n \xrightarrow{L_1} X$ ) entraîne la convergence en probabilité de  $X_n$  vers  $X$ .

**Démonstration :** voir exercice [VI.9].

**Exercice VI.9 convergence en moyenne implique convergence en probabilité**

Soit  $X, X_1, X_2, \dots$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Montrer que la convergence en moyenne de  $X_n$  vers  $X$  (notée  $X_n \xrightarrow{L_1} X$ ) entraîne la convergence en probabilité de  $X_n$  vers  $X$ .

**Corrigé :**

- $\|X_n - X\| = \int |X_n(\omega) - X(\omega)| dP(\omega)$ .
- Soit  $\varepsilon > 0$  quelconque.
- $\int |X_n(\omega) - X(\omega)| dP(\omega) \geq \int_{|X_n(\omega) - X(\omega)| > \varepsilon} |X_n(\omega) - X(\omega)| dP(\omega)$ .
- $\int_{|X_n(\omega) - X(\omega)| > \varepsilon} |X_n(\omega) - X(\omega)| dP(\omega) \geq \varepsilon P(|X_n(\omega) - X(\omega)| > \varepsilon)$ .
- On en conclut,  $\|X_n - X\| \rightarrow 0 \Rightarrow P(|X_n(\omega) - X(\omega)| > \varepsilon) \rightarrow 0$  pour tout  $\varepsilon > 0$ .
- La dernière propriété n'étant autre que la convergence en probabilité. Nous rappelons quelques propriétés étudiées dans le chapitre consacré aux espaces  $L^p$ .

**Propriété VI.2** Soit  $(X_n)$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$  définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On suppose que la série de terme général  $X_n$  est absolument convergente (on rappelle que l'absolue convergence est définie par  $\sum_n \|X_n\|_1 < \infty$ ), alors  $\sum_n X_n$  converge presque partout et dans  $L_1$ , c'est-à-dire qu'il existe un vecteur aléatoire  $X \in L_1$ , tel que :

- (a)  $\sum_{n \leq m} X_n \xrightarrow{p.s.} X$ ,
  - (b)  $\sum_{n \leq m} X_n \xrightarrow{L_1} X$ ,
- quand  $m \rightarrow \infty$ .

**Propriété VI.3** *Toute suite de cauchy dans  $L_1$  est convergente.*

Si  $X_n$  est une suite de cauchy dans  $L_1$ , il existe donc  $X$  vers lequel  $X_n$  converge en moyenne (et a fortiori en probabilité).

Nous énonçons maintenant un théorème de convergence énonçant des conditions suffisantes pour la convergence en moyenne. Par rapport au théorème de convergence dominée de Lebesgue, on ne suppose ici que la convergence en probabilité. En revanche, les variables aléatoires sont majorées par une constante et non simplement par une variable aléatoire intégrable.

**Théorème VI.1 convergence**

*Soit  $X_n$  une suite de vecteurs aléatoires définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  à valeurs dans  $\mathbb{R}^k$  ( $k \in \mathbb{N}$ ) et  $X$  un vecteur aléatoire. On suppose que les  $|X_n|$  sont presque sûrement majorés par une constante  $K$ ,  $K \geq 0$ , et que les  $X_n$  convergent en probabilité vers  $X$ , alors  $X_n$  converge en moyenne vers  $X$ .*

**démonstration :**

- Montrons d'abord que  $X \leq^{P_{p.s.}} K$ .  $|X| \leq |X - X_n| + |X_n| \leq |X - X_n| + K$ . Soit  $k \in \mathbb{N}$  et  $\omega \in \Omega$ .
- $|X(\omega)| > K + \frac{1}{k} \Rightarrow |X(\omega) - X_n(\omega)| + K > K + \frac{1}{k}$ . On a alors  $\{|X| > K + \frac{1}{k}\} \subset \{|X - X_n| > \frac{1}{k}\}$ .  $P(|X| > K + \frac{1}{k}) \leq P(|X - X_n| > \frac{1}{k})$ . De plus par la convergence en probabilité,  $\lim_{n \rightarrow \infty} P(|X - X_n| > \frac{1}{k}) = 0$ . D'où  $P(|X| > K + \frac{1}{k}) = 0$ . Comme  $\{|X| > K\} = \bigcup_{k \in \mathbb{N}} \{|X| > K + \frac{1}{k}\}$ ,  $P(|X| > K) = P\left(\bigcup_{k \in \mathbb{N}} \{|X| > K + \frac{1}{k}\}\right) \leq \sum_{k \in \mathbb{N}} P\left(|X| > K + \frac{1}{k}\right) = 0$ , on a  $P(|X| > K) = 0$ , ce qui montre que  $X \leq^{P_{p.s.}} K$ .
- Montrons maintenant le théorème. Soit  $\varepsilon > 0$ . Comme  $X_n$  converge en probabilité vers  $X$ ,  $\lim_{n \rightarrow \infty} P(|X_n - X| > \frac{\varepsilon}{3}) = 0$ . Donc il existe  $n_0 \in \mathbb{N}$ , tel que  $n \geq n_0 \Rightarrow P(|X_n - X| > \frac{\varepsilon}{3}) < \frac{\varepsilon}{3K}$ . Pour  $n \geq n_0$ ,  $E(|X_n - X|) = E(|X_n - X| \mathbb{I}_{|X_n - X| > \frac{\varepsilon}{3}}) + E(|X_n - X| \mathbb{I}_{|X_n - X| \leq \frac{\varepsilon}{3}}) \leq E\left((|X_n| + |X|)\mathbb{I}_{|X_n - X| > \frac{\varepsilon}{3}}\right) + \frac{\varepsilon}{3}E(\mathbb{I}_{|X_n - X| \leq \frac{\varepsilon}{3}})$ . Cette dernière quantité est majorée par  $2KP(|X_n - X| > \frac{\varepsilon}{3}) + \frac{\varepsilon}{3}$ , et donc par  $\varepsilon$ , ce qui montre que  $X_n$  converge en moyenne vers  $X$ .

Pour aller plus loin dans les liens entre convergence en probabilité et convergence en moyenne, il est utile d'introduire le concept d'uniforme intégrabilité. On rappelle au préalable que si  $|X|$  est un vecteur aléatoire intégrable ( $E^P(|X|) < \infty$ ), alors  $E^P(|X| \mathbb{I}_{|X| > \alpha}) \rightarrow 0$  quand  $\alpha \rightarrow 0$ . En effet, pour toute suite  $\alpha_n$  tendant vers 0, la suite  $X_n = |X| \mathbb{I}_{|X| > \alpha_n}$  converge  $P$  presque-sûrement vers 0 et les  $X_n$  sont majorés par  $X$  intégrable, ce qui montre l'assertion par le théorème de convergence dominée. Étant donné une suite de vecteurs aléatoires, le concept d'uniforme intégrabilité exprime que cette décroissance vers 0 se fait de manière uniforme.

