

# Cours de probabilités, ECS deuxième année

Alain TROESCH

26 septembre 2012



# Table des matières

<b>1</b>	<b>Rappels de probabilités générales et discrètes</b>	<b>5</b>
1.1	Principes généraux du calcul des probabilités . . . . .	5
1.1.1	Probabilité d'un événement décrit par une expérience concrète . . . . .	5
1.1.2	Cas d'une expérience dont l'issue dépend du résultat d'une expérience antérieure . . . . .	7
1.1.3	Calcul de probabilités conditionnelles . . . . .	8
1.2	Rappels et compléments sur les variables aléatoires réelles discrètes . . . . .	8
1.2.1	Définitions . . . . .	8
1.2.2	Espérance . . . . .	9
1.2.3	Espérance conditionnelle . . . . .	9
1.2.4	Variance (dispersion) . . . . .	10
1.3	Rappel des lois classiques discrètes . . . . .	11
<b>2</b>	<b>Vecteurs aléatoires</b>	<b>13</b>
2.1	Lois d'un couple aléatoire discret . . . . .	13
2.1.1	Tribu associée à un couple de v.a.r.d. . . . .	13
2.1.2	Loi conjointe, lois marginales . . . . .	14
2.1.3	Loi d'un vecteur aléatoire . . . . .	15
2.1.4	Lois conditionnelles . . . . .	15
2.2	Indépendance de variables aléatoires . . . . .	16
2.2.1	Couples de v.a.r. indépendantes . . . . .	16
2.2.2	Familles de v.a.r. indépendantes . . . . .	16
2.2.3	Événements construits à l'aide de v.a.r. indépendantes . . . . .	17
2.3	Étude de $Z = g(X, Y)$ , généralisations à $Z = g(X_1, \dots, X_n)$ . . . . .	17
2.3.1	Tribu associée à $g(X, Y)$ . . . . .	17
2.3.2	Loi de $Z = g(X, Y)$ . . . . .	18
2.3.3	Espérance de $g(X, Y)$ . . . . .	18
2.3.4	Exemples : Espérance de $X + Y$ , de $XY$ . . . . .	18
2.3.5	Covariance, variance d'une somme . . . . .	19
2.3.6	Matrice des variances-covariances . . . . .	20
2.4	Stabilité des lois classiques . . . . .	21
2.4.1	Stabilité des lois binomiales . . . . .	21
2.4.2	Stabilité des lois de Pascal . . . . .	21
2.4.3	Stabilité des lois binomiales négatives . . . . .	21
2.4.4	Stabilité des lois de Poisson . . . . .	21

<b>3</b>	<b>Variables aléatoires à densité</b>	<b>23</b>
3.1	Densités . . . . .	23
3.1.1	Fonctions de répartition et densités . . . . .	23
3.1.2	Premier théorème de transfert : densité de $\varphi(X)$ dans des cas simples . . .	26
3.2	Moments d'une variable aléatoire réelle à densité . . . . .	27
3.2.1	Espérance . . . . .	27
3.2.2	Moments d'ordre 2 . . . . .	29
3.3	Somme de deux variables aléatoires à densité . . . . .	30
3.3.1	Densité d'une somme . . . . .	30
3.3.2	Espérance d'une somme . . . . .	30
3.3.3	Variance d'une somme . . . . .	31
<b>4</b>	<b>Loi continues classiques</b>	<b>33</b>
4.1	Loi uniforme . . . . .	33
4.1.1	Description et transformations simples . . . . .	33
4.1.2	Moments . . . . .	34
4.2	Loi exponentielle . . . . .	34
4.2.1	Description et transformations simples . . . . .	34
4.2.2	Moments . . . . .	35
4.2.3	Caractérisation par l'absence de mémoire . . . . .	35
4.3	Lois Gamma ( $\Gamma$ ) et gamma ( $\gamma$ ) . . . . .	36
4.3.1	Description et transformations simples . . . . .	36
4.3.2	Influence des paramètres sur la densité . . . . .	36
4.3.3	Moments . . . . .	37
4.3.4	Stabilité de la loi Gamma . . . . .	38
4.3.5	Cas particuliers de la loi $\Gamma$ . . . . .	38
4.4	Loi normale (ou loi gaussienne, ou loi de Laplace-Gauss) . . . . .	38
4.4.1	Description et transformations simples . . . . .	38
4.4.2	Étude de la densité et de la fonction de répartition de la loi normale centrée réduite . . . . .	39
4.4.3	Influence des paramètres . . . . .	39
4.4.4	Moments . . . . .	41
4.4.5	Stabilité de la loi normale . . . . .	41
4.5	Tableaux récapitulatif . . . . .	41
<b>5</b>	<b>Convergences et approximations en probabilité</b>	<b>45</b>
5.1	Convergence en probabilité . . . . .	45
5.1.1	Définition . . . . .	45
5.1.2	Inégalité de Bienaymé-Tchebychev . . . . .	45
5.1.3	Loi faible des grands nombres . . . . .	46
5.1.4	Un exemple classique (premier exemple d'estimation) . . . . .	46
5.1.5	Une condition suffisante de convergence en probabilité . . . . .	46
5.2	Convergence en loi . . . . .	47
5.2.1	Définition . . . . .	47
5.2.2	Théorème de la limite centrée . . . . .	47
5.3	Approximations . . . . .	47
5.3.1	Approximation d'une loi hypergéométrique par une loi binomiale . . . . .	47

---

5.3.2	Approximation d'une loi binomiale par une loi de Poisson . . . . .	48
5.3.3	Approximation d'une loi binomiale par une loi normale . . . . .	48
5.3.4	Approximation d'une loi de Poisson par une loi normale . . . . .	48
<b>6</b>	<b>Estimateurs</b> . . . . .	<b>49</b>
6.1	Échantillons d'une loi de probabilité . . . . .	49
6.2	Estimateurs, estimation ponctuelle . . . . .	50
6.2.1	Définitions . . . . .	50
6.2.2	Biais d'un estimateur . . . . .	50
6.2.3	Risque quadratique . . . . .	51
6.2.4	Convergence d'une suite d'estimateurs . . . . .	51
6.3	Estimation par intervalles de confiance . . . . .	52
6.3.1	Notion d'intervalle de confiance . . . . .	52
6.3.2	Exemple important : espérance d'une loi normale de $\sigma^2$ donné . . . . .	53
6.3.3	Cas typique : estimation d'une espérance à l'aide de la moyenne empirique . . . . .	54
6.3.4	Exemple : estimation par intervalle de confiance du paramètre d'une loi de Bernoulli . . . . .	54
<b>7</b>	<b>Éléments de statistiques</b> . . . . .	<b>55</b>
7.1	Définitions – Terminologie élémentaire . . . . .	55
7.1.1	Statistiques . . . . .	55
7.1.2	Effectif, fréquence . . . . .	56
7.1.3	Séries statistiques . . . . .	56
7.1.4	Représentations graphiques . . . . .	57
7.2	Paramètres de position . . . . .	58
7.2.1	Modes, moyennes . . . . .	58
7.2.2	Médiane . . . . .	59
7.3	Paramètres de dispersion . . . . .	59
7.3.1	Écart-type . . . . .	59
7.3.2	Quantiles . . . . .	60
7.4	Statistiques descriptives bivariées . . . . .	61
7.4.1	Définitions et représentations . . . . .	61
7.4.2	Fréquences marginales, fréquences conditionnelles . . . . .	63
7.4.3	Paramètres associés à une série statistique bivariée . . . . .	65
7.4.4	Droites de régression (ou droite des moindres carrés) . . . . .	66



# Probabilités – Chapitre 1

## Rappels de probabilités générales et discrètes

Le but de ce chapitre n'est pas de reprendre l'ensemble des résultats vu en première année, mais de les réordonner dans le but de forger certains automatismes de calcul. L'ordre dans lequel sont rappelés ces différents résultats n'est ici pas nécessairement l'ordre naturel de démonstration. Certains résultats ont été regroupés pour une meilleure analyse synthétique.

Dans tout ce qui suit, on suppose donné un espace probabilisé  $(\Omega, \mathcal{T}, P)$ .

### 1.1 Principes généraux du calcul des probabilités

#### 1.1.1 Probabilité d'un événement décrit par une expérience concrète

Nous exposons ici le principe général du calcul (et de la rédaction de ce calcul) d'une probabilité. Évidemment, dans la pratique, les situations sont très variées, et on peut être amené à s'écarter légèrement de la voie tracée ci-dessous.

La rédaction d'un calcul de probabilité se fait généralement en 4 étapes

1. Définition d'événements élémentaires issus de l'expérience décrite, en n'oubliant pas la numérotation de ces événements en cas d'expérience répétée. Il ne faut pas hésiter à introduire des événements par vous-même, à donner des notations. Ce travail n'est pas nécessairement fait pour vous dans le sujet.
2. Description à l'aide d'une **phrase** de l'événement dont on recherche la probabilité, sous la forme :

« L'événement  $A$  est réalisé si et seulement si ... »

Ce qu'on recherche est ici une condition nécessaire et suffisante de réalisation de l'événement  $A$  en fonction de la réalisation ou non des événements élémentaires définis dans la première étape. Il s'agit donc de « décomposer » l'événement  $A$  en événements élémentaires.

3. Traduction ensembliste (c'est-à-dire sous forme d'union, intersection ou complémentaires d'événements élémentaires) de la description verbale donnée ci-dessus. Les deux étapes sont nécessaires : la première (description verbale) fournit une explication facilement lisible, et éclaire la formule, la seconde (description ensembliste) donne de la rigueur à l'argument, et permet ensuite de s'orienter vers la bonne méthode de calcul.

4. Calcul de la probabilité  $P(A)$ , en se servant de la décomposition ensembliste précédente, et des règles de calcul des probabilités d'unions, d'intersections ou de complémentaires. Nous rappelons l'essentiel de ces règles ci-dessous.

Ainsi, il est important de savoir s'orienter rapidement vers la bonne technique de calcul (probabilité d'une union, d'une intersection), suivant la situation rencontrée. Voici les différents cas rencontrés les plus fréquemment :

1. **Calcul de la probabilité d'un complémentaire.**

Une seule formule à connaître, issue de la définition d'une mesure de probabilité :

$$\forall A \in \mathcal{T}, \quad P(\bar{A}) = 1 - P(A).$$

2. **Calcul de la probabilité d'une intersection.**

(a) **Intersection dénombrable d'événements décroissants pour l'inclusion.**

On utilise le théorème de la limite monotone :

THÉORÈME 1.1.1 *Soit  $(A_n)_{n \in \mathbb{N}^*}$  une suite décroissante d'événements, c'est-à-dire telle que pour tout  $n \in \mathbb{N}^*$ ,  $A_{n+1} \subset A_n$ . Alors*

$$P\left(\bigcap_{n=1}^{+\infty} A_n\right) = \lim_{n \rightarrow +\infty} P(A_n).$$

(b) **Cas d'un nombre fini d'événements mutuellement indépendants.**

Soit  $(A_1, \dots, A_n)$  une famille d'événements mutuellement indépendants. Alors :

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2) \dots P(A_n).$$

(c) **Cas d'un nombre fini d'événements non nécessairement indépendants.**

On utilise dans ce cas la formule des probabilités composées, obtenue par généralisation de la formule définissant les probabilités conditionnelles :

THÉORÈME 1.1.2 *Soit  $(A_1, \dots, A_n)$  une famille d'événements telle que  $P(A_1 \cap \dots \cap A_{n-1}) \neq 0$ . Alors :*

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P_{A_1}(A_2)P_{A_1 \cap A_2}(A_3) \dots P_{A_1 \cap \dots \cap A_{n-1}}(A_n).$$

Cette formule est surtout utile lorsqu'il y a un enchaînement temporel des événements ( $A_1$  étant le premier à advenir,  $A_n$  le dernier), chaque résultat influant sur les suivants. C'est le cas par exemple dans le cadre de tirages successifs dans une urne évolutive, l'évolution se faisant différemment suivant la boule tirée à un tirage donné.

3. **Calcul de la probabilité d'une union.**

(a) **Union dénombrable d'événements croissants pour l'inclusion.**

On utilise la deuxième version du théorème de la limite monotone :

THÉORÈME 1.1.3 *Soit  $(A_n)_{n \in \mathbb{N}^*}$  une suite croissante d'événements, c'est-à-dire telle que pour tout  $n \in \mathbb{N}^*$ ,  $A_n \subset A_{n+1}$ . Alors*

$$P\left(\bigcup_{n=1}^{+\infty} A_n\right) = \lim_{n \rightarrow +\infty} P(A_n).$$

**(b) Cas d'un nombre fini ou dénombrable d'événements 2 à 2 incompatibles.**

Dans le cas d'un nombre fini, c'est l'additivité d'une mesure de probabilité, dans le cas d'un nombre dénombrable, c'est la propriété de  $\sigma$ -additivité, issue de la définition d'une mesure de probabilité :

Soit  $(A_i)_{i \in I}$  une famille finie ou dénombrable d'événements 2 à 2 incompatibles. Alors

$$P\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} P(A_i).$$

**(c) Cas d'un nombre fini d'événements mutuellement indépendants.**

On a intérêt dans ce cas à se ramener à l'événement complémentaire. En effet, si  $A_1, \dots, A_n$  sont mutuellement indépendants, alors  $\overline{A_1}, \dots, \overline{A_n}$  aussi, et par conséquent :

$$P\left(\bigcup_{i=1}^n A_i\right) = 1 - P\left(\bigcap_{i=1}^n \overline{A_i}\right) = 1 - \prod_{i=1}^n (1 - P(A_i)),$$

la dernière égalité provenant de l'indépendance et de la règle des complémentaires.