### Definition VI.2 uniforme intégrabilité.

Soit  $X_i, i \in I$ , une famille de vecteurs aléatoires définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  à valeurs dans  $\mathbb{R}^k$  ( $k \in \mathbb{N}$ ). On dit que cette famille est uniformément intégrable si pour tout  $\varepsilon > 0$ , il existe  $K > 0$  tel que :

$$\forall i \in I, E^P(|X_i| \mathbb{I}_{|X_i|>K}) < \varepsilon.$$

La famille  $I$  n'est pas forcément dénombrable. L'uniforme intégrabilité implique l'intégrabilité et que la famille  $X_i$  est bornée dans  $L^1$ , c'est-à-dire que toutes les espérances des  $|X_i|$  sont majorées par une constante (indépendante de  $i$ ). En effet, prenons  $\varepsilon = 1$  et notons  $K_1$  tel que  $\forall i \in I, E^P(|X_i| \mathbb{I}_{|X_i|>K_1}) < 1$ . Comme  $E^P(|X_i|) = E^P(|X_i| \mathbb{I}_{|X_i|\leq K_1}) + E^P(|X_i| \mathbb{I}_{|X_i|>K_1}) < 1 + K_1$ . En revanche, une famille de vecteurs aléatoires intégrables n'est pas forcément uniformément intégrable. Il se peut même qu'une famille de vecteurs aléatoires intégrables soit bornée dans  $L^1$  sans être uniformément intégrable. Considérons par exemple le cas où  $P$  est la mesure uniforme sur  $[0, 1]$  et une suite de variables aléatoires  $X_n = n \mathbb{I}_{[0, 1/n]}$ . On a  $E^P[X_n] = 1, \forall n \in \mathbb{N}$ . Soit  $K > 0$  quelconque et  $n > K$ . On a  $E^P(|X_n| \mathbb{I}_{|X_n|>K}) = 1$ , ce qui montre que la suite  $X_n$  n'est pas uniformément intégrable. On remarque au passage que  $X_n \xrightarrow{P.p.s} 0$ , mais que  $E^P(X_n) = 1$  ne tend pas vers 0.

**Propriété VI.4 condition suffisante d'uniforme intégrabilité** Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et Soit  $X_i, i \in I$  une famille de vecteurs aléatoires définis sur  $(\Omega, \mathcal{A}, P)$  à valeurs dans  $\mathbb{R}^k$ ,  $k \in \mathbb{N}$ .

- (a) Si la famille  $X_i, i \in I$  est bornée dans  $L^p$  pour  $p > 1$ , c'est-à-dire s'il existe une constante  $A > 0$  telle que  $\forall i \in I, E^P(|X_i|^p) < A$ , alors la famille  $X_i$  est uniformément intégrable.
- (b) Si la famille  $X_i$  est dominée par une variable intégrable  $Y$ , c'est-à-dire si  $\forall i \in I, |X_i| \leq Y$  et  $E^P(Y) < \infty$ , alors la famille  $X_i$  est uniformément intégrable.

**démonstration** : la démonstration est laissée à titre d'exercice.

On remarque que la condition 2 est celle qui intervient dans le théorème de convergence dominée de Lebesgue. Énonçons maintenant le théorème qui relie convergence en probabilité en convergence en moyenne au moyen du concept d'uniforme intégrabilité.

**Théorème VI.2** Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé,  $X_n$  une suite de vecteurs aléatoires intégrables définis sur  $(\Omega, \mathcal{A}, P)$  à valeur dans  $\mathbb{R}^k$  ( $k \in \mathbb{N}$ ) et  $X$  un vecteur aléatoire intégrable. Alors  $X_n$  converge en moyenne (ou dans  $L^1$ ) vers  $X$  si et seulement si les deux conditions suivantes sont vérifiées.

- (a)  $X_n$  converge vers  $X$  en probabilité.
- (b) la suite  $X_n$  est uniformément intégrable.

## Chapitre 35

# Théorèmes de convergence

### 35.1 Loi des grands nombres

On rappelle pour mémoire le lemme de Borel-Cantelli démontré au chapitre I.

#### Propriété VI.1 Lemme de Borel-Cantelli

*Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et  $A_n$  une suite d'évènements de  $\mathcal{A}$ . On suppose que  $\sum P(A_n) < \infty$ . Alors  $\limsup A_n$  est négligeable.*

Il est également possible d'établir le résultat suivant :

#### Corollaire VI.1 théorème de Borel-Cantelli

*Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et  $A_n$  une suite d'évènements indépendants de  $\mathcal{A}$ .*

- (a) *Si  $\sum P(A_n) < \infty$ , alors  $\limsup A_n$  est négligeable.*
- (b) *Si  $\sum P(A_n)$  diverge, alors  $P(\limsup A_n) = 1$ .*

#### Démonstration :

- Vérifions la partie 2 du corollaire.
- On rappelle que  $P(\limsup A_n) = \lim_{n \rightarrow \infty} \downarrow P\left(\bigcup_{m \geq n} A_m\right)$ .
- Il faut donc montrer que pour tout  $n$   $P\left(\bigcup_{m \geq n} A_m\right) = 1$ .
- D'après les propriétés des opérations sur les ensembles,  $\left(\bigcup_{m \geq n} A_m\right)^c = \bigcap_{m \geq n} A_m^c$  et par conséquent  $P\left(\bigcup_{m \geq n} A_m\right) = 1 - P\left(\bigcap_{m \geq n} A_m^c\right)$ .
- $P\left(\bigcap_{m \geq n} A_m^c\right) = \lim_{N \rightarrow \infty} \downarrow P\left(\bigcap_{n \leq m \leq N} A_m^c\right)$ .
- D'après l'indépendance des  $A_n$ ,  $P\left(\bigcap_{n \leq m \leq N} A_m^c\right) = \prod_{n \leq m \leq N} P(A_m^c) = \prod_{n \leq m \leq N} (1 - P(A_m))$ .

- Comme  $1 = x \leq \exp x$ ,  $0 \leq 1 - P(A_m) \leq \exp(-P(A_m))$ .
- $P\left(\bigcap_{n \leq m \leq N} A_m^c\right) \leq \exp\left(-\sum_{n \leq m \leq N} P(A_m)\right)$ .
- Comme  $\sum P(A_n)$  diverge,  $\lim_{N \rightarrow \infty} \exp\left(-\sum_{n \leq m \leq N} P(A_m)\right) = 0$ .
- Donc,  $P\left(\bigcap_{m \geq n} A_m^c\right) = 0$ , ce qui montre le corollaire.

**Propriété VI.2** Soit  $X_n$  une suite de variables aléatoires réelles définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  et  $\mu$  une constante. On a :

$$X_n \xrightarrow{P} \mu \Leftrightarrow X_n \xrightarrow{\mathcal{L}} \mu.$$

**Démonstration :**

- L'implication est un résultat général déjà démontré. Montrons la réciproque.
- Soit  $\varepsilon > 0$ . On veut montrer que  $P(|X_n - \mu| > \varepsilon) \rightarrow 0$  quand  $n \rightarrow \infty$ , ou de manière équivalente que  $P(|X_n - \mu| \leq \varepsilon) \rightarrow 1$ .
- $P(|X_n - \mu| \leq \varepsilon) = E^P(\mathbb{1}_{|X_n - \mu| \leq \varepsilon}) = E^{P^{X_n}}(\mathbb{1}_{|x - \mu| \leq \varepsilon}) = \int \mathbb{1}_{|x - \mu| \leq \varepsilon} dP^{X_n}(x)$ .
- Considérons la fonction réelle continue  $f$ , nulle en dehors de  $[\mu - \varepsilon, \mu + \varepsilon]$ , telle que  $f(\mu) = 1$  et linéaire sur  $[\mu - \varepsilon, \mu]$ ,  $[\mu, \mu + \varepsilon]$ .  $f$  est continue, bornée et  $f(x) \leq \mathbb{1}_{|x - \mu| \leq \varepsilon}$ .
- D'où  $P(|X_n - \mu| \leq \varepsilon) \geq \int f(x) dP^{X_n}(x)$ .
- Comme  $P^{X_n} \xrightarrow{e} P^\mu = \delta(\mu)$ ,  $\int f(x) dP^{X_n}(x) \rightarrow \int f(x) dP^\mu(x) = f(\mu) = 1$ .
- Comme  $1 \geq P(|X_n - \mu| \leq \varepsilon) \geq \int f(x) dP^{X_n}(x)$ , on en déduit que  $P(|X_n - \mu| \leq \varepsilon) \rightarrow 1$ , ce qui montre la propriété.

**Théorème VI.1 Loi faible des grands nombres de Khintchine**

Soit  $X_n$  une suite de variables aléatoires réelles définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , indépendantes, de même loi, intégrables d'espérance  $E(X_n) = \mu$ . Alors,

$$\frac{1}{n} \sum_{m \leq n} X_m = \frac{S_n}{n} \xrightarrow{P} \mu.$$

**Démonstration :** voir exercice VI.10.

**Exercice VI.10 Loi faible des grands nombres de Khintchine**

Soit  $X_n$  une suite de variables aléatoires réelles définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , indépendantes, de même loi, intégrables d'espérance  $E(X_n) = \mu$ . Montrer en utilisant le théorème de Lévy et en faisant un développement limité que,

$$\frac{1}{n} \sum_{m \leq n} X_m = \frac{S_n}{n} \xrightarrow{P} \mu.$$

**Corrigé :**

- Montrons d'abord la convergence en loi de  $\frac{S_n}{n}$  vers  $\mu$ .
- On utilise les fonctions caractéristiques ; on note  $\varphi$ , la fonction caractéristique de  $X_n$  (ne dépendant pas de  $n$ ).

- D'après l'indépendance des  $\frac{X_m}{n}$ , la fonction caractéristique de  $\sum_{m \leq n} \frac{X_m}{n}$  s'écrit comme  $(\varphi(\frac{t}{n}))^n$ .
- Comme le premier moment existe, on en déduit que  $\varphi^{(1)}(0) = i\mu$ .
- En faisant un développement limité autour de zéro,  $\varphi(t) = 1 + i\mu t + o(t)$ .
- $(\varphi(\frac{t}{n}))^n = \exp(n \ln(1 + i\mu \frac{t}{n} + o(\frac{t}{n})))$  dont la limite quand  $n \rightarrow \infty$  est  $e^{it\mu}$ .
- La limite est donc la fonction caractéristique d'une variable aléatoire constante égale à  $\mu$ , ce qui montre la convergence en loi vers  $\mu$ .
- D'après la propriété précédente, la convergence en loi vers une constante implique la convergence en probabilité, ce qui montre la loi faible des grands nombres.

**Théorème VI.2 Loi forte des grands nombres de Kolmogorov**

Soit  $X_n$  une suite de variables aléatoires réelles définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , indépendantes, de même loi, intégrables d'espérance  $E(X_n) = \mu$ . Alors,

$$\frac{1}{n} \sum_{m \leq n} X_m = \frac{S_n}{n} \xrightarrow{p.s.} \mu.$$

**Théorème VI.3 Loi des grands nombres dans  $L_2$**

Soit  $X_n$  une suite de variables aléatoires réelles définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , non corrélées, de carré intégrable, et de même espérance  $\mu$  et variance  $\sigma^2$ . Alors,

$$\frac{1}{n} \sum_{m \leq n} X_m = \frac{S_n}{n} \xrightarrow{L_2} \mu.$$

**Démonstration :** voir exercice VI.11

**Remarques :** on demande ici l'existence des moments d'ordre 2. Par ailleurs, on n'a pas besoin de l'indépendance, il suffit que les  $X_n$  soient non corrélés. On n'a pas non plus besoin que les  $X_n$  soient de même loi, mais simplement que les deux premiers moments soient invariants. Bien sûr si les  $X_n$  sont iid, les hypothèses du théorème sont vérifiées. Enfin, rappelons que la convergence dans  $L^2$  implique la convergence en probabilité, mais que convergence en moyenne quadratique et convergence presque sûre ne sont pas comparables en toute généralité.

**Exercice VI.11 Loi des grands nombres dans  $L_2$**

Soit  $X_n$  une suite de variables aléatoires réelles définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , non corrélées, de carré intégrable, et de même espérance  $\mu$  et variance  $\sigma^2$ . Alors,

$$\frac{1}{n} \sum_{m \leq n} X_m = \frac{S_n}{n} \xrightarrow{L_2} \mu.$$

**Corrigé :**

- On cherche à montrer que  $E \left[ \left( \frac{S_n}{n} - \mu \right)^2 \right]$  tend vers zéro quand  $n \rightarrow \infty$ .
- Comme  $E \left[ \frac{S_n}{n} - \mu \right] = 0$ ,  $E \left[ \left( \frac{S_n}{n} - \mu \right)^2 \right] = \text{Var} \left[ \frac{S_n}{n} - \mu \right]$ .
- $\text{Var} \left[ \frac{S_n}{n} - \mu \right] = \text{Var} \left[ \frac{S_n}{n} \right] = \frac{1}{n^2} \text{Var} [S_n]$ .
- Comme les  $X_n$  sont non corrélés,  $\text{Var} [S_n] = n\sigma^2$ .
- D'où  $E \left[ \left( \frac{S_n}{n} - \mu \right)^2 \right] = \frac{\sigma^2}{n} \rightarrow 0$  quand  $n \rightarrow \infty$ , ce qui montre le théorème.