**(d) Cas général lorsqu'on a des informations sur les intersections.**

On utilise la formule du crible de Poincaré : on donne les cas particuliers pour  $n = 2$ ,  $n = 3$  à connaître parfaitement, puis le cas général :

- $n = 2$  :  $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$
- $n = 3$  :  $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_2 \cap A_3) - P(A_1 \cap A_3) - P(A_1 \cap A_2) + P(A_1 \cap A_2 \cap A_3)$
- Plus généralement :

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{\ell=1}^n (-1)^{\ell+1} \sum_{1 \leq i_1 < \dots < i_\ell \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_\ell}).$$

**1.1.2 Cas d'une expérience dont l'issue dépend du résultat d'une expérience antérieure**

Dans certaines situations, par exemple lorsque le mode opératoire d'une expérience (par exemple la composition d'une urne) dépend du résultat d'une expérience précédente, il peut être nécessaire de discuter suivant le résultat obtenu lors de l'expérience antérieure : autrement dit, dans ces situations, il est aisé de calculer des probabilités conditionnellement à chacun des résultats possibles de la première expérience. Il faut ensuite récupérer la probabilité globale à partir de toutes ces probabilités conditionnelles. On utilise pour cela la formule des probabilités totales :

**THÉORÈME 1.1.4** *Soit  $(A_i)_{i \in I}$  un système complet ou quasi-complet, fini ou dénombrable, tel que pour tout  $i \in I$ ,  $P(A_i) \neq 0$ . Soit  $B \in \mathcal{T}$ . Alors la somme ci-dessous converge (absolument), et :*

$$P(B) = \sum_{i \in I} P(A_i) P_{A_i}(B).$$

A retenir : discussion à faire  $\rightarrow$  formule des probabilités totales

### 1.1.3 Calcul de probabilités conditionnelles

Puisque la probabilité conditionnelle  $P_B$  sachant un événement  $B$  donné est une mesure de probabilité, tous les résultats, et toutes les règles vues ci-dessus pour le calcul des probabilités s'applique au calcul des probabilités conditionnelles.

Quelques remarques et avertissements s'imposent néanmoins :

- Attention, on ne peut pas définir de notion d'événement conditionnel. Ce n'est pas l'événement qui est conditionnel, mais sa réalisation. Ainsi, si vous adoptez la démarche générale de rédaction :
  - \* ne décrivez pas l'événement conditionnel (cela n'a pas de sens), mais donnez bien une condition nécessaire et suffisante de réalisation, conditionnellement à un certain événement :
    - « Sachant que  $B$  est réalisé,  $A$  est réalisé si et seulement si ... »
  - \* Gardez-vous bien de transcrire cette phrase de manière ensembliste; cela vous amènerait inévitablement à considérer des « événements conditionnels ». Sautez cette étape et passez directement aux probabilités. Vous pouvez transcrire votre phrase de façon ensembliste à l'intérieur des probabilités, la condition étant alors précisée non pas au niveau des ensembles, mais au niveau de la probabilité.
- Lorsqu'on rajoute via certaines formules une condition à la mesure de probabilité conditionnelle initiale, cela revient à faire l'intersection des conditions. Ainsi, par exemple, la formule des probabilités totales est valable aussi pour le calcul d'une probabilité conditionnelle, et on obtient,  $(A_i)_{i \in I}$  étant un système complet au plus dénombrable tel que pour tout  $i \in I$   $P_C(A_i) \neq 0$  :

$$P_C(B) = \sum_{i \in I} P_C(A_i) P_{C \cap A_i}(B).$$

De même pour la formule des probabilités composées.

## 1.2 Rappels et compléments sur les variables aléatoires réelles discrètes

### 1.2.1 Définitions

DÉFINITION 1.2.1 Un *aléa numérique* est une application  $X$  de  $\Omega'$  dans  $\mathbb{R}$ , tel que :

- $\Omega'$  est un sous-ensemble de  $\Omega$
- pour tout  $a \in \mathbb{R}$ ,  $X^{-1}(] - \infty, a]) \in \mathcal{T}$  (donc l'événement  $X^{-1}(] - \infty, a])$  admet une probabilité).

Un aléa numérique  $X$  associe donc à certains éléments de  $\Omega$  une certaine valeur numérique, mais tous les éléments de  $\Omega$  n'ont pas forcément d'image par  $X$ .

DÉFINITION 1.2.2 Une *variable aléatoire réelle*  $X$  (v.a.r. en abrégé) est un aléa numérique tel que  $\Omega' = \Omega$ ; autrement dit, tous les éléments de  $\Omega$  ont une image par  $X$ .

DÉFINITION 1.2.3 Une *variable aléatoire réelle discrète*  $X$  (v.a.r.d. en abrégé) est une v.a.r. telle que  $X(\Omega)$  soit un ensemble fini ou dénombrable.

D'après les règles de cardinalité concernant les ensembles finis ou dénombrables, on obtient immédiatement :

PROPOSITION 1.2.4 Soit  $X$  et  $Y$  deux v.a.r.d.,  $\lambda$  un réel quelconque, et  $f$  une application de  $X(\Omega)$  dans  $\mathbb{R}$ . Alors, les v.a.r. définies de la manière suivante sont encore discrètes :

1.  $X + Y : \omega \mapsto X(\omega) + Y(\omega)$  ;
2.  $\lambda X : \omega \mapsto \lambda X(\omega)$  ;
3.  $XY : \omega \mapsto X(\omega)Y(\omega)$  ;
4.  $f(X) : \omega \mapsto f(X(\omega))$ .

PROPOSITION 1.2.5 Soit  $X$  une v.a.r.d. La famille d'événements  $([X = x])_{x \in X(\Omega)}$  est un système complet (ou quasi-complet) au plus dénombrable d'événements,

On rappelle que deux variables  $X$  et  $Y$  sont indépendantes si tous les couples d'événements  $([X = k], [Y = \ell])_{(k, \ell) \in X(\Omega) \times Y(\Omega)}$  sont indépendants. On définit de même des familles de variables aléatoires mutuellement indépendantes.

### 1.2.2 Espérance

On rappelle que pour que la nature et la somme d'une série ne dépende pas de l'ordre de sommation (donc d'une rénumération du terme général), il faut et il suffit que la série (pour le choix d'une énumération) soit absolument convergente. Pour toutes les sommes relatives aux variables aléatoires discrètes, cette absolue convergence est nécessaire, du fait qu'on n'impose pas d'ordre privilégié sur les éléments de  $X(\Omega)$ .

DÉFINITION 1.2.6 Soit  $X$  une v.a.r.d.. L'espérance de  $X$  est, sous réserve de convergence absolue :

$$E(X) = \sum_{x \in X(\Omega)} xP(X = x).$$

REMARQUE 1.2.7 L'espérance peut ne pas exister :

- série divergente :  $X(\Omega) = \mathbb{N}^*$  et  $\forall n \in \mathbb{N}^*, P(X = n) = \frac{6}{\pi^2 n^2}$ .
- série semi-convergente :  $X(\Omega) = \{(-1)^n n, n \in \mathbb{N}^*\}$ , et :  $\forall n \in \mathbb{N}^*, P(X = (-1)^n n) = \frac{6}{\pi^2 n^2}$ .

THÉORÈME 1.2.8 Soit  $X$  et  $Y$  deux v.a.r.d. admettant une espérance. Alors  $\lambda X + Y$  admet une espérance, et  $E(\lambda X + Y) = \lambda E(X) + E(Y)$ .

THÉORÈME 1.2.9 (de transfert) Soit  $X$  une v.a.r.d, et  $g : X(\Omega) \rightarrow \mathbb{R}$ . Alors  $E(g(X))$  existe si et seulement si  $\sum_{x \in X(\Omega)} g(x)P(X = x)$  converge absolument, et dans ce cas :

$$E(g(X)) = \sum_{x \in X(\Omega)} g(x)P(X = x)$$

### 1.2.3 Espérance conditionnelle

On rappelle que pour tout événement  $A$  non presque-impossible, la probabilité conditionnelle  $P_A$  est une mesure de probabilité. Cela justifie la définition suivante.

DÉFINITION 1.2.10 Soit  $A$  un événement non presque-impossible et  $X$  une variable aléatoire réelle discrète. Alors, on dit que  $X$  admet une espérance conditionnelle sachant  $A$  (ou espérance conditionnée à  $A$ ) si  $X$  admet une espérance conditionnelle pour la mesure de probabilité  $P_A$ , donc si la série

$$\sum_{x \in X(\Omega)} x P_A([X = x]) = \sum_{x \in X(\Omega)} x P(X = x | A)$$

est absolument convergente. Dans ce cas, on note  $E(X | A)$  l'espérance conditionnelle de  $X$  sachant  $A$ , définie par :

$$E(X | A) = \sum_{x \in X(\Omega)} x P(X = x | A).$$

LEMME 1.2.11 Soit  $X$  une variable aléatoire réelle discrète. Alors  $E(X)$  existe si et seulement si  $E(|X|)$  existe. En particulier, pour tout événement non presque-impossible  $A$ ,  $E(X | A)$  existe si et seulement si  $E(|X| | A)$  existe.

THÉORÈME 1.2.12 (Formule de l'espérance totale) Soit  $(A_n)_{n \in I}$  un système quasi-complet fini ou dénombrable, constitué d'événements non quasi-impossibles, et  $X$  une v.a.r.d.. Alors  $X$  admet une espérance si et seulement si les deux conditions ci-dessous sont satisfaites :

- (i) pour tout  $n \in I$ ,  $X$  admet une espérance conditionnelle sachant  $A_n$ , (donc  $|X|$  aussi) ;
- (ii) la série  $\sum_{n \in I} E(|X| | A_n) P(A_n)$  est convergente

Dans ce cas :

$$E(X) = \sum_{n \in I} E(X | A_n) P(A_n)$$

REMARQUES 1.2.13 1. Si on dispose d'un système complet (ou quasi-complet)  $(A_n)_{n \in I}$  d'événements dont certains sont presque-impossibles, on peut se ramener au cas précédent, en considérant le système obtenu en enlevant les parts de probabilité nulle. On obtient de la sorte encore un système quasi-complet, pour lequel on peut utiliser le théorème précédent, qui nous donne alors, en cas de convergence absolue, la formule :

$$E(X) = \sum_{\substack{n \in I \\ P(A_n) \neq 0}} E(X | A_n) P(A_n)$$

- 2. Dans le cas où le système complet  $(A_i)_{i \in I}$  est fini, si la condition (i) est satisfaite, la condition (ii) l'est aussi automatiquement, puisque la somme considérée est alors finie !
- 3. La série de la condition (ii) est positive, donc sa convergence équivaut à sa convergence absolue.

#### 1.2.4 Variance (dispersion)

DÉFINITION 1.2.14 Soit  $X$  une v.a.r.d. admettant une espérance  $E(X)$ . Alors, sous réserve de convergence de la série, la variance de  $X$  est :

$$V(X) = \sum_{x \in X(\Omega)} (x - E(X))^2 P(X = x) = E((X - E(X))^2)$$

L'écart-type est alors défini par :  $\sigma(X) = \sqrt{V(X)}$ .

REMARQUE 1.2.15 La variance peut ne pas exister même si l'espérance existe.

THÉORÈME 1.2.16 (Koenig-Huyghens) *Soit  $X$  admettant une variance. Alors :*

$$V(X) = E(X^2) - E(X)^2 \quad \text{et} \quad \forall a \in \mathbb{R}, V(X) = E((X - a)^2) - (E(X) - a)^2.$$

COROLLAIRE 1.2.17  $\forall (a, b) \in \mathbb{R}^2, V(X + b) = V(X)$  et  $V(aX) = a^2V(X)$ .

THÉORÈME 1.2.18 *Si  $X$  et  $Y$  sont indépendantes, alors  $V(X + Y) = V(X) + V(Y)$ .*

Cette formule se généralise au cas d'une famille de  $n$  variables mutuellement indépendantes.

On rappelle aussi (cela fera l'objet du chapitre suivant) que dans le cas d'une somme de variables non indépendantes, la variance peut se calculer à l'aide des covariances (cf chapitre suivant). Nous rappelons la formule :

$$V(X + Y) = V(X) + V(Y) + 2\text{cov}(X, Y).$$

Plus généralement :

$$V(X_1 + \dots + X_n) = \sum_{1 \leq i, j \leq n} \text{cov}(X_i, X_j) = \sum_{i=1}^n V(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j).$$

DÉFINITION 1.2.19 Si  $E(X) = 0$ , on dit que  $X$  est centrée ; si  $V(X) = 1$ , on dit que  $X$  est réduite.

De manière générale si  $E(X)$  et  $V(X)$  existent, la v.a.r.d.  $X^* = \frac{X - E(X)}{\sigma(X)}$  est centrée et réduite.

On l'appelle *v.a.r.d. centrée réduite associée à  $X$* .

## 1.3 Rappel des lois classiques discrètes

On se contente de redonner la description rapide des expériences associées aux différentes lois classiques, ainsi que le tableau récapitulatif :

- Loi issu d'un tirage unique
  - \* Loi de Bernoulli : succès (valeur 1) ou échec (valeur 0) lors d'un tirage déséquilibré
  - \* Loi uniforme  $\mathcal{U}(\llbracket 1, N \rrbracket)$  : valeur obtenue lors d'un tirage équiprobable dans une urne contenant  $N$  boules numérotées de 1 à  $N$ .
- Lois issues d'une répétition de tirages avec remise
  - \* Loi binomiale  $\mathcal{B}(n, p)$  : nombre de succès lors d'une répétition de  $n$  expériences de Bernoulli indépendantes de même paramètre.
  - \* Loi géométrique  $\mathcal{G}(p)$  : temps d'attente du premier succès lors d'une répétition d'expériences de Bernoulli indépendantes de même paramètre.
  - \* Loi de Pascal  $\mathcal{P}(r, p)$  (hors-programme) : temps d'attente du  $r$ -ième succès lors d'une répétition d'expériences de Bernoulli indépendantes de même paramètre.
  - \* Loi binomiale négative  $\mathcal{J}(r, p)$  (hors-programme) : Nombre d'échecs précédant le  $r$ -ième succès lors d'une répétition d'expériences de Bernoulli indépendantes de même paramètre.
- Lois issues d'une répétition de tirages sans remise
  - \* Loi hypergéométrique  $\mathcal{H}(N, n, p)$  : nombre de succès (boules blanches tirées) lors d'une succession de  $n$  tirages sans remise dans une urne contenant initialement  $N$  boules dont exactement  $Np$  boules blanches.
  - \* Temps d'attente du premier ou du  $r$ -ième succès, dans le même contexte (loi sans nom, hors-programme)

Nom	Paramètres	Notation	Valeurs	Loi	$E(X)$	$V(X)$
<b>Uniforme</b>	$n \in \mathbb{N}^*$	$\mathcal{U}(n)$	$\llbracket 1, n \rrbracket$	$P(X = k) = \frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
<b>Bernoulli</b>	$p \in ]0, 1[$	$\mathcal{B}(1, p)$	$\{0, 1\}$	$P(X = 1) = p$	$p$	$pq$
<b>Binomiale</b>	$(n, p)$	$\mathcal{B}(n, p)$	$\llbracket 0, n \rrbracket$	$P(X = k) = \binom{n}{k} p^k q^{n-k}$	$np$	$npq$
<b>Géométrique</b>	$p$	$\mathcal{G}(p)$	$\mathbb{N}^*$	$P(X = k) = pq^{k-1}$	$\frac{1}{p}$	$\frac{q}{p^2}$
Pascal	$(r, p)$	$\mathcal{P}(r, p)$	$\llbracket r, +\infty \llbracket$	$P(X = k) = \binom{k-1}{r-1} p^r q^{k-r}$	$\frac{r}{p}$	$\frac{rq}{p^2}$
Binomiale négative	$(r, p)$	$\mathcal{J}(r, p)$	$\mathbb{N}$	$P(X = k) = \binom{k+r-1}{k} p^r q^k$	$\frac{rq}{p}$	$\frac{rq}{p}$
<b>Hypergéométrique</b>	$(N, n, p)$	$\mathcal{H}(N, n, p)$	$\subset \llbracket 0, n \rrbracket$	$P(X = k) = \frac{\binom{Np}{k} \binom{Nq}{n-k}}{\binom{N}{n}}$	$np$	$npq \frac{N-n}{N-1}$
Attente du 1 <sup>er</sup> succès (tirage sans remise)	$(N, p)$		$\llbracket 1, Nq+1 \rrbracket$	$P(X = k) = \frac{Np}{k} \cdot \frac{\binom{Nq}{k-1}}{\binom{N}{k}}$		
Attente du $r^e$ succès (tirage sans remise)	$(r, N, p)$		$\llbracket r, Nq+r \rrbracket$	$P(X = k) = \frac{r}{k} \frac{\binom{Np}{r} \binom{Nq}{k-r}}{\binom{N}{k}}$		
<b>Poisson</b>	$\lambda \in \mathbb{R}_+$	$\mathcal{P}(\lambda)$	$\mathbb{N}$	$P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$	$\lambda$	$\lambda$

# Probabilités – Chapitre 2

## Vecteurs aléatoires

On s'intéresse à la corrélation existant entre plusieurs variables aléatoires, en nombre fini (vecteurs aléatoires) ou infini (familles de v.a.r.d.).

DÉFINITION 2.0.1 On appelle *couple aléatoire* sur l'espace probabilisé  $(\Omega, \mathcal{T}, P)$ , une application  $V : \Omega \rightarrow \mathbb{R}^2$ , donc qui à toute épreuve  $\omega$  de  $\Omega$ , associe un vecteur  $V(\omega)$  de  $\mathbb{R}^2$ , et telle que chaque coordonnée définit une variable aléatoire réelle. Ainsi, on peut écrire  $V = \begin{pmatrix} X \\ Y \end{pmatrix}$ , où  $X$  et  $Y$  sont

deux variables aléatoires, et où, pour tout  $\omega \in \Omega$ ,  $V(\omega) = \begin{pmatrix} X(\omega) \\ Y(\omega) \end{pmatrix}$ .

DÉFINITION 2.0.2 Plus généralement, un *vecteur aléatoire* à valeurs dans  $\mathbb{R}^n$  est une application  $V : \Omega \rightarrow \mathbb{R}^n$ , telle que chaque coordonnée définit une variable aléatoire réelle. Ainsi, on peut écrire  $V = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ , où  $X_1, \dots, X_n$  sont des v.a.r., et où, pour tout  $\omega \in \Omega$ ,  $V(\omega) = \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_n(\omega) \end{pmatrix}$ .

DÉFINITION 2.0.3 On dit que  $V$  est un couple (resp. un vecteur) aléatoire discret, si les coordonnées de  $V$  définissent des variables aléatoires réelles discrètes, donc, si, avec les notations ci-dessus,  $X$  et  $Y$  (resp.  $X_1, \dots, X_n$ ) sont des variables aléatoires réelles discrètes.

### 2.1 Lois d'un couple aléatoire discret

Dans toute cette section,  $X$  et  $Y$  désignent deux v.a.r.d..

#### 2.1.1 Tribu associée à un couple de v.a.r.d.

PROPOSITION 2.1.1 La famille d'événements  $([X = x] \cap [Y = y])_{(x,y) \in X(\Omega) \times Y(\Omega)}$  est une famille d'événements deux à deux incompatibles, dont l'union est égale à  $\Omega$ .

REMARQUE 2.1.2 Il ne s'agit en général pas d'un système complet, car certaines parts peuvent être vides.

DÉFINITION 2.1.3 Le *système complet associé au couple*  $(X, Y)$  est le système complet obtenu à partir de la famille  $([X = x] \cap [Y = y])_{(x,y) \in X(\Omega) \times Y(\Omega)}$  en supprimant les parts vides.

DÉFINITION 2.1.4 La *tribu associée au couple*  $(X, Y)$  est la tribu engendrée par le système complet associé à  $(X, Y)$ . On la note souvent  $\mathcal{A}_{(X,Y)}$  ou  $\mathcal{T}_{(X,Y)}$  (notation à redéfinir dans un devoir).

PROPOSITION 2.1.5 On a les inclusions suivantes :  $\mathcal{A}_X \subset \mathcal{A}_{(X,Y)}$  et  $\mathcal{A}_Y \subset \mathcal{A}_{(X,Y)}$ .

La définition de la tribu associée à un couple se généralise immédiatement à toute famille  $(X_k)_{k \in K}$  de variables aléatoires. On notera  $\mathcal{A}_{(X_k)_{k \in K}}$ .

## 2.1.2 Loi conjointe, lois marginales

NOTATION 2.1.6 Pour alléger les notations, on note  $P(X = x, Y = y)$  pour  $P([X = x] \cap [Y = y])$ .

DÉFINITION 2.1.7 La *loi conjointe* du couple  $(X, Y)$  est la donnée de la fonction :

$$\begin{aligned} p : X(\Omega) \times Y(\Omega) &\longrightarrow [0, 1] \\ (x, y) &\longmapsto P(X = x, Y = y). \end{aligned}$$

Étant donné des énumérations  $(x_i)_{i \in I}$  et  $(y_j)_{j \in J}$  de  $X(\Omega)$  et  $Y(\Omega)$  ( $I$  fini ou  $I = \mathbb{N}$ ; de même pour  $J$ ), la donnée de la loi conjointe de  $X$  et  $Y$  est équivalente à la donnée de la suite doublement indexée  $(p_{i,j})_{(i,j) \in I \times J}$ , où :

$$\forall (i, j) \in I \times J, \quad p_{i,j} = P(X = x_i, Y = y_j).$$

DÉFINITION 2.1.8 1. La *première loi marginale* du couple  $(X, Y)$  est la loi de la variable aléatoire  $X$ . Étant donné une énumération  $(x_i)_{i \in I}$  de  $X(\Omega)$ , la donnée de la première loi marginale est équivalente à la donnée d'une suite, généralement notée  $(p_{i,\bullet})_{i \in I}$ , où :

$$\forall i \in I, \quad p_{i,\bullet} = P(X = x_i).$$

2. La *seconde loi marginale* du couple  $(X, Y)$  est la loi de la variable aléatoire  $Y$ . Étant donné une énumération  $(y_j)_{j \in J}$  de  $Y(\Omega)$ , la donnée de la seconde loi marginale est équivalente à la donnée d'une suite, généralement notée  $(p_{\bullet,j})_{j \in J}$ , où :

$$\forall j \in J, \quad p_{\bullet,j} = P(Y = y_j).$$

PROPOSITION 2.1.9 1.  $\forall i \in I, \quad p_{i,\bullet} = \sum_{j \in J} p_{i,j}$     2.  $\forall j \in J, \quad p_{\bullet,j} = \sum_{i \in I} p_{i,j}$ .

REMARQUE 2.1.10 Le point dans la notation des lois marginales indique qu'on a sommé sur toutes les valeurs possibles de l'indice qu'il remplace.

REMARQUE 2.1.11 Les lois marginales sont déterminées par la loi conjointe. **Réciproque fautive !**

EXEMPLE 2.1.12 Soit une urne contenant 1 boule blanche, 1 boule noire. On effectue un tirage, avec équiprobabilité. On note  $X_1, Y_1$  et  $X_2$  les trois variables (égales l'une à l'autre) égales à 1 si on tire la boule noire, et 0 sinon. On note  $Y_2$  la variable égale à 1 si on tire la boule blanche,

et 0 sinon. Alors, les lois marginales de  $(X_1, Y_1)$  et de  $(X_2, Y_2)$  sont les mêmes, pourtant les lois conjointes sont distinctes.

La loi conjointe de  $(X, Y)$  est souvent plus facile à déterminer que les lois de  $X$  et  $Y$ . On se sert donc du calcul de cette loi conjointe pour déterminer ensuite les lois de  $X$  et  $Y$ .

EXEMPLE 2.1.13 On lance deux dés équilibrés, on note  $X$  le maximum, et  $Y$  le minimum. Déterminer la loi conjointe de  $(X, Y)$ , puis les lois marginales.

PROPOSITION 2.1.14 (*Propriété caractéristique d'une loi conjointe*) (*réciproque admise*)  
Une suite  $(p_{i,j})_{i \in I, j \in J}$  ( $I$  et  $J$  finis ou dénombrables) est la loi d'un couple  $(X, Y)$  ssi :

1.  $\forall (i, j) \in I \times J, p_{i,j} \geq 0$ ;
2.  $\sum_{(i,j) \in I \times J} p_{i,j} = 1$ .

REMARQUE 2.1.15 Toutes les sommes considérées dans ce paragraphe sont à termes positifs. Donc leur convergence équivaut à leur convergence absolue, et l'ordre de sommation importe peu. Ainsi, toutes les sommes qu'on a considérées sont bien définies.

### 2.1.3 Loi d'un vecteur aléatoire

Dans tout ce paragraphe,  $V = (X_1, \dots, X_n)$  est un vecteur aléatoire discret.

DÉFINITION 2.1.16 La loi du vecteur  $V$  est la donnée de l'application

$$\varphi_V : X_1(\Omega) \times \dots \times X_n(\Omega) \rightarrow \mathbb{R}$$

qui à  $(x_1, \dots, x_n)$  associe  $\varphi_V(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$ .

PROPOSITION 2.1.17 On a : 
$$\sum_{(x_1, \dots, x_n) \in X_1(\Omega) \times \dots \times X_n(\Omega)} \varphi_V(x_1, \dots, x_n) = 1.$$

DÉFINITION 2.1.18 La  $k$ -ième loi marginale du vecteur  $V$  est la loi de sa  $k$ -ième coordonnée  $X_k$ .

PROPOSITION 2.1.19 On a, pour tout  $x_k \in X_k(\Omega)$  :

$$P(X_k = x_k) = \sum_{(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n) \in X_1(\Omega) \times \dots \times X_{k-1}(\Omega) \times X_{k+1}(\Omega) \times \dots \times X_n(\Omega)} \varphi_V(x_1, \dots, x_n)$$

### 2.1.4 Lois conditionnelles

DÉFINITION 2.1.20 Soit  $A$  un événement non quasi-impossible. Alors  $(\Omega, \mathcal{T}, P_A)$  est un espace probabilisé, et la loi conditionnelle de  $X$  sachant  $A$  est la loi de  $X$  dans cet espace probabilisé, c'est-à-dire la loi obtenue pour la mesure de probabilité  $P_A$ . Il s'agit donc de la fonction  $x \mapsto P_A(X = x)$ , pour  $x \in X(\Omega)$ .

En particulier, étant donné une deuxième variable aléatoire  $Y$ , et  $y \in Y(\Omega)$  tel que  $P(Y = y)$  est non nul, la loi conditionnelle de  $X$  sachant  $Y = j$  est la loi de  $X$  pour la mesure  $P_{[Y=j]}$ .