## 35.2 Théorème central limite

**Théorème VI.1 théorème central limite (Lindeberg-Lévy) unidimensionnel**

Soit  $(X_n)$  une suite de variables aléatoires réelles définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , indépendantes et de même loi, de carré intégrable ( $E(X_n^2) < \infty$ ). On note  $\mu$  et  $\sigma^2$  respectivement la moyenne et la variance de  $X_n$  et  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Alors,

$$\sqrt{n} \left( \frac{\bar{X}_n - \mu}{\sigma} \right) \xrightarrow{\mathcal{L}} N(0, 1).$$

**Remarque :** on rappelle que  $\xrightarrow{\mathcal{L}} N(0, 1)$  signifie la convergence étroite des mesures images associées aux variables aléatoires  $\sqrt{n} \left( \frac{\bar{X}_n - \mu}{\sigma} \right)$  vers la loi normale centrée réduite.

**Démonstration :** voir exercice VI.12.

**Exercice VI.12 théorème central limite (Lindeberg-Lévy) unidimensionnel**

Soit  $(X_n)$  une suite de variables aléatoires réelles définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , indépendantes et de même loi, de carré intégrable ( $E(X_n^2) < \infty$ ). On note  $\mu$  et  $\sigma^2$  respectivement la moyenne et la variance de  $X_n$  et  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ . En utilisant  $\varphi$ , la fonction caractéristique de  $\frac{X_i - \mu}{\sigma}$  et en faisant un développement limité à l'ordre 2 de la fonction caractéristique de  $\sqrt{n} \left( \frac{\bar{X}_n - \mu}{\sigma} \right)$ , montrer que :

$$\sqrt{n} \left( \frac{\bar{X}_n - \mu}{\sigma} \right) \xrightarrow{\mathcal{L}} N(0, 1).$$

**Corrigé :**

- Notons  $\varphi$  la fonction caractéristique de la variable aléatoire,  $\frac{X_i - \mu}{\sigma}$ . Comme les  $X_i$  sont de même loi, cette fonction caractéristique ne dépend pas de  $i$ .
- On peut réécrire  $\sqrt{n} \left( \frac{\bar{X}_n - \mu}{\sigma} \right)$  comme  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma}$ .

- La fonction caractéristique de  $\frac{1}{\sqrt{n}} \frac{X_i - \mu}{\sigma}$  s'écrit en fonction de  $\varphi$  comme  $t \rightarrow \varphi\left(\frac{t}{\sqrt{n}}\right)$ .
- Les variables aléatoires  $\frac{1}{\sqrt{n}} \frac{X_i - \mu}{\sigma}$  sont indépendantes. La fonction caractéristique de leur somme est égale à  $\left(\varphi\left(\frac{t}{\sqrt{n}}\right)\right)^n$ .
- On remarque que  $\frac{X_i - \mu}{\sigma}$  est d'espérance égale à zéro et de variance égale à 1. On en déduit les dérivées première et seconde de la fonction caractéristique en zéro :  $\varphi^{(1)}(0) = 0$  et  $\varphi^{(2)}(0) = -1$ .
- Un développement limité à l'ordre 2 en  $\frac{t}{\sqrt{n}}$  de  $\varphi$  donne :
- $\varphi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + \frac{t^2}{n} o\left(\frac{t}{\sqrt{n}}\right)$ , avec  $\lim_{x \rightarrow 0} o(x) = 0$ . D'où :
- $\left(\varphi\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 + \frac{1}{n} \left(-\frac{t^2}{2} + o\left(\frac{t}{\sqrt{n}}\right)\right)\right)^n$ . Cette dernière quantité s'écrit comme :
- $\exp\left[n \ln\left(1 + \frac{1}{n} \left[-\frac{t^2}{2} + o\left(\frac{t}{\sqrt{n}}\right)\right]\right)\right]$ .
- En utilisant les règles usuelles sur la composition des développements limités, on obtient que la dernière quantité converge vers  $e^{-t^2/2}$  quand  $n \rightarrow \infty$ , ce qui montre la convergence en loi vers la loi normale centrée réduite.

### Théorème VI.2 théorème central limite (Lindeberg-Levy) multidimensionnel

Soit  $(X_n)$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^p$ ,  $p \in \mathbb{N}$ , définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , indépendants et de même loi, de carré intégrable ( $E(|X_n|^2) < \infty$ ). On note  $\mu$  et  $\Sigma^2$  respectivement le vecteur moyenne et la matrice de variance-covariance de  $X_n$  et  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Alors,

$$\sqrt{n} (\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} N(0, \Sigma).$$

#### Démonstration :

- Soit  $t \in \mathbb{R}^p$ . Posons  $Y_n = t' X_n$  et  $\bar{Y}_n = \frac{1}{n} \sum_{j \leq n} Y_j$ . Les  $Y_n$  sont des variables aléatoires indépendantes et de même loi, d'espérance égale à  $t' \mu$  et de variance égale à  $t' \Sigma t$ .
- En utilisant le théorème central limite (unidimensionnel),  $\sqrt{n} \left( \frac{\bar{Y}_n - t' \mu}{(t' \Sigma t)^{1/2}} \right) \xrightarrow{\mathcal{L}} N(0, 1)$ .
- On vérifie facilement en utilisant les propriétés des fonctions caractéristiques que si une suite de variables aléatoires réelles  $Z_n$  converge en loi vers  $N(0, 1)$  alors pour tout  $\lambda \in \mathbb{R}$ ,  $\lambda Z_n \xrightarrow{\mathcal{L}} N(0, \lambda^2)$ .
- On en déduit,  $\sqrt{n} (\bar{Y}_n - t' \mu) \xrightarrow{\mathcal{L}} N(0, t' \Sigma t)$ .
- Notons  $X$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^p$ , distribué comme  $N(0, \Sigma)$  (on a vu l'existence de tels vecteurs aléatoires).  $t' X$  est une variable aléatoire gaussienne distribuée comme  $N(0, t' \Sigma t)$ .
- Comme  $\bar{Y}_n = t' \bar{X}_n$ , la convergence en loi de  $\sqrt{n} (\bar{Y}_n - t' \mu)$  vers une loi  $N(0, t' \Sigma t)$  signifie que  $t' [\sqrt{n} (\bar{X}_n - \mu)]$  converge en loi vers  $t' X$ .
- Comme la propriété précédente est vraie pour tout  $t \in \mathbb{R}^p$ , on en déduit le résultat annoncé d'après le théorème de Cramer-Wold.

### Propriété VI.1 approximation de la loi normale par une loi binomiale

Soit  $X_n$  une suite de variables aléatoires réelles, définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ ,  $X_n$  suivant une loi  $\mathcal{B}(n, p)$  (où  $\mathcal{B}(n, p)$  désigne la loi binomiale de paramètres  $n$  et  $p$ ). Alors,

$$\frac{X_n - np}{\sqrt{npq}} \xrightarrow{\mathcal{L}} N(0, 1),$$

où  $q = 1 - p$ .