PROPOSITION 2.1.21 *Étant donné une énumération  $(x_i)_{i \in I}$  de  $X(\Omega)$  et une énumération  $(y_j)_{j \in J}$  de  $Y(\Omega)$ , ainsi qu'un indice  $j \in J$ , la loi conditionnelle de  $X$  sachant que  $Y = y_j$  est donnée par :*

$$\forall i \in I, P_{[Y=y_j]}(X = x_i) = \frac{p_{i,j}}{p_{\bullet,j}}.$$

## 2.2 Indépendance de variables aléatoires

### 2.2.1 Couples de v.a.r. indépendantes

Dans ce paragraphe,  $X$  et  $Y$  désignent deux v.a.r.d..

DÉFINITION 2.2.1 On dit que  $X$  et  $Y$  sont *indépendantes* si pour tout  $x \in X(\Omega)$  et pour tout  $y \in Y(\Omega)$ , les événements  $[X = x]$  et  $[Y = y]$  sont indépendants.

PROPOSITION 2.2.2 *Les propositions suivantes sont équivalentes :*

1.  $X$  et  $Y$  sont indépendantes ;
2. les systèmes complets  $([X = x])_{x \in X(\Omega)}$  et  $([Y = y])_{y \in Y(\Omega)}$  sont indépendants ;
3. Les  $\sigma$ -algèbres  $\mathcal{A}_X$  et  $\mathcal{A}_Y$  sont indépendantes.

#### Rappels

- Si  $(A_i)_{i \in I}$  est un système complet fini ou dénombrable, alors la  $\sigma$ -algèbre engendrée par  $(A_i)_{i \in I}$  est constitué des ensembles  $\bigcup_{i \in J} A_i$ , pour tout sous-ensemble  $J$  de  $I$ .
- Si deux mesures de probabilités coïncident sur un système complet  $(A_i)_{i \in I}$ , elles coïncident sur la  $\sigma$ -algèbre engendrée par ce système complet.

PROPOSITION 2.2.3 *Si  $X$  et  $Y$  sont indépendantes, alors :*

$$\forall (x, y) \in X(\Omega) \times Y(\Omega), P(X = x, Y = y) = P(X = x)P(Y = y).$$

*Autrement dit, étant donné des énumérations  $(x_i)_{i \in I}$  et  $(y_j)_{j \in J}$  :*

$$\forall (i, j) \in I \times J, p_{i,j} = p_{i,\bullet} \cdot p_{\bullet,j}.$$

Ainsi, si  $X$  et  $Y$  sont indépendantes (et seulement dans ce cas), les lois marginales déterminent la loi conjointe.

### 2.2.2 Familles de v.a.r. indépendantes

Soit  $(X_k)_{k \in K}$  une famille quelconque de v.a.r.d..

DÉFINITION 2.2.4 On dit que les variables aléatoires  $X_k$ , pour  $k \in K$ , sont mutuellement indépendantes si pour toute famille  $(x_k)_{k \in K}$  telle que pour tout  $k \in K$ ,  $x_k \in X_k(\Omega)$ , les événements  $[X_k = x_k]$ , pour  $k \in K$ , sont mutuellement indépendants.

PROPOSITION 2.2.5 *Les propositions suivantes sont équivalentes :*

1. les variables  $X_k$ , pour  $k \in K$ , sont mutuellement indépendantes ;
2. les systèmes complets  $([X_k = x])_{x \in X_k(\Omega)}$ , pour  $k \in K$ , sont mutuellement indépendants ;
3. les  $\sigma$ -algèbres  $\mathcal{A}_{X_k}$ , pour  $k \in K$ , sont mutuellement indépendantes.

### 2.2.3 Événements construits à l'aide de v.a.r. indépendantes

Soit  $(X_k)_{k \in K}$  une famille de v.a.r.d..

On formalise la notion d'événement ne dépendant que d'un certain nombre des variables  $X_k$  : soit  $K' \subset K$  ; dire que  $E$  ne dépend que des variables  $X_k$ , pour  $k \in K'$  signifie que  $E \in \mathcal{A}_{(X_k)_{k \in K'}}$ .

Supposons que les variables  $X_k$ ,  $k \in K$ , sont mutuellement indépendantes. Le lemme 2.2.6 permet de formaliser le raisonnement intuitif suivant : si  $E$  et  $F$  sont deux événements dépendant uniquement de variables  $X_k$  sans variable commune à  $E$  et  $F$ , alors  $E$  et  $F$  sont indépendants.

**LEMME 2.2.6 (Lemme des coalitions)** *Soit  $(X_k)_{k \in K}$  une famille de v.a.r.d. mutuellement indépendantes, et soit  $K'$  et  $K''$  des sous-ensembles de  $K$  tels que  $K' \cap K'' = \emptyset$ . Alors les  $\sigma$ -algèbres  $\mathcal{A}_{(X_k)_{k \in K'}}$  et  $\mathcal{A}_{(X_k)_{k \in K''}}$  sont indépendantes.*

Ce lemme se généralise bien sûr à un nombre plus important de sous-ensembles deux à deux disjoints de  $K$ . On obtient alors l'indépendance mutuelle des  $\sigma$ -algèbres correspondantes.

Ainsi, dans la situation décrite plus haut pour les événements  $E$  et  $F$ , il existe deux sous-ensembles disjoints  $K'$  et  $K''$  de  $K$  tels que  $E$  ne dépend que des  $X_k$  pour  $k \in K'$  et  $F$  ne dépend que des  $X_k$  pour  $k \in K''$ . Autrement dit :  $E \in \mathcal{A}_{(X_k)_{k \in K'}}$  et  $F \in \mathcal{A}_{(X_k)_{k \in K''}}$ . Mais d'après le lemme,  $\mathcal{A}_{(X_k)_{k \in K'}}$  et  $\mathcal{A}_{(X_k)_{k \in K''}}$  sont indépendantes, donc  $E$  et  $F$  sont indépendantes.

## 2.3 Étude de $Z = g(X, Y)$ , généralisations à $Z = g(X_1, \dots, X_n)$

Dans cette section, on désigne par  $X$  et  $Y$  deux v.a.r.d., et par  $g$  une fonction de  $X(\Omega) \times Y(\Omega)$  dans  $\mathbb{R}$ . On définit la variable aléatoire  $g(X, Y)$  par :

$$\forall \omega \in \Omega, \quad g(X, Y)(\omega) = g(X(\omega), Y(\omega)).$$

On a :  $g(X, Y)(\Omega) \subset \text{Im } g$ .

Dans les généralisations à des vecteurs aléatoires,  $n$  désignera un entier supérieur ou égal à 2,  $g$  désignera une fonction de  $\mathbb{R}^n$  dans  $\mathbb{R}$ , et  $X_1, \dots, X_n$  seront  $n$  variables aléatoires réelles discrètes

### 2.3.1 Tribu associée à $g(X, Y)$

**PROPOSITION 2.3.1** *Soit  $z \in \text{Im } g$ . Alors :  $[g(X, Y) = z] = \bigcup_{(x, y) \in g^{-1}(z)} [X = x] \cap [Y = y]$ .*

**PROPOSITION 2.3.2**  $\mathcal{A}_{g(X, Y)} \subset \mathcal{A}_{(X, Y)}$ .

Plus généralement, on obtient une description de l'événement  $[g(X_1, \dots, X_n) = z]$  :

**PROPOSITION 2.3.3** *Soit  $z \in \text{Im } g$ . Alors :*

$$[g(X_1, \dots, X_n) = z] = \bigcup_{(x_1, \dots, x_n) \in g^{-1}(z)} [X_1 = x_1] \cap \dots \cap [X_n = x_n].$$

On obtient alors de même :

**PROPOSITION 2.3.4**  $\mathcal{A}_{g(X_1, \dots, X_n)} \subset \mathcal{A}_{(X_1, \dots, X_n)}$ .

### 2.3.2 Loi de $Z = g(X, Y)$

PROPOSITION 2.3.5 *La loi de  $g(X, Y)$  est donnée par :*

$$g(X, Y)(\Omega) \subset \text{Im } g \quad \text{et} \quad \forall z \in \text{Im } g, \quad P(g(X, Y) = z) = \sum_{(x, y) \in g^{-1}(z)} P(X = x, Y = y).$$

*En particulier, si  $X$  et  $Y$  sont indépendantes :*

$$\forall z \in \text{Im } g, \quad P(g(X, Y) = z) = \sum_{(x, y) \in g^{-1}(z)} P(X = x)P(Y = y).$$

EXEMPLES 2.3.6 1.  $\forall z \in \mathbb{R}, \quad P(X + Y = z) = \sum_{x+y=z} P(X = x, Y = y).$

2.  $\forall z \in X(\Omega) \cup Y(\Omega), \quad P(\min(X, Y) = z) = \sum_{y \geq z} P(X = z, Y = y) + \sum_{x > z} P(X = x, Y = z).$

Exemple : tirage de deux dés.

3. Si  $X \hookrightarrow \mathcal{G}(p)$  et  $Y \hookrightarrow \mathcal{G}(p)$ , et si  $X$  et  $Y$  sont indépendantes, alors  $X + Y \hookrightarrow \mathcal{P}(2, p).$

Plus généralement :

PROPOSITION 2.3.7 *La loi de  $g(X_1, \dots, X_n)$  est donnée par :  $g(X_1, \dots, X_n)(\Omega) \subset \text{Im } g$  et*

$$\forall z \in \text{Im } g, \quad P(g(X_1, \dots, X_n) = z) = \sum_{(x_1, \dots, x_n) \in g^{-1}(z)} P(X_1 = x_1, \dots, X_n = x_n).$$

*En particulier, si  $X_1, \dots, X_n$  sont mutuellement indépendantes :*

$$\forall z \in \text{Im } g, \quad P(g(X_1, \dots, X_n) = z) = \sum_{(x_1, \dots, x_n) \in g^{-1}(z)} P(X_1 = x_1) \cdots P(X_n = x_n).$$

### 2.3.3 Espérance de $g(X, Y)$

THÉORÈME 2.3.8 (théorème de transfert) *L'espérance de  $g(X, Y)$  existe si et seulement si la série ci-dessous converge absolument, et dans ce cas :*

$$E(g(X, Y)) = \sum_{(x, y) \in X(\Omega) \times Y(\Omega)} g(x, y)P(X = x, Y = y).$$

Plus généralement :

THÉORÈME 2.3.9 (théorème de transfert) *L'espérance de  $g(X_1, \dots, X_n)$  existe si et seulement si la série ci-dessous converge absolument, et dans ce cas :*

$$E(g(X_1, \dots, X_n)) = \sum_{(x_1, \dots, x_n) \in X_1(\Omega) \times \cdots \times X_n(\Omega)} g(x_1, \dots, x_n)P(X_1 = x_1, \dots, X_n = x_n).$$

### 2.3.4 Exemples : Espérance de $X + Y$ , de $XY$

THÉORÈME 2.3.10 *Soit  $X$  et  $Y$  deux v.a.r.d. admettant une espérance. Alors l'espérance de  $X + Y$  existe, et  $E(X + Y) = E(X) + E(Y).$*

COROLLAIRE 2.3.11 (linéarité de l'espérance) *Soit  $X$  et  $Y$  deux v.a.r.d., et  $\lambda$  et  $\mu$  deux réels. Alors  $E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y).$*

COROLLAIRE 2.3.12 Si  $X$  et  $Y$  admettent des espérances, et que  $X \leq Y$  (c'est-à-dire pour tout  $\omega \in \Omega$ ,  $X(\omega) \leq Y(\omega)$ ), alors  $E(X) \leq E(Y)$ .

REMARQUE 2.3.13 Attention, de manière générale,  $E(XY) \neq E(X)E(Y)$ .

EXEMPLE 2.3.14 Soit  $X_1$ ,  $X_2$ , et  $X_3$ , indépendantes, suivant des lois de Bernoulli de paramètre  $p$ . Soit  $Y_1 = X_1X_2$  et  $Y_2 = X_2X_3$ . Alors  $E(Y_1Y_2) \neq E(Y_1)E(Y_2)$ .

THÉORÈME 2.3.15 Si  $X$  et  $Y$  sont indépendantes et admettent une espérance, alors  $XY$  admet une espérance, et  $E(XY) = E(X)E(Y)$ .

### 2.3.5 Covariance, variance d'une somme

DÉFINITION 2.3.16 Sous réserve d'existence, on définit la *covariance* de  $X$  et  $Y$  par :

$$\text{cov}(X, Y) = E((X - E(X))(Y - E(Y))).$$

REMARQUE 2.3.17 Si  $\text{cov}(X, Y) > 0$ ,  $X - E(X)$  et  $Y - E(Y)$  ont tendance à être de même signe, donc  $X$  et  $Y$  ont tendance à se situer du même côté de leur espérance. Ainsi :

- $\text{cov}(X, Y) > 0$  signifie que  $X$  et  $Y$  ont tendance à être de variation parallèle;
- $\text{cov}(X, Y) < 0$  signifie que  $X$  et  $Y$  ont tendance à être de variation opposée;

LEMME 2.3.18 Si  $X$  et  $Y$  admettent des moments d'ordre 2, alors  $XY$  admet une espérance.

PROPOSITION 2.3.19 Si  $X$  et  $Y$  admettent des moments d'ordre 2, alors  $\text{cov}(X, Y)$  existe.

PROPRIÉTÉS 2.3.20 Soit  $X$  et  $Y$  admettant des moments d'ordre 2.

1.  $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$  ;
2.  $\text{cov}(X, Y) = \text{cov}(Y, X)$  (*symétrie*)
3.  $\text{cov}$  est bilinéaire. Autrement dit, pour toutes v.a.r.d.  $X, X', Y$  et  $Y'$  admettant des moments d'ordre 2 :

(a) (*linéarité par rapport à la première variable*)

$$\forall (\lambda, \mu) \in \mathbb{R}^2, \quad \text{cov}(\lambda X + \mu X', Y) = \lambda \text{cov}(X, Y) + \mu \text{cov}(X', Y);$$

(b) (*linéarité par rapport à la seconde variable*)

$$\forall (\lambda, \mu) \in \mathbb{R}^2, \quad \text{cov}(X, \lambda Y + \mu Y') = \lambda \text{cov}(X, Y) + \mu \text{cov}(X, Y');$$

4. Si  $X$  et  $Y$  sont indépendantes,  $\text{cov}(X, Y) = 0$ . **Réciproque fausse !**

PROPOSITION 2.3.21 Si  $X$  et  $Y$  admettent une variance, alors  $X + Y$  également, et :

$$V(X + Y) = V(X) + V(Y) + 2 \cdot \text{cov}(X, Y).$$

THÉORÈME 2.3.22 Si  $X$  et  $Y$  sont indépendantes, et admettent une variance, alors  $X + Y$  aussi, et  $V(X + Y) = V(X) + V(Y)$ .

PROPOSITION 2.3.23 Soit  $X_1, \dots, X_n$  des v.a.r.d. ayant un moment d'ordre 2;  $X_1 + \dots + X_n$  admet une variance et :

$$V(X_1 + \dots + X_n) = V(X_1) + \dots + V(X_n) + 2 \cdot \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j).$$

EXEMPLE 2.3.24 Calcul de la variance d'une variable aléatoire suivant une loi hypergéométrique.

THÉORÈME 2.3.25 Sous les mêmes hypothèses, si  $X_1, \dots, X_n$  sont deux à deux indépendantes :

$$V(X_1 + \dots + X_n) = V(X_1) + \dots + V(X_n).$$

PROPOSITION 2.3.26 Soit  $X$  et  $Y$  deux v.a.r.d. admettant des moments d'ordre 2. Alors

$$\text{cov}(X, Y)^2 \leq V(X)V(Y).$$

DÉFINITION 2.3.27 Le coefficient de corrélation linéaire de  $X$  et  $Y$  est :  $\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$ .

PROPOSITION 2.3.28 Soit  $X$  et  $Y$  deux v.a.r.d. admettant des moments d'ordre 2. Alors

$$|\rho_{X,Y}| \leq 1.$$

On a l'égalité  $|\rho_{X,Y}| = 1$  si et seulement si il existe  $a$  et  $b$  deux réels tels que presque sûrement,  $Y = aX + b$ . Ainsi, il existe une relation affine entre  $X$  et  $Y$ . De plus, le signe de  $\rho_{X,Y}$  donne le signe de  $a$ .

### 2.3.6 Matrice des variances-covariances

De ce qui précède, la forme cov est bilinéaire symétrique positive (mais pas définie) sur l'espace des variables aléatoires réelles discrètes

PROPOSITION 2.3.29 Si  $X$  et  $Y$  sont deux variables linéairement indépendantes, elles forment une base de Vect( $X, Y$ ), et la matrice de la forme bilinéaire cov dans cette base est :

$$\begin{pmatrix} V(X) & \text{cov}(X, Y) \\ \text{cov}(X, Y) & V(Y) \end{pmatrix}.$$

Cette matrice est appelée matrice des variances-covariances de  $(X, Y)$ .

COROLLAIRE 2.3.30 Pour tout  $(a, b, a', b') \in \mathbb{R}^4$  :

$$\begin{cases} \text{cov}(aX + bY, a'X + b'Y) &= \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} V(X) & \text{cov}(X, Y) \\ \text{cov}(X, Y) & V(Y) \end{pmatrix} \begin{pmatrix} a' \\ b' \end{pmatrix} \\ V(aX + bY) &= \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} V(X) & \text{cov}(X, Y) \\ \text{cov}(X, Y) & V(Y) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \end{cases}$$

La définition des matrices de variances-covariances s'étend au cas où  $X$  et  $Y$  sont liés, ainsi que la propriété ci-dessus.

DÉFINITION 2.3.31 On définit de même la matrice des variances-covariances d'un vecteur aléatoire discret  $X = (X_1, \dots, X_n)$  par :

$$V(X) = \begin{pmatrix} V(X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & V(X_2) & \cdots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \cdots & V(X_n) \end{pmatrix}$$

PROPOSITION 2.3.32 Pour tout  $(\lambda_1, \dots, \lambda_n, \mu_1, \dots, \mu_n) \in \mathbb{R}^{2n}$  :

$$\left\{ \begin{array}{l} \text{cov}(\lambda_1 X_1 + \cdots + \lambda_n X_n, \mu_1 X_1 + \cdots + \mu_n X_n) = (\lambda_1 \ \cdots \ \lambda_n) \begin{pmatrix} V(X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & V(X_2) & \cdots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \cdots & V(X_n) \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} \\ \\ V(\lambda_1 X_1 + \cdots + \lambda_n X_n) = (\lambda_1 \ \cdots \ \lambda_n) \begin{pmatrix} V(X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & V(X_2) & \cdots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \cdots & V(X_n) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \end{array} \right.$$

## 2.4 Stabilité des lois classiques

Par stabilité, on entend la chose suivante : si  $X$  et  $Y$  suit un certain type de loi, alors  $X + Y$  aussi.

### 2.4.1 Stabilité des lois binomiales

PROPOSITION 2.4.1 Soit  $(m, n) \in (\mathbb{N}^*)^2$  et  $p \in ]0, 1[$ . Si  $X \hookrightarrow \mathcal{B}(n, p)$  et  $Y \hookrightarrow \mathcal{B}(m, p)$  sont indépendantes, alors  $X + Y \hookrightarrow \mathcal{B}(n + m, p)$ .

En particulier si  $X \hookrightarrow \mathcal{B}(n, p)$ ,  $X$  est la somme de  $n$  variables de Bernoulli de paramètre  $p$  mutuellement indépendantes.

### 2.4.2 Stabilité des lois de Pascal

PROPOSITION 2.4.2 Soit  $(r, s) \in (\mathbb{N}^*)^2$ , et  $p \in ]0, 1[$ . Si  $X \hookrightarrow \mathcal{P}(r, p)$  et  $Y \hookrightarrow \mathcal{P}(s, p)$  sont indépendantes, alors  $X + Y \hookrightarrow \mathcal{P}(r + s, p)$ .

En particulier si  $X \hookrightarrow \mathcal{P}(r, p)$ ,  $X$  est la somme de  $r$  variables suivant une loi géométrique de paramètre  $p$ , mutuellement indépendantes.

### 2.4.3 Stabilité des lois binomiales négatives

PROPOSITION 2.4.3 Soit  $(r, p) \in (\mathbb{N}^*)^2$  et  $p \in ]0, 1[$ . Si  $X \hookrightarrow \mathcal{J}(r, p)$  et  $Y \hookrightarrow \mathcal{J}(s, p)$  sont indépendantes, alors  $X + Y \hookrightarrow \mathcal{J}(r + s, p)$ .

### 2.4.4 Stabilité des lois de Poisson

PROPOSITION 2.4.4 Soit  $(\lambda, \mu) \in (\mathbb{R}_+)^2$ . Si  $X \hookrightarrow \mathcal{P}(\lambda)$  et  $Y \hookrightarrow \mathcal{P}(\mu)$  sont indépendantes, alors  $X + Y \hookrightarrow \mathcal{P}(\lambda + \mu)$ .



# Probabilités – Chapitre 3

## Variables aléatoires à densité

Vous connaissez une première grande famille de variables aléatoires : les variables aléatoires réelles discrètes, prenant un nombre restreint de valeurs possibles (nombre fini, ou dénombrable, donc notamment pas une valeur quelconque dans un intervalle donné). Nous présentons dans ce chapitre une deuxième grande famille de variables aléatoires réelles, pour lesquelles les valeurs prises ne sont pas en nombre au plus dénombrable (ce sont généralement des intervalles de  $\mathbb{R}$ ). Nous définissons pour ces variables aléatoires les notions d'espérance et de variance. Il n'est plus possible pour cela de considérer des sommes, qui ne sont définies que pour un nombre fini, ou dénombrable (en se ramenant à des séries) de termes. Nous utiliserons par conséquent des intégrales, parfois impropres, notamment si l'ensemble des valeurs prises par une variable aléatoire  $X$  est un intervalle non borné.

### 3.1 Densités

Pour la plupart des variables aléatoires, la notion de probabilité ponctuelle (c'est-à-dire  $P(X = \alpha)$ ) n'a pas de sens, car lorsque  $X$  prend ses valeurs dans un intervalle  $[a, b]$ , imposer une égalité  $X = \alpha$  est trop précis. Il peut arriver que  $X$  soit très proche de  $\alpha$ , mais il est très peu probable que  $X$  prenne, avec une précision infinie, la valeur exacte de  $\alpha$ . Pour les variables à densité telles que nous les définiront, nous aurons même, pour tout  $\alpha$  dans  $X(\Omega)$ ,  $P(X = \alpha) = 0$ .

Ainsi, il ne sera pas possible de définir de manière générale une loi de probabilité par la seule donnée des probabilités ponctuelles. Nous décrirons donc une loi de probabilité par sa fonction de répartition, ce qui nous amènera ensuite à la notion de densité.

#### 3.1.1 Fonctions de répartition et densités

**DÉFINITION 3.1.1** Soit  $X$  une variable aléatoire réelle. La fonction de répartition  $F_X$  de  $X$  est la fonction définie sur  $\mathbb{R}$  par :

$$\forall x \in \mathbb{R}, \quad F_X(x) = P(X \leq x).$$

**PROPOSITION 3.1.2 (Rappel : caractérisation d'une fonction de répartition)**

Une fonction  $F : \mathbb{R} \rightarrow \mathbb{R}$  est la fonction de répartition d'une variable aléatoire  $X$  si et seulement si :

- (i)  $F$  est croissante
- (ii)  $F$  est continue à droite en tout point
- (iii)  $F$  admet des limites en  $-\infty$  et  $+\infty$ , et  $\lim_{x \rightarrow -\infty} F(x) = 0$ , et  $\lim_{x \rightarrow +\infty} F(x) = 1$ .

DÉFINITION 3.1.3 On dit que  $X$  est une *variable aléatoire réelle à densité* si sa fonction de répartition  $F_X$  est :

- (i) continue sur  $\mathbb{R}$  ;
- (ii) de classe  $\mathcal{C}^1$  sur  $\mathbb{R}$  sauf éventuellement en un nombre fini de points, c'est-à-dire qu'il existe  $\{x_1, \dots, x_n\}$  tels que  $F$  est de classe  $\mathcal{C}^1$  sur  $\mathbb{R} \setminus \{x_1, \dots, x_n\}$ .

DÉFINITION 3.1.4 Soit  $X$  une variable aléatoire réelle à densité, et soit  $f$  une fonction réelle définie sur  $\mathbb{R}$ . On dit que  $f$  est **une densité** de  $X$  si :

- (i)  $f$  est positive ou nulle sur  $\mathbb{R}$  ;
- (ii) il existe un nombre fini de points  $x_1 < \dots < x_n$  tels que

$$\forall x \in \mathbb{R} \setminus \{x_1, \dots, x_n\}, \quad f(x) = F'_X(x).$$

Ainsi, une densité  $f$  est **presque partout** la dérivée de la fonction de répartition.

REMARQUES 3.1.5 1. L'hypothèse de positivité de  $f$  est nécessairement vérifiée sur  $\mathbb{R} \setminus \{x_1, \dots, x_n\}$ , puisque  $F_X$  est croissante

- 2. On n'a pas unicité d'une densité de  $f$ . Toute variable aléatoire réelle à densité admet même une infinité de densités. **Par abus de notation, on désignera tout de même par  $f_X$  une densité de  $X$** , mais il faut être conscient que cette notation n'est définie correctement qu'à un nombre fini de points près.