**Démonstration :** Voir exercice (??).

**Corrigé :**

- On rappelle que si  $X$  est distribué comme une loi binomiale de paramètres  $n$  et  $p$ , alors  $P(X = x) = C_n^x p^x (1-p)^{n-x}$  pour  $x = 0, 1, 2, \dots, n$ . Si l'on considère  $n$  tirages avec remise parmi une urne contenant une proportion  $p$  de boules rouges, on a une probabilité  $P(X = x)$  de tirer  $x$  boules rouges. La fonction caractéristique de  $X$ ,  $\varphi_X$  est telle que  $\varphi_X(t) = E(e^{itX}) = \sum_{x=1}^n P(X = x)e^{itx} = \sum_{x=1}^n C_n^x (pe^{it})^x q^{n-x} = (pe^{it} + q)^n$  en vertu de la formule du binôme. En considérant les dérivées de  $\varphi$  en zéro, on en déduit les deux premiers moments de  $X$  :  $E(X) = np$ ,  $\text{Var}(X) = npq$ .
- Considérons une suite de variables aléatoires indépendantes, de Bernoulli,  $Y_i$ ,  $i = 1, 2, \dots$ , prenant la valeur 1 avec la probabilité  $p$  et 0 avec la probabilité  $q$  (il s'agit en fait d'une loi  $\mathcal{B}(1, p)$ ).
- On remarque que la fonction caractéristique de  $Y_i$  est égale à  $pe^{it} + q$ .
- $\sum_{i \leq n} Y_i$  admet pour fonction caractéristique  $(pe^{it} + q)^n$  et a donc la même loi que  $X_n$ .
- Par la définition même de la convergence en loi si une suite de variables aléatoires converge en loi vers une loi limite, toute suite de variables aléatoires de même loi converge en loi vers la même loi limite.
- Il suffit donc de vérifier que  $\frac{\sum_{i \leq n} Y_i - np}{\sqrt{npq}} \xrightarrow{\mathcal{L}} N(0, 1)$ .
- En notant  $\bar{Y}_n = \frac{1}{n} \sum_{i \leq n} Y_i$ , le terme de gauche s'écrit comme  $\sqrt{n} \frac{\bar{Y}_n - p}{\sqrt{pq}}$ .
- La propriété énoncée est alors une application directe du théorème central limite dans le cas de variables de Bernoulli.
- On peut ainsi simuler une loi normale à partir d'un simple jeu de pile ou face.

### Propriété VI.2 approximation de la loi normale multivariée par une loi multinomiale

Soit  $X_n$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^m$  ( $m \in \mathbb{N}$ ), définis sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ ,  $X_n$  suivant une loi multinomiale  $\mathcal{M}(n, p_1, \dots, p_m)$  de paramètres  $n$  et  $p_1, \dots, p_i, \dots, p_m$  ( $p_i > 0$ ,  $\sum_{i=1}^m p_i = 1$ ). On note  $p$  le vecteur de dimension  $m$  et de composantes  $p_i$ ,  $i = 1, \dots, m$  et on note  $q_i = 1 - p_i$ . Alors,

$$\frac{X_n - np}{\sqrt{n}} \xrightarrow{\mathcal{L}} N(0, \Sigma),$$

où  $\Sigma$  est la matrice de termes diagonaux  $p_i q_i$  et de termes croisés  $-p_i p_j$ .

– **Démonstration :**

- Quelques rappels sur la loi multinomiale.
- Si le vecteur aléatoire  $X = (X_1, \dots, X_m)$  est distribué comme une loi multinomiale de paramètres  $n$  et  $p = (p_1, \dots, p_m)$ , alors :
- $P(X_1 = x_1 \text{ et } \dots \text{ et } X_m = x_m) = \frac{n!}{x_1! \dots x_m!} p_1^{x_1} \dots p_m^{x_m}$
- pour  $x_1 = 0, 1, 2, \dots, n, \dots, x_m = 0, 1, 2, \dots, n$  et  $\sum_{i=1}^n x_i = n$ .
- La fonction caractéristique (vectorielle) de  $X$ ,  $\varphi_X$  est telle que :
- $\varphi_X(t_1, \dots, t_m) = \left( \sum_{j=1}^m p_j e^{it_j} \right)^n$ .
- On remarque que la loi de  $X$  est la loi de la somme de  $n$  vecteurs aléatoires  $Y_i = (Y_{i,1}, \dots, Y_{i,m})$  indépendants identiquement distribués suivant une loi multinomiale  $\mathcal{M}(1, p)$ .
- En considérant les dérivées en zéro de la fonction caractéristique, on en déduit que les deux premiers moments du vecteur aléatoire  $Y_i$  sont tels que :
- $E(Y_{i,j}) = p_j$ ,  $\text{Var}(Y_{i,j}) = p_j q_j$  (où  $q_j = 1 - p_j$ ) et  $\text{Cov}(Y_{i,j}, Y_{i,l}) = -p_j p_l$  pour  $j = 1, \dots, m$  et  $l = 1, \dots, m$ .
- On reconnaît la matrice  $\Sigma$  introduite dans l'énoncé de la propriété.
- Comme  $X_n \sim \sum_{i \leq n} Y_i$ , il suffit donc d'étudier la convergence (en loi) de  $\frac{\sum_{i \leq n} Y_i - np}{\sqrt{n}}$ .
- En notant  $\bar{Y}_n = \frac{1}{n} \sum_{i \leq n} Y_i$ , le terme précédent s'écrit comme  $\sqrt{n} (\bar{Y}_n - p)$ .
- La propriété énoncée est alors une application directe du théorème central limite (cas multidimensionnel).

**Propriété VI.3 Convergence de la loi de Poisson vers la loi normale**

Soit  $X$  une variable aléatoire distribuée selon une loi de Poisson  $\mathcal{P}(\lambda)$ .

Alors  $\frac{X-\lambda}{\sqrt{\lambda}} \xrightarrow{\mathcal{L}} N(0, 1)$  quand  $\lambda \rightarrow \infty$ .

– **Démonstration :**

- On rappelle qu'une variable aléatoire  $X$  suit une loi de Poisson de paramètre  $\lambda$  ( $\lambda > 0$ ), notée  $\mathcal{P}(\lambda)$  si :
- $P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$  pour  $x \in \mathbb{N}$ . On a par ailleurs,  $E(X) = \lambda$  et  $\text{Var}(X) = \lambda$ .
- La fonction caractéristique de  $X$  se calcule facilement comme  $\varphi_X(t) = e^{\lambda(e^{it}-1)}$ .
- Posons  $Y = \frac{X-\lambda}{\sqrt{\lambda}} = \frac{X}{\sqrt{\lambda}} - \sqrt{\lambda}$ . En utilisant les propriétés des fonctions caractéristiques,
- $\varphi_Y(t) = e^{-it\sqrt{\lambda}} \varphi_X\left(\frac{t}{\sqrt{\lambda}}\right) = \exp\left(-it\sqrt{\lambda} + \lambda(e^{it/\sqrt{\lambda}} - 1)\right)$ .
- $e^{it/\sqrt{\lambda}} - 1 = \frac{it}{\sqrt{\lambda}} + \frac{1}{2} \left(\frac{it}{\sqrt{\lambda}}\right)^2 + \left(\frac{it}{\sqrt{\lambda}}\right)^2 o\left(\frac{1}{\sqrt{\lambda}}\right)$ .
- $\lambda \left(e^{it/\sqrt{\lambda}} - 1\right) = it\sqrt{\lambda} - \frac{t^2}{2} \left(1 + o\left(\frac{1}{\sqrt{\lambda}}\right)\right)$ .
- On en conclut que  $\varphi_Y(t)$  tend vers  $e^{-t^2/2}$  quand  $\lambda \rightarrow \infty$ , ce qui montre la propriété.