TERMINOLOGIE 3.1.6 On dira par la suite qu'une propriété définie sur  $\mathbb{R}$  est vraie *presque partout*, si elle est vraie en tout point de  $\mathbb{R}$ , sauf éventuellement en un nombre **fini** de points. On abrègera par p.p. Par exemple, l'égalité  $f(x) \underset{\text{p.p.}}{=} g(x)$  signifie que  $f$  et  $g$  sont égales sur  $\mathbb{R}$ , sauf éventuellement en un nombre fini de points. (Notation à redéfinir dans un devoir)

EXEMPLES 3.1.7 1. Plusieurs densités de la variable aléatoire réelle à densité dont la fonction de répartition est donnée par :

$$\forall x \in \mathbb{R}, \quad F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1. \end{cases}$$

- 2. Les v.a.r.d. ne sont pas des variables aléatoires réelles à densité.
- 3. Une variable aléatoire réelle peut n'être ni une variable aléatoire réelle discrète ni une variable aléatoire réelle à densité. Par exemple la v.a.r. dont la fonction de répartition est donnée par :

$$\forall x \in \mathbb{R}, \quad F_X(x) = \begin{cases} 0 & \text{si } x < -1 \\ \frac{x+2}{4} & \text{si } -1 \leq x < 1 \\ 1 & \text{si } x \geq 1. \end{cases}$$

(Vérifiez que ceci est bien une fonction de répartition, notamment qu'elle est continue à droite)

4. Exemple classique (un cas particulier de loi de Pareto) :

$$\forall x \in \mathbb{R}, F_X(x) = \begin{cases} 0 & \text{si } x < 1 \\ 1 - \frac{1}{x} & \text{si } x \geq 1, \end{cases} \quad f_X(x) \stackrel{\text{p.p.}}{=} \begin{cases} 0 & \text{si } x < 1 \\ \frac{1}{x^2} & \text{si } x \geq 1, \end{cases}$$

5. Exemple d'une densité non bornée

$$\forall x \in \mathbb{R}, F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \sqrt{x} & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

6. Loi uniforme sur  $[a, b]$ ,  $X \hookrightarrow \mathcal{U}([a, b])$ , exemple à connaître

$$\forall x \in \mathbb{R}, f_X(x) \stackrel{\text{p.p.}}{=} \begin{cases} 0 & \text{si } x < a \text{ ou } x > b \\ \frac{1}{b-a} & \text{si } x \in [a, b] \end{cases}$$

7. Loi exponentielle de paramètre  $c > 0$ ,  $X \hookrightarrow \mathcal{E}(c)$ , exemple à connaître

$$\forall x \in \mathbb{R}, f_X(x) \stackrel{\text{p.p.}}{=} \begin{cases} 0 & \text{si } x < 0 \\ ce^{-cx} & \text{si } x \geq 0 \end{cases}$$

**PROPOSITION 3.1.8 (interprétation de la densité)**

Soit  $X$  une variable aléatoire réelle à densité. Pour tout  $x \in \mathbb{R}$  en lequel  $f(x) = F'(x) \neq 0$ , on a :

$$P(x < X \leq x + h) \underset{h \rightarrow 0}{\sim} f_X(x) \cdot h.$$

Ainsi, pour  $h$  petit  $f_X(x)h$  est à peu près la probabilité d'obtenir une valeur proche de  $x$  dans un diamètre de longueur  $h$ . Ainsi, plus la densité au point  $x$  est élevée, plus la probabilité d'obtenir une valeur proche de  $x$  est forte.

**PROPOSITION 3.1.9** Soit  $X$  une variable aléatoire réelle à densité. Alors :  $\forall x \in \mathbb{R}, P(X = x) = 0$ .

**THÉORÈME 3.1.10** Soit  $X$  une variable aléatoire réelle à densité, de fonction de répartition  $F_X$ , et dont une densité est  $f_X$ . Alors, pour tout  $x \in \mathbb{R}$ , l'intégrale ci-dessous converge, et on a l'égalité :

$$F_X(x) = \int_{-\infty}^x f_X(t) dt.$$

**COROLLAIRE 3.1.11** Soit  $X$  une variable aléatoire réelle à densité et  $f_X$  une densité de  $X$ . Alors  $\int_{-\infty}^{+\infty} f_X(x) dx$  converge et :  $\int_{-\infty}^{+\infty} f_X(x) dx = 1$ .

**COROLLAIRE 3.1.12** Soit  $X$  une variable aléatoire réelle à densité, dont une densité est  $f_X$ , et soit  $(a, b) \in \mathbb{R}^2$ ,  $a < b$ . Alors :

- $P(X \leq a) = P(X < a) = \int_{-\infty}^a f_X(t) dt$  ;
- $P(X \geq a) = P(X > a) = \int_a^{+\infty} f_X(t) dt$  ;
- $P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = \int_a^b f_X(t) dt$ .

**THÉORÈME 3.1.13 (Caractérisation d'une densité)**

Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$ . Les propositions suivantes sont équivalentes :

- (i)  $f$  est la densité d'une variable aléatoire réelle  $X$  ;
- (ii)  $f$  vérifie les trois propriétés suivantes :
  - $f$  est positive ;
  - $f$  est continue sur  $\mathbb{R}$  sauf éventuellement en un nombre fini de points ;
  - l'intégrale  $\int_{-\infty}^{+\infty} f(x) dx$  converge, et  $\int_{-\infty}^{+\infty} f(x) dx = 1$ .

**3.1.2 Premier théorème de transfert : densité de  $\varphi(X)$  dans des cas simples**

Avant d'étudier un théorème général, nous voyons quelques exemples. Un conseil : refaites le raisonnement à chaque fois dans ces cas simples, plutôt que d'utiliser le théorème général, avec lequel on risque plus de s'embrouiller qu'autre chose. Il est d'ailleurs stipulé dans le programme que les formules obtenues dans ces cas particuliers doivent être retrouvées.

Remarquez par ailleurs que l'exemple 2 donné ci-dessous ne rentre pas dans les hypothèses du premier théorème de transfert. Il est donc capital d'être familiarisé avec les calculs de ces exemples, pour pouvoir les adapter au moins aux cas qui sortent du cadre du théorème général.

Dans ce qui suit,  $X$  désigne une variable aléatoire réelle à densité.

EXEMPLES 3.1.14 1.  $\varphi(X) = aX + b$ . On obtient :

$$\forall x \in \mathbb{R}, \quad F_{\varphi(X)}(x) = 1 - F_X\left(\frac{x-b}{a}\right) \quad f_{\varphi(X)}(x) \underset{\text{p.p.}}{=} -\frac{1}{|a|} f_X\left(\frac{x-b}{a}\right).$$

2.  $\varphi(X) = X^2$ . On obtient :  $F_{\varphi(X)}(x) = 0$  sur  $\mathbb{R}_-^*$ , et :

$$\forall x \in \mathbb{R}^+, \quad F_{\varphi(X)}(x) = F_X(\sqrt{x}) - F_X(-\sqrt{x}) \quad f_{\varphi(X)}(x) \underset{\text{p.p.}}{=} \frac{1}{2\sqrt{x}} (f_X(\sqrt{x}) + f_X(-\sqrt{x})).$$

3.  $\varphi(X) = e^X$ . On obtient :  $F_{\varphi(X)}(x) = 0$  sur  $\mathbb{R}^-$  et :

$$\forall x \in \mathbb{R}_+^*, \quad F_{\varphi(X)}(x) = F_X(\ln x) \quad f_{\varphi(X)}(x) \underset{\text{p.p.}}{=} \frac{1}{x} f_X(\ln x).$$

**THÉORÈME 3.1.15 (Premier théorème de transfert)**

Soit  $X$  une variable aléatoire réelle à densité de densité  $f$ , et soit  $\varphi$  une application de classe  $\mathcal{C}^1$  telle que  $\varphi'$  soit strictement positive, sauf éventuellement en un nombre **fini** de points. Alors :

- (i)  $\varphi$  se corestreint en une bijection de  $\mathbb{R}$  sur  $\varphi(\mathbb{R})$  ; on note  $\varphi^{-1} : \varphi(\mathbb{R}) \rightarrow \mathbb{R}$  sa réciproque ;
- (ii)  $\varphi(X)$  est une variable aléatoire réelle à densité, dont toute densité  $g$  est nulle p.p. en dehors de  $\varphi(\mathbb{R})$ , et vérifie :

$$\forall y \in \varphi(\mathbb{R}), \quad g(y) \underset{\text{p.p.}}{=} (f \circ \varphi^{-1})(y) (\varphi^{-1})'(y).$$

REMARQUES 3.1.16 1. L'égalité définissant  $g(y)$  sur  $\varphi(\mathbb{R})$  n'est pas définie en tout point en lequel s'annule  $\varphi'$ . Mais ces points sont en nombre fini.

- 2. Si  $\varphi$  est strictement décroissante ( $\varphi'$  strictement négative sauf en un nombre fini de points), alors on obtient de même :

$$\forall y \in \varphi(\mathbb{R}), \quad g(y) \underset{\text{p.p.}}{=} -(f \circ \varphi^{-1})(y) (\varphi^{-1})'(y).$$

3. Dans ce théorème  $\varphi$  peut n'être défini que sur un intervalle ouvert contenant  $X(\Omega)$ .

EXEMPLES 3.1.17 Retrouver les formules pour les densités de  $aX + b$ , de  $e^X$ .

## 3.2 Moments d'une variable aléatoire réelle à densité

### 3.2.1 Espérance

LEMME 3.2.1 Soit  $f$  une fonction positive, continue sur  $\mathbb{R}$ , sauf en un nombre fini de points, telle que  $\int_{-\infty}^{+\infty} f(t) dt$  converge. Alors les seuls défauts de convergence de l'intégrale  $\int_{-\infty}^{+\infty} tf(t) dt$  sont aux bornes  $-\infty$  et  $+\infty$ .

DÉFINITION 3.2.2 Soit  $X$  une variable aléatoire réelle à densité de densité  $f_X$ . On dit que  $X$  admet une espérance si et seulement si l'intégrale  $\int_{-\infty}^{+\infty} tf_X(t) dt$  converge, et dans ce cas, on définit l'espérance (mathématique) de  $X$  par :

$$E(X) = \int_{-\infty}^{+\infty} tf_X(t) dt.$$

REMARQUE 3.2.3 Contrairement aux v.a.r.d., il est inutile ici d'avoir une hypothèse de convergence absolue. D'ailleurs, on peut remarquer que, puisque  $t \mapsto tf_X(t)$  est de signe constant sur  $\mathbb{R}_+^*$  et sur  $\mathbb{R}_-^*$ , la convergence de  $\int_{-\infty}^{+\infty} tf_X(t) dt$  équivaut en fait à sa convergence absolue.

EXEMPLES 3.2.4 1.  $X \hookrightarrow \mathcal{U}([a, b])$ , alors  $E(X) = \frac{a+b}{2}$  (à connaître)

2.  $X \hookrightarrow \mathcal{E}(c)$ , alors  $E(X) = \frac{1}{c}$  (à connaître)

3.  $X$  défini par la densité  $t \mapsto f(t) = \frac{1}{\pi} \cdot \frac{1}{1+t^2}$  (Loi de Cauchy)

La variable aléatoire réelle à densité  $X$  n'admet pas d'espérance.

4. Une variable aléatoire réelle à densité de densité non bornée peut admettre une espérance.

Exemple  $X$  de densité  $f$  :

$$\forall x \in \mathbb{R}, f(x) = \begin{cases} 0 & \text{si } x \leq 0 \text{ ou si } x > 1 \\ \frac{1}{2\sqrt{x}} & \text{si } 0 < x \leq 1. \end{cases}$$

On obtient  $E(X) = \frac{1}{3}$ .

PROPOSITION 3.2.5 (**Existence de l'espérance d'une v.a.r. à densité bornée p.s.**)

Si  $X(\Omega)$  est borné presque sûrement, i.e. s'il existe un intervalle borné  $I$  tel que  $P(X \in I) = 1$ , alors  $E(X)$  existe.

Les hypothèses ci-dessus sont équivalentes à dire qu'une densité  $f_X$  de  $X$  est à support borné. Remarquez que si cette propriété est vraie pour une densité, elle l'est pour toutes les densités.

**PROPOSITION 3.2.6 (Positivité de l'espérance)**

Soit  $X$  une variable aléatoire réelle à densité admettant une espérance. Si  $X \geq 0$  ps, c'est-à-dire si  $P(X < 0) = 0$ , alors  $E(X) \geq 0$ .

**PROPOSITION 3.2.7** Soit  $X$  une variable aléatoire réelle à densité de densité  $f$ , et admettant une espérance. Si  $f$  est pair (densité symétrique), alors  $E(X) = 0$ .

**THÉORÈME 3.2.8 (Second théorème de transfert, admis en général)**

Soit  $X$  une variable aléatoire réelle à densité, et  $f$  une densité de  $X$ . Soit  $I = ]a, b[$  un intervalle ouvert contenant  $X(\Omega)$ , et  $\varphi$  une application de  $I$  dans  $\mathbb{R}$ , continue presque partout. Alors :

(i)  $Y = \varphi(X)$  est une variable aléatoire réelle.

(ii)  $Y$  admet une espérance mathématique si et seulement si  $\int_a^b \varphi(t)f(t) dt$  converge **absolument**

(iii) dans ce cas,  $E(X) = \int_a^b \varphi(t)f(t) dt$ .

Ce théorème est admis en général. Conformément au programme, nous nous contentons de le démontrer dans le cas particulier où  $\varphi$  est de classe  $\mathcal{C}^1$ , et  $\varphi'$  strictement positive.

**REMARQUE 3.2.9** L'énoncé général du second de transfert dépasse largement le cadre du programme, puisque  $\varphi(X)$  n'est avec ces hypothèses pas nécessairement une variable aléatoire réelle à densité (et donc nous n'avons pas défini la notion d'espérance pour cette situation). C'est pour cette raison qu'ici **l'hypothèse de convergence absolue est indispensable**.

**EXEMPLE 3.2.10**  $X \hookrightarrow \mathcal{U}([0, 1])$ , et  $\varphi$  définie sur  $\mathbb{R}$  par :

$$\forall x \in \mathbb{R}, \quad \varphi(x) = \begin{cases} 0 & \text{si } x < \frac{1}{2} \\ 2x - 1 & \text{si } \frac{1}{2} \leq x \leq 1 \\ 1 & \text{si } x > 1. \end{cases}$$

En revanche, d'après le premier théorème de transfert, si  $\varphi$  est de classe  $\mathcal{C}^1$ , et  $\varphi'$  strictement positive, alors  $\varphi(X)$  est une variable aléatoire réelle à densité.

**EXEMPLES 3.2.11** 1. Soit  $X \hookrightarrow \mathcal{E}(1)$ , alors  $e^X$  n'admet pas d'espérance.

2. Soit  $X \hookrightarrow \mathcal{E}(1)$ , alors  $X^2$  admet une espérance, et  $E(X^2) = \Gamma(3) = 2$ .

Ce dernier cas entre dans les hypothèses du théorème, mais pas dans le cadre de la démonstration restreinte effectuée : pour vérification de la pertinence du théorème, on fait aussi le calcul à la main, pour vérifier qu'on obtient bien la même chose

3. Soit  $X$  une variable aléatoire réelle à densité de densité  $f$  définie par :

$$\forall x \in \mathbb{R}, \quad f(x) = \begin{cases} 0 & \text{si } x \leq 1 \\ \frac{2}{x^3} & \text{si } x > 1. \end{cases}$$

Soit  $Y = X^2 \sin X$ . Alors  $Y$  n'admet pas de variance.

**COROLLAIRE 3.2.12** Soit  $X$  une variable aléatoire réelle à densité, et  $a$  et  $b$  deux réels. Si  $X$  admet une espérance, alors  $aX + b$  aussi, et  $E(aX + b) = aE(X) + b$ .

### 3.2.2 Moments d'ordre 2

DÉFINITION 3.2.13 Soit  $X$  une variable aléatoire réelle à densité, soit  $f_X$  une densité de  $X$ . On dit que  $X$  admet un moment d'ordre 2 si l'intégrale  $\int_{-\infty}^{+\infty} t^2 f_X(t) dt$  converge, et le cas échéant, on définit le moment d'ordre 2 par :

$$M_2(X) = \int_{-\infty}^{+\infty} t^2 f_X(t) dt$$

PROPOSITION 3.2.14 Si  $X$  admet un moment d'ordre 2, alors  $X$  admet une espérance.

PROPOSITION 3.2.15 Soit  $X$  une variable aléatoire réelle à densité. La variable aléatoire  $X$  admet un moment d'ordre 2 si et seulement si  $X^2$  admet une espérance, et dans ce cas,  $M_2(X) = E(X^2)$ .

DÉFINITION 3.2.16 On dit que  $X$  admet une variance (ou un moment centré d'ordre 2) si  $X$  admet une espérance, et si  $(X - E(X))^2$  admet une espérance. Dans ce cas,

$$V(X) = E((X - E(X))^2) = \int_{-\infty}^{+\infty} (t - E(X))^2 f_X(t) dt.$$

REMARQUE 3.2.17 Soit  $X$  une variable aléatoire réelle à densité, admettant une variance. Alors  $V(X) \geq 0$

PROPOSITION 3.2.18 Soit  $X$  une variable aléatoire réelle à densité. Les propositions suivantes sont équivalentes :

- (i)  $X$  admet un moment d'ordre 2,
- (ii)  $X$  admet une variance.

DÉFINITION 3.2.19 Soit  $X$  une variable aléatoire réelle à densité admettant une variance. Alors on définit l'écart-type de  $X$  par  $\sigma(X) = \sqrt{V(X)}$ .

#### THÉORÈME 3.2.20 (Formule de König-Huygens)

Si  $X$  est une variable aléatoire réelle à densité admettant un moment d'ordre 2, alors

$$V(X) = M_2(X) - E(X)^2.$$

THÉORÈME 3.2.21 Soit  $X$  une variable aléatoire réelle à densité admettant une variance, et  $a$  et  $b$  deux réels. Alors  $aX + b$  admet aussi une variance, et

$$V(aX + b) = a^2 V(X).$$

EXEMPLES 3.2.22 1. Soit  $X \hookrightarrow \mathcal{U}([a, b])$ , alors  $V(X) = \frac{(b-a)^2}{12}$

2. Soit  $X \hookrightarrow \mathcal{E}(c)$ , alors  $V(X) = \frac{1}{c^2}$ .

3. Soit  $X$  de densité  $f$  définie par :

$$\forall x \in \mathbb{R}, f(x) = \begin{cases} 0 & \text{si } x \leq 1 \\ \frac{2}{x^3} & \text{si } x > 1. \end{cases}$$

Alors  $X$  admet une espérance mais pas une variance.

DÉFINITION 3.2.23 Soit  $X$  une variable aléatoire réelle à densité.

1. On dit que  $X$  est centrée si  $X$  admet une espérance et  $E(X) = 0$  ;
2. On dit que  $X$  est centrée réduite si  $X$  admet un moment d'ordre 2, et  $E(X) = 0$ ,  $\sigma(X) = 1$ .

PROPOSITION 3.2.24 Soit  $X$  une variable aléatoire réelle à densité admettant un moment d'ordre 2. Alors  $X - E(X)$  est centrée, et  $\frac{X - E(X)}{\sigma(X)}$  est centrée réduite.

DÉFINITION 3.2.25 La variable  $X^* = \frac{X - E(X)}{\sigma(X)}$  est appelée variable centrée réduite associée à  $X$ .

### 3.3 Somme de deux variables aléatoires à densité

#### 3.3.1 Densité d'une somme

DÉFINITION 3.3.1 Soit  $X$  deux v.a.r. (à densité ou non). On dit que  $X$  et  $Y$  sont indépendantes si pour tout  $t \in \mathbb{R}$ , et tout  $t' \in \mathbb{R}$ , on a :

$$P(X \leq t, Y \leq t') = P(X \leq t)P(Y \leq t') = F_X(t)F_Y(t').$$

DÉFINITION 3.3.2 Soit  $f$  et  $g$  deux application de  $\mathbb{R}$  dans  $\mathbb{R}$ . On appelle produit de convolution, et on note  $f \star g$ , la fonction définie pour tout  $x$  en lequel on a convergence par

$$f \star g(x) = \int_{-\infty}^{+\infty} f(t)g(x-t) dt.$$

THÉORÈME 3.3.3 (Densité d'une somme, admis)

Soit  $X$  et  $Y$  deux variables aléatoires réelles à densité **indépendantes**. Si le produit de convolution  $f_X \star f_Y$  est continu presque partout, alors la variable  $X + Y$  est une variable aléatoire réelle à densité, dont une densité est donnée par :

$$\forall x \in \mathbb{R}, \quad f_{X+Y}(x) \underset{p.p.}{=} f_X \star f_Y(x) \underset{p.p.}{=} \int_{-\infty}^{+\infty} f_X(t)f_Y(x-t) dt.$$

EXEMPLES 3.3.4 1.  $X, Y \hookrightarrow \mathcal{E}(1)$ , indépendantes. Densité de  $X + Y$  ?

2.  $X, Y \hookrightarrow \mathcal{U}([0, 1])$ , indépendantes. Densité de  $X + Y$  ?

#### 3.3.2 Espérance d'une somme

On a vu que le second théorème de transfert, que l'on a admis, permet d'obtenir des espérances de variables aléatoires qui ne sont pas nécessairement des v.a.r.d. ou des variables aléatoires réelles à densité. Le programme fournit également un théorème permettant d'exploiter ces espérances, dont l'énoncé (et donc la démonstration) utilise des objets que nous n'avons pas forcément bien définis.

**THÉORÈME 3.3.5 (Espérance d'une somme – démonstration hors-programme)**

Soit  $X$  et  $Y$  deux v.a.r. quelconques admettant une espérance (obtenues par exemple à l'aide du théorème de transfert). Alors  $X + Y$  admet une espérance et

$$E(X + Y) = E(X) + E(Y).$$

**COROLLAIRE 3.3.6** Si  $X \leq Y$  presque sûrement, alors  $E(X) \leq E(Y)$ .

**3.3.3 Variance d'une somme**

De même que pour les espérances, grâce au théorème de transfert, on peut être amené à calculer la variance de variables qui ne sont pas des variables aléatoires réelles à densité. Le théorème suivant fournit des règles de calcul dans ce cadre sortant des définitions du programme.

**THÉORÈME 3.3.7 (Variance d'une somme – démonstration hors-programme)**

Soit  $X$  et  $Y$  deux variables aléatoires réelles quelconques, admettant des variances.

1. Si  $X$  et  $Y$  sont indépendantes, alors  $E(XY) = E(X)E(Y)$
2. Si  $X$  et  $Y$  sont indépendantes, alors  $V(X + Y) = V(X) + V(Y)$ .

**REMARQUE 3.3.8** On pourrait définir pour les variables aléatoires réelles à densité, exactement de la même façon que pour les v.a.r.d., la notion de covariance. On obtiendrait alors de même une formule donnant  $V(X + Y)$  en fonction de  $V(X)$ ,  $V(Y)$  et  $\text{cov}(X, Y)$ . La notion de covariance (et donc la formule générale de la variance de la somme) pour les variables aléatoires réelles à densité est hors-programme. Elle est parfois introduite dans certains problèmes.



# Probabilités – Chapitre 4

## Loi continues classiques

### 4.1 Loi uniforme

#### 4.1.1 Description et transformations simples

PROPOSITION/DÉFINITION 4.1.1 Soit  $a < b$ , et  $f$  la fonction définie sur  $\mathbb{R}$  par :

$$\forall x \in \mathbb{R}, f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b], \\ 0 & \text{sinon.} \end{cases}$$

Alors  $f$  est une densité de probabilité. Si  $X$  est une variable aléatoire dont  $f$  est une densité, on dit que  $X$  suit une loi uniforme sur  $[a, b]$ , et on note  $X \hookrightarrow \mathcal{U}([a, b])$ .

PROPOSITION 4.1.2 Soit  $X \hookrightarrow \mathcal{U}([a, b])$ . Alors la fonction de répartition de  $X$  est :

$$\forall t \in \mathbb{R}, F_X(t) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } x \in [a, b] \\ 1 & \text{si } x > b. \end{cases}$$

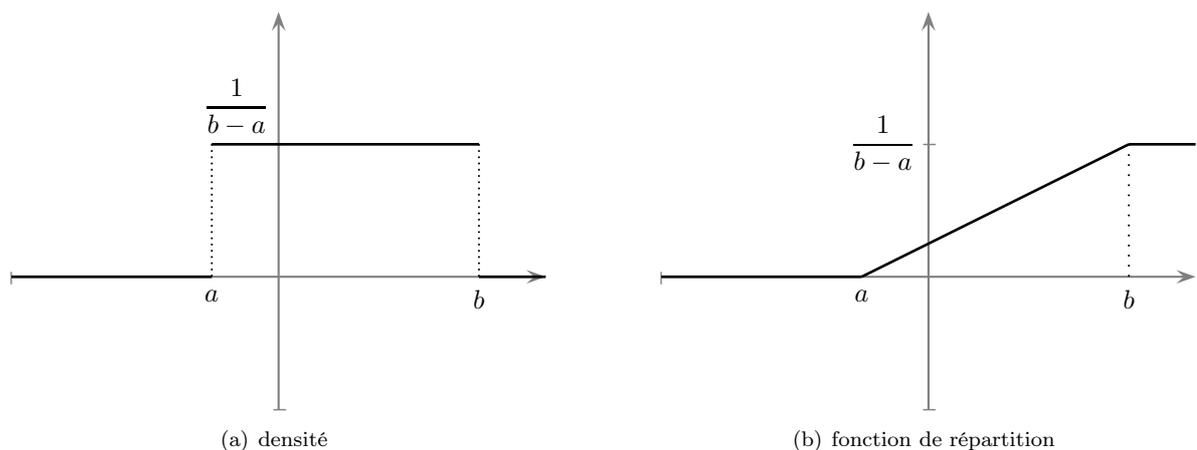


FIGURE 4.1 – Loi uniforme sur  $[a, b]$

THÉORÈME 4.1.3 Soit  $a < b$ , et  $(\alpha, \beta) \in \mathbb{R}^2, \alpha \neq 0$ . Alors :

- (i) si  $\alpha > 0$ ,  $(X \hookrightarrow \mathcal{U}([a, b])) \iff (\alpha X + \beta \hookrightarrow \mathcal{U}([\alpha a + \beta, \alpha b + \beta]))$ .  
(ii) si  $\alpha < 0$ ,  $(X \hookrightarrow \mathcal{U}([a, b])) \iff (\alpha X + \beta \hookrightarrow \mathcal{U}([\alpha b + \beta, \alpha a + \beta]))$ .

COROLLAIRE 4.1.4 Soit  $a < b$ . Alors :

- (i)  $(X \hookrightarrow \mathcal{U}([0, 1])) \iff (a + (b - a)X \hookrightarrow \mathcal{U}([a, b]))$ .  
(ii)  $(X \hookrightarrow \mathcal{U}([a, b])) \iff (\frac{X}{b-a} - \frac{a}{b-a} \hookrightarrow \mathcal{U}([0, 1]))$ .  
(iii) si  $b > 0$ ,  $(X \hookrightarrow \mathcal{U}([0, 1])) \iff (a + bX \hookrightarrow \mathcal{U}([a, a + b]))$ .  
(iv) si  $b < 0$ ,  $(X \hookrightarrow \mathcal{U}([0, 1])) \iff (a + bX \hookrightarrow \mathcal{U}([a + b, a]))$ .