**Propriété VI.4** Soit  $X_n$  une suite de variables aléatoires  $X_n$  distribuées selon des lois binomiales  $\mathcal{B}(n, p_n)$ . Si  $np_n \rightarrow \lambda$  quand  $n$  tend vers l'infini, alors,

$$X_n \xrightarrow{\mathcal{L}} \mathcal{P}(\lambda),$$

quand  $n$  tend vers l'infini ( $\mathcal{P}(\lambda)$  est la loi de Poisson de paramètre  $\lambda$ ).

– **Démonstration :**

- Notons  $\varphi_{X_n}(t)$  la fonction caractéristique de  $X_n$ .
- On a  $\varphi_{X_n}(t) = (q_n + p_n e^{it})^n = (1 + p_n(e^{it} - 1))^n$ .
- Comme  $np_n = \lambda + o\left(\frac{1}{n}\right)$ , on peut écrire  $\varphi_{X_n}(t)$  comme :
- $\varphi_{X_n}(t) = \exp(n \log(1 + \frac{1}{n}(\lambda + o(\frac{1}{n}))(e^{it} - 1)))$ .
- On voit alors que  $\varphi_{X_n}(t)$  tend vers  $\exp(\lambda(e^{it} - 1))$ , ce qui montre la propriété.

### 35.3 Convergence de la fonction de répartition empirique

**Definition VI.1 Empirical c.d.f.** Let  $X_n$  be a sequence of independent, identically distributed, real-valued random variables. The empirical cumulative distribution function  $F_n$  is defined by :

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{I}_{\{X_k \leq x\}}$$

For all  $n$ ,  $F_n$  is a simple function. You can use the  $F_n$  to approximate the c.d.f. of the  $X_n$  thanks to the following theorem.

**Théorème VI.3 (Gnedenko-Cantelli theorem)** Let  $X_n$  be a sequence of i.i.d. random variables and  $F_n$  the associated empirical c.d.f.'s. Then

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0 \quad \text{as } n \rightarrow \infty.$$

# Chapitre 36

## Espérance conditionnelle

### 36.1 Definition of Conditional Expectation

#### 36.1.1 General definition

Recall the definition of conditional probability associated with Bayes' Rule

$$\mathbb{P}(A|B) \equiv \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

For a discrete random variable  $X$  we have

$$\mathbb{P}(A) = \sum_x \mathbb{P}(A, X = x) = \sum_x \mathbb{P}(A|X = x)\mathbb{P}(X = x)$$

and the resulting formula for conditional expectation

$$\begin{aligned} E(Y|X = x) &= \int_{\Omega} Y(\omega)\mathbb{P}(dw|X = x) \\ &= \frac{\int_{X=x} Y(\omega)\mathbb{P}(dw)}{\mathbb{P}(X = x)} \\ &= \frac{E(Y\mathbf{1}_{(X=x)})}{\mathbb{P}(X = x)} \end{aligned}$$

We would like to extend this to handle more general situations where densities don't exist or we want to condition on very "complicated" sets.

**Definition VI.2** *Given a random variable  $Y$  with  $E|Y| < \infty$  defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  and some sub- $\sigma$ -field  $\mathcal{G} \subset \mathcal{A}$  we will define the **conditional expectation** as the almost surely unique random variable  $E(Y|\mathcal{G})$  which satisfies the following two conditions*

1.  $E(Y|\mathcal{G})$  is  $\mathcal{G}$ -measurable
2.  $E(YZ) = E(E(Y|\mathcal{G})Z)$  for all  $Z$  which are bounded and  $\mathcal{G}$ -measurable

Remark : one could replace 2. in the previous definition with :

$$\forall G \in \mathcal{G}, \quad E(Y\mathbf{1}_G) = E(E(Y|\mathcal{G})\mathbf{1}_G).$$

### Proof of existence and unicity

- **Existence** Using linearity, we need only consider  $X \geq 0$ . Define a measure  $Q$  on  $\mathcal{F}$  by  $Q(A) = E[X\mathbf{1}_A]$  for  $A \in \mathcal{F}$ . This is trivially absolutely continuous with respect to  $P|_{\mathcal{F}}$ , the restriction of  $P$  to  $\mathcal{F}$ . Let  $E[X|\mathcal{F}]$  be the Radon-Nikodym derivative of  $Q$  with respect to  $P|_{\mathcal{F}}$ . The Radon-Nikodym derivative is  $\mathcal{F}$ -measurable by construction and so provides the desired random variable.
- **Unicity** : If  $Y_1, Y_2$  are two  $\mathcal{F}$ -measurable random variables with  $E[Y_1\mathbf{1}_A] = E[Y_2\mathbf{1}_A]$  for all  $A \in \mathcal{F}$ , then  $Y_1 = Y_2$ , a.s., or conditional expectation is unique up to a.s. equivalence.

For  $\mathcal{G} = \sigma(X)$  when  $X$  is a discrete variable, the space  $\Omega$  is simply partitioned into disjoint sets  $\Omega = \sqcup G_n$ . Our definition for the discrete case gives

$$\begin{aligned} E(Y|\sigma(X)) &= E(Y|X) \\ &= \sum_n \frac{E(Y\mathbf{1}_{X=x_n})}{\mathbb{P}(X=x_n)} \mathbf{1}_{X=x_n} \\ &= \sum_n \frac{E(Y\mathbf{1}_{G_n})}{\mathbb{P}(G_n)} \mathbf{1}_{G_n} \end{aligned}$$

which is clearly  $\mathcal{G}$ -measurable. In general for  $\mathcal{G} = \sigma(X)$  :

#### Definition VI.3 Conditional expectation of $Y$ given $X$

Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$  and  $X$  another random variable defined on  $(\Omega, \mathcal{A}, P)$ . Define then  $E(Y | X)$  the conditional expectation of  $Y$  given  $X$  as  $E(Y | \sigma(X))$ .

**Proposition VI.1** Let  $(\Omega, \mathcal{A})$  be a measurable space,

$$Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$$

and  $X$  another real-valued random variable defined on  $(\Omega, \mathcal{A}, P)$ . As  $X = f(Y)$ , where  $f$  is measurable, real-valued function if and only if  $\sigma(X) \subset \sigma(Y)$ , we get that  $E(Y | X)$  is a measurable function of  $X$ .

**Proposition VI.2** Let  $(\Omega, \mathcal{A}, P)$  be a probability space, and  $X$  and  $Y$  two independent random variables such that  $Y$  is  $P$ -integrable. Then  $E(Y | X) = E(Y)$ ,  $P$ -almost surely.

Do not mix this notion with the following :

#### 36.1.2 Couples of random variables with p.d.f.

**Proposition VI.3** Let  $(X, Y)$  be a couple of real-valued random variables with p.d.f.  $f_{X,Y}(x, y)$  w.r.t. the Lebesgue measure on  $\mathbb{R}^2$ . Denote the respective marginal p.d.f. of  $X$  and  $Y$  as  $f_X(x)$  and  $f_Y(y)$ . Consider  $f_{X|Y}(x |$

$y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ . Then almost surely

$$\forall C \in \mathcal{B}, P(X \in C \mid Y = y) = \int_C f_{X|Y}(x \mid y) dx.$$

If besides  $X$  is  $P$ -integrable, then

$$E(X \mid Y = y) = \int_{\mathbb{R}} x f_{X|Y}(x \mid y) dx.$$

If  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$  is a measurable function such that  $g(X, Y)$  is integrable, then

$$E(g(X, Y) \mid Y = y) = \int_{\mathbb{R}} g(x, y) f_{X|Y}(x \mid y) dx.$$

**Remarks :** As soon as  $f_Y(y) > 0$ , this defines the distribution of  $X$  given that  $Y = y$ , described by p.d.f  $f_{X|Y}(x \mid y)$ , which is nonnegative and of integral 1.

If  $X$  and  $Y$  are independent,  $f_{X|Y} = f_X$  and  $f_{Y|X} = f_Y$ . To make the link with  $E[X|Y]$  would require to introduce the concept of regular conditional distribution.

Equation (VI.3) may be useful to compute the mathematical expectation of  $g(X, Y)$  as

$$E(g(X, Y)) = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} g(x, y) f_{X|Y}(x \mid y) dx \right) f_Y(y) dy.$$

## 36.2 Properties of Conditional Expectation

### 36.2.1 Conditional expectation

$E(\cdot|\mathcal{G})$  may be seen as an operator on random variables that transforms  $\mathcal{A}$ -measurable variables into  $\mathcal{G}$ -measurable ones.

Let us recall the basic properties of conditional expectation :

(a)  $E(\cdot|\mathcal{G})$  is positive :

$$Y \geq 0 \rightarrow E(Y|\mathcal{G}) \geq 0$$

(b)  $E(\cdot|\mathcal{G})$  is linear :

$$E(aX + bY|\mathcal{G}) = aE(X|\mathcal{G}) + bE(Y|\mathcal{G})$$

(c)  $E(\cdot|\mathcal{G})$  is a projection :

$$E(E(X|\mathcal{G})|\mathcal{G}) = E(X|\mathcal{G})$$

(d) More generally, the “tower property”. If  $\mathcal{H} \subset \mathcal{G}$  then

$$E(E(X|\mathcal{G})|\mathcal{H}) = E(X|\mathcal{H}) = E(E(X|\mathcal{H}) | \mathcal{G})$$

Proof : The right equality holds because  $E[X|\mathcal{H}]$  is  $\mathcal{H}$ - measurable, hence  $\mathcal{G}$ -measurable. To show the left equality, let  $A \in \mathcal{H}$ . Then since  $A$  is also in  $\mathcal{G}$ ,

$$E[E[X|\mathcal{G}]|\mathcal{H}]\mathbf{1}_A = E[E[X|\mathcal{G}]\mathbf{1}_A] = E[X\mathbf{1}_A] = E[E[X|\mathcal{H}]\mathbf{1}_A].$$

Since both sides are  $\mathcal{H}$ - measurable, the equality follows.

(e)  $E(\cdot|\mathcal{G})$  commutes with multiplication by  $\mathcal{G}$ -measurable variables :

$$E(XY|\mathcal{G}) = E(X|\mathcal{G})Y \text{ for } E|XY| < \infty \text{ and } Y \mathcal{G}\text{-measurable}$$

Proof : If  $A \in \mathcal{G}$ , then for any  $B \in \mathcal{G}$ ,

$$E[\mathbf{1}_A E[X|\mathcal{G}]\mathbf{1}_B] = E[E[X|\mathcal{G}]\mathbf{1}_{A \cap B}] = E[X\mathbf{1}_{A \cap B}] = E[(\mathbf{1}_A X)\mathbf{1}_B].$$

Since  $\mathbf{1}_A E[X|\mathcal{G}]$  is  $\mathcal{G}$ -measurable, this shows that the required equality holds when  $Y = \mathbf{1}_A$  and  $A \in \mathcal{G}$ . Using linearity and taking limits shows that the equality holds whenever  $Y$  is  $\mathcal{G}$ -measurable and  $X$  and  $XY$  are integrable.

(f)  $E(\cdot|\mathcal{G})$  respects monotone convergence :

$$0 \leq X_n \uparrow X \implies E(X_n|\mathcal{G}) \uparrow E(X|\mathcal{G})$$

(g) If  $\varphi$  is convex (in particular if  $\varphi(x) = x^2$ ) and  $E|\varphi(X)| < \infty$  then a conditional form of Jensen's inequality holds :

$$\varphi(E(X|\mathcal{G})) \leq E(\varphi(X)|\mathcal{G})$$

(h)  $E(\cdot|\mathcal{G})$  is a continuous contraction of  $\mathcal{L}^p$  for  $p \geq 1$  :

$$\|E(X|\mathcal{G})\|_p \leq \|X\|_p$$

and

$$X_n \xrightarrow{L^p} X \text{ implies } E(X_n|\mathcal{G}) \xrightarrow{L^p} E(X|\mathcal{G})$$

(i) Repeated Conditioning. For  $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \dots, \mathcal{G}_\infty = \sigma(\cup \mathcal{G}_i)$ , and  $X \in \mathcal{L}^p$  with  $p \geq 1$  then

$$\begin{aligned} E(X|\mathcal{G}_n) &\xrightarrow{\text{a.s.}} E(X|\mathcal{G}_\infty) \\ E(X|\mathcal{G}_n) &\xrightarrow{L^p} E(X|\mathcal{G}_\infty) \end{aligned}$$

(j) Best approximation property :

Suppose that the random variable  $X$  is square-integrable, but not measurable with respect to  $\mathcal{G}$ . That is, the information in  $\mathcal{G}$  does not completely determine the values of  $X$ . The conditional expectation,  $Y = E[X | \mathcal{G}]$ , has the property that it is the best approximation to  $X$  among functions measurable with respect to  $\mathcal{G}$ , in the least squares sense. That is, if  $\tilde{Y}$  is  $\mathcal{G}$ -measurable, then

$$E[(\tilde{Y} - X)^2] \geq E[(Y - X)^2].$$

It thus realizes the orthogonal projection of  $X$  onto a convex closed subset of a Hilbert space. This predicts the variance decomposition theorem that we shall see in a further section.