### 4.1.2 Moments

THÉORÈME 4.1.5 Soit  $X \hookrightarrow \mathcal{U}([0, 1])$ . Alors  $X$  admet une espérance et une variance, et :

$$E(X) = \frac{1}{2} \quad \text{et} \quad V(X) = \frac{1}{12}.$$

COROLLAIRE 4.1.6 Soit  $X \hookrightarrow \mathcal{U}([a, b])$ . Alors  $X$  admet une espérance et une variance, et

$$E(X) = \frac{a+b}{2} \quad \text{et} \quad V(X) = \frac{(b-a)^2}{12}.$$

## 4.2 Loi exponentielle

### 4.2.1 Description et transformations simples

PROPOSITION/DÉFINITION 4.2.1 Soit  $c \in \mathbb{R}_+^*$ , et  $f$  la fonction définie sur  $\mathbb{R}$  par :

$$\forall x \in \mathbb{R}, \quad f(x) = \begin{cases} ce^{-cx} & \text{si } x \geq 0, \\ 0 & \text{si } x < 0 \end{cases}.$$

Alors  $f$  est une densité de probabilité. Si  $X$  est une variable aléatoire dont  $f$  est une densité, on dit que  $X$  suit une loi exponentielle de paramètre  $c$ , et on note  $X \hookrightarrow \mathcal{E}(c)$ .

PROPOSITION 4.2.2 Soit  $X \hookrightarrow \mathcal{E}(c)$ . Alors, la fonction de répartition de  $X$  est :

$$\forall x \in \mathbb{R}, \quad F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-cx} & \text{si } x \geq 0. \end{cases}$$

PROPOSITION 4.2.3 Soit  $c \in \mathbb{R}_+^*$ , et  $\lambda \in \mathbb{R}_+^*$ . Alors :  $X \hookrightarrow \mathcal{E}(c) \iff \frac{X}{\lambda} \hookrightarrow \mathcal{E}(\lambda c)$ .

COROLLAIRE 4.2.4 Soit  $\lambda \in \mathbb{R}_+^*$ . Alors :

- (i)  $X \hookrightarrow \mathcal{E}(1) \iff \frac{X}{\lambda} \hookrightarrow \mathcal{E}(\lambda)$   
(ii)  $X \hookrightarrow \mathcal{E}(\lambda) \iff \lambda X \hookrightarrow \mathcal{E}(1)$

Plus  $c$  est grand, plus la variable est concentrée au voisinage à droite de 0. Nous illustrons ceci sur le représentation de la densité correspondant à deux paramètres  $c$  et  $c'$ , avec  $c < c'$  (figure (4.2)). Cette propriété se traduit aussi par la valeur de l'espérance et de la variance (voir paragraphe suivant)

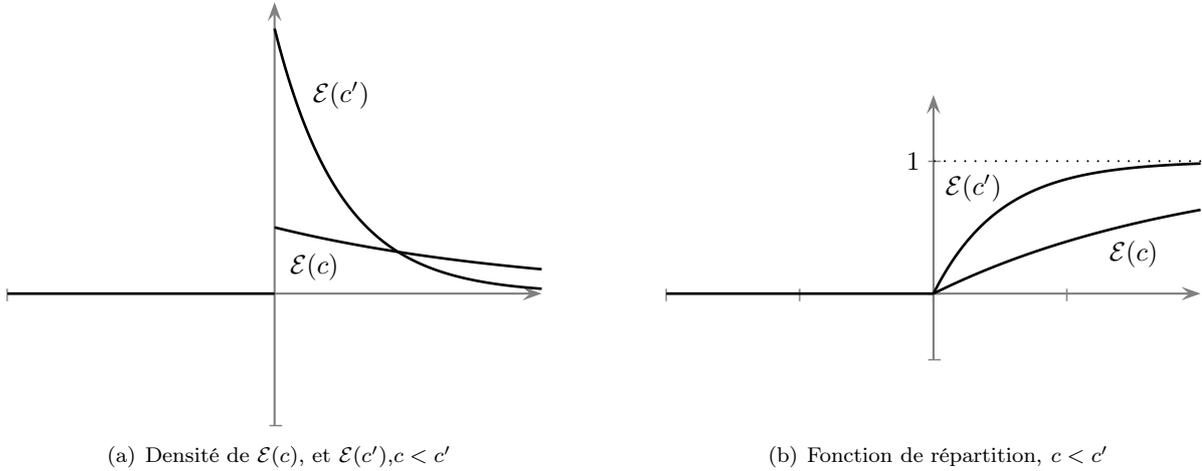


FIGURE 4.2 – Loi exponentielle

### 4.2.2 Moments

THÉORÈME 4.2.5 Soit  $X \hookrightarrow \mathcal{E}(1)$ . Alors  $X$  admet une espérance et une variance, et

$$E(X) = 1 \quad \text{et} \quad V(X) = 1.$$

COROLLAIRE 4.2.6 Soit  $c \in \mathbb{R}_+^*$ , et  $X \hookrightarrow \mathcal{E}(c)$ . Alors  $X$  admet une espérance et une variance, et :

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

REMARQUE 4.2.7 Analogie avec la loi géométrique.

### 4.2.3 Caractérisation par l'absence de mémoire

DÉFINITION 4.2.8 Soit  $X$  une variable aléatoire réelle. On dit que  $X$  suit une loi sans mémoire si  $X$  est positive ou nulle, et si :

$$\forall (x, y) \in \mathbb{R}_+, \quad P(X > x + y) = P(X > x)P(X > y).$$

PROPOSITION 4.2.9 Soit  $X$  suivant une loi sans mémoire, et  $(x, y, z) \in \mathbb{R}_+^*$ , tels que  $[X > x]$  et  $[X > y]$ , soient non quasi-impossibles. Alors :

- $P(X > x + z \mid X > x) = P(X > z)$
- $P(X > x + z \mid X > x) = P(X > y + z \mid X > y)$ .

THÉORÈME 4.2.10 Soit  $X$  une variable aléatoire réelle. Alors  $X$  suit une loi sans mémoire si et seulement si  $X$  est nulle presque sûrement, ou  $X$  suit une loi exponentielle.

### 4.3 Lois Gamma ( $\Gamma$ ) et gamma ( $\gamma$ )

#### 4.3.1 Description et transformations simples

PROPOSITION 4.3.1 Soit  $(b, \nu) \in (\mathbb{R}_+^*)^2$ , et  $f$  la fonction définie sur  $\mathbb{R}$  par :

$$\forall x \in \mathbb{R}, \quad f(x) = \begin{cases} \frac{x^{\nu-1}}{\Gamma(\nu)b^\nu} e^{-\frac{x}{b}} = \frac{1}{b\Gamma(\nu)} \cdot \left(\frac{x}{b}\right)^{\nu-1} e^{-\frac{x}{b}} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0 \end{cases}.$$

Alors  $f$  est une densité de probabilité.

DÉFINITION 4.3.2 1. On dit qu'une variable aléatoire suit une loi  $\Gamma$  de paramètres  $(b, \nu)$  si  $X$  admet la fonction  $f$  de la proposition précédente pour densité. On note  $X \hookrightarrow \Gamma(b, \nu)$ . Le réel  $b$  est appelé paramètre d'échelle.

2. On dit que  $X$  suit la loi  $\gamma$  de paramètre  $\nu$  si  $X$  suit la loi  $\Gamma(1, \nu)$ . On note  $X \hookrightarrow \gamma(\nu)$ . On dit aussi que  $X$  suit la loi gamma-standard de paramètre  $\nu$ .

Ainsi, si  $X \rightarrow \gamma(\nu)$ , une densité de  $X$  est donnée par :

$$\forall x \in \mathbb{R}, \quad f(x) = \begin{cases} \frac{x^{\nu-1}}{\Gamma(\nu)} e^{-x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0 \end{cases}$$

PROPOSITION 4.3.3 Soit  $(b, b', \nu) \in (\mathbb{R}_+^*)^3$ .

1. Soit  $X \hookrightarrow \gamma(\nu)$ ,  $Y \hookrightarrow \Gamma(b, \nu)$ . Alors :

$$\forall x \in \mathbb{R}, \quad f_Y(x) \underset{p.p.}{=} \frac{1}{b} f_X\left(\frac{x}{b}\right).$$

2. Plus généralement, soit  $X \hookrightarrow \Gamma(b, \nu)$ ,  $Y \hookrightarrow \Gamma(b', \nu)$ . Alors :

$$\forall x \in \mathbb{R}, \quad f_Y(x) \underset{p.p.}{=} \frac{b}{b'} f_X\left(\frac{xb}{b'}\right).$$

PROPOSITION 4.3.4 Soit  $(b, b', \nu) \in (\mathbb{R}_+^*)^3$ .

1.  $X \hookrightarrow \gamma(\nu) \iff bX \hookrightarrow \Gamma(b, \nu)$ .

2.  $X \hookrightarrow \Gamma(b, \nu) \iff \frac{X}{b} \hookrightarrow \gamma(\nu)$ .

3. Plus généralement :  $X \hookrightarrow \Gamma(b, \nu) \iff b'X \hookrightarrow \Gamma(bb', \nu)$ .

PROPOSITION 4.3.5  $X$  suit la loi  $\Gamma(b, 1)$  si et seulement si  $X$  suit la loi  $\mathcal{E}(\frac{1}{b})$ .

#### 4.3.2 Influence des paramètres sur la densité

Soit  $X \hookrightarrow \Gamma(b, \nu)$ , et  $f$  la densité de  $X$  décrite plus haut. Alors :

$$f(x) \underset{0+}{\sim} \frac{x^{\nu-1}}{b^\nu \Gamma(\nu)}$$

PROPOSITION 4.3.6 1.  $\lim_{x \rightarrow +\infty} f(x) = 0$

2. • si  $\nu \in ]0, 1[$ ,  $\lim_{x \rightarrow 0^+} f(x) = +\infty$  et  $f$  est décroissante sur  $\mathbb{R}_+^*$  ;
- si  $\nu = 1$ ,  $\lim_{x \rightarrow 0^+} f(x) = \frac{1}{b}$  et  $f$  est décroissante sur  $\mathbb{R}_+^*$  ;
- si  $\nu \in ]1, 2[$  :
  - \*  $\lim_{x \rightarrow 0^+} f(x) = 0$ ,
  - \* la courbe de  $f$  admet en  $0^+$  une tangente verticale ;
- si  $\nu = 2$  :
  - \*  $\lim_{x \rightarrow 0^+} f(x) = 0$ ,
  - \* la courbe de  $f$  admet en  $0^+$  une tangente d'équation  $y = \frac{x}{b^2}$  ;
- si  $\nu > 2$  :
  - \*  $\lim_{x \rightarrow 0^+} f(x) = 0$ ,
  - \* la courbe de  $f$  admet en  $0^+$  une tangente horizontale.

Nous illustrons ce résultat dans la figure (4.3).

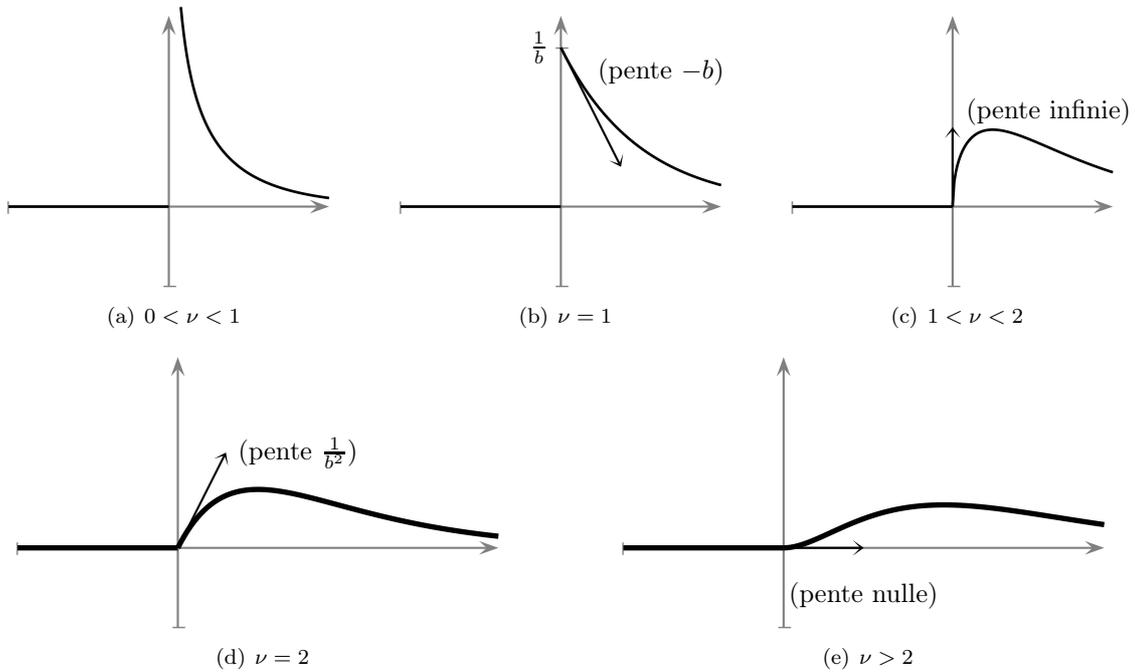


FIGURE 4.3 – Densité de la loi  $\Gamma(b, \nu)$

PROPOSITION 4.3.7 En notant  $F_{b,\nu}$  la fonction de répartition d'une variable aléatoire suivant la loi  $\Gamma(b, \nu)$ ,

$$\forall x \in \mathbb{R}, \quad F_{b,\nu}(x) = F_{1,\nu}\left(\frac{x}{b}\right).$$

### 4.3.3 Moments

PROPOSITION 4.3.8 Soit  $X \hookrightarrow \gamma(\nu)$ . Alors  $X$  admet une espérance et une variance, et :

$$E(X) = \nu \quad \text{et} \quad