### 36.2.2 Conditional variance

**Definition VI.4** Let  $X$  be a square-integrable, real-valued random variable defined on a probability space  $(\Omega, \mathcal{A}, P)$ , and let  $\mathcal{F}$  be a sub- $\sigma$ -algebra of  $\mathcal{A}$ . Define the **conditional variance of  $X$  given  $\mathcal{F}$**  (denoted by  $Var(X | \mathcal{F})$ ) as the random variable  $E((X - E(X | \mathcal{F}))^2 | \mathcal{F})$ .

Define also the **conditional variance of  $X$  given a real-valued random variable  $Y$**  defined on  $(\Omega, \mathcal{A}, P)$  (denoted by  $Var(X | Y)$ ) as the random variable  $E((X - E(X | Y))^2 | Y)$ .

**Proposition VI.4**  $Var(X | \mathcal{F})$  and  $Var(X | Y)$  are well-defined, almost surely nonnegative and finite.

$$Var(X | \mathcal{F}) = E(X^2 | \mathcal{F}) - E(X | \mathcal{F})^2,$$

and

$$Var(X | Y) = E(X^2 | Y) - E(X | Y)^2.$$

#### Proposition VI.5 Variance decomposition formula

Let  $(X, Y)$  be a couple of random variables defined on a probability space  $(\Omega, \mathcal{A}, P)$ , such that  $X$  is square-integrable. Then

$$Var(X) = E(Var(X | Y)) + Var(E(X | Y)).$$

This may be very useful in non-life insurance to find the variance of a compound distribution.

**Proof :**

- $Var(X | Y) = E(X^2 | Y) - (E(X | Y))^2$ .
- $E[Var(X | Y)] = E[E(X^2 | Y)] - E[(E(X | Y))^2]$ .
- $E[E(X^2 | Y)] = E[X^2]$ .
- $E[Var(X | Y)] = E[X^2] - E[(E(X | Y))^2]$ .
- $Var(E(X | Y)) = E[(E(X | Y))^2] - (E[E(X | Y)])^2$ .
- $E[E(X | Y)] = E[X]$ .
- Hence  $Var(E(X | Y)) = E[(E(X | Y))^2] - (E[X])^2$ .

### 36.2.3 Compound distributions

Let  $(\Omega, \mathcal{A}, P)$  be a probability space, and

- $(X_n)_{n \in \mathbb{N}}$  a sequence of i.i.d., nonnegative random variables defined on  $(\Omega, \mathcal{A}, P)$ .  $X_n$  represents the severity of the  $n^{th}$  claim in the collective risk model.
- $N$  an random variable defined on  $(\Omega, \mathcal{A}, P)$  and taking values in  $\mathbb{N}$ , independent from the  $X_n$ . It represents the number of claims.
- Let  $S_N = X_1 + \dots + X_N$  represent the aggregate claim amount.

In many models we may know the mean and variance of  $N$  and  $X_1$ . How can we then get the mean and variance of  $S_N$ ? Simply by conditioning on the number of claims, and using conditional expectation and variance given  $N$ .

**Proposition VI.6** *First,*

$$ES_N = EN \cdot EX_1$$

*Moreover, thanks to the variance decomposition theorem, we may decompose  $Var(S_N)$  into two parts : the first one represents the part due to variability in claim amounts; the second one represents the part due to variability in the number of claims :*

$$Var(S_N) = EN \cdot Var(X_1) + (EX_1)^2 \cdot Var(N)$$

#### Formule de décomposition de la variance

Soit  $(X, Y)$  un couple de variables aléatoires réelles définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ ,  $X$  de carré intégrable. Montrer que :

$$\text{Var}(X) = E(\text{Var}(X | Y)) + \text{Var}(E(X | Y)).$$

#### Preuve :

- $\text{Var}(X | Y) = E(X^2 | Y) - (E(X | Y))^2$ .
- $E[\text{Var}(X | Y)] = E[E(X^2 | Y)] - E[(E(X | Y))^2]$ .
- $E[E(X^2 | Y)] = E[X^2]$ .
- $E[\text{Var}(X | Y)] = E[X^2] - E[(E(X | Y))^2]$ .
- $\text{Var}(E(X | Y)) = E[(E(X | Y))^2] - (E[E(X | Y)])^2$ .
- $E[E(X | Y)] = E[X]$ .
- D'où  $\text{Var}(E(X | Y)) = E[(E(X | Y))^2] - (E[X])^2$  et le résultat annoncé.



## Septième partie

# Bibliographie

- Basel-Committee (2001a) Operational risk, consultative document. Supporting Document to the New Basel Capital Accord.
- Basel-Committee (2001b) Sound practices for the management and supervision of operational risk Consultative Document.
- Basel-Committee (2001c) Working paper on the regulatory treatment of operational risk .
- Basel-Committee (2003) Supervisory guidance on operational risk advanced measurement approaches for regulatory capital .
- Charpentier, A., Denuit, M. (2004) Mathématiques de l'assurance non-vie, tome I. Economica.
- Durrett, R. (1999) Essentials of stochastic processes. Springer Texts in Statistics. Springer-Verlag, New York.
- Frachot, A., Moudoulaud, O., Roncalli, T. (2003) Loss distribution approach in practice. dans le livre : The Basel Handbook : A guide for financial practitioners .
- Giuffre, S. E. (2004) Operational risk : How good is the coverage ? .
- Jordan, J. S. (2003) Quantifying operational risk : A supervisors perspective. présentation lors du séminaire "Operational Risk in Banks and Financial Institutions" à Londres .
- Kuritzkes, A., Ziff, B. (2004) Operational risk : New approaches to measurement and modeling .
- Lopez, J. A. (2002) What is operational risk ? FRBSF Economic Letter (2).
- McNee, A. (2002) Rating operational risk. disponible en ligne sur le site [www.erisk.com](http://www.erisk.com) .
- OBrien, N., Smith, B., Allen, M. (2002) Operational risk - models. disponible en ligne sur le site [www.financewise.com](http://www.financewise.com) .
- Partrat, C., Besson, J.-L. (2004) Assurance non-vie. Modélisation, simulation. Economica.
- Pennequin, M., Roncalli, T., Salomon, E. (2004) La prise en compte de la diversification des risques opérationnels. Banque Magazine 660.
- Rolski, T., Schmidli, H., Schmidt, V., Teugels, J. (1999) Stochastic processes for insurance and finance. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester.

- Roncalli, T. (2004) Gestion des risques financiers. Economica.
- Rudin, W. (1987) Real and complex analysis, 3e Édition. McGraw-Hill Book Co., New York.
- Shaked, M. (1980) On mixtures from exponential families. *J. Roy. Statist. Soc. Ser. B* 42, 192–198.
- Tripp, M. H., Bradley, H. L., Devitt, R., Orros, G. C., Overton, G. L., Pryor, L. M., Shaw, R. A. (2004) Quantifying operational risk in general insurance companies. présenté à l'Institute of Actuaries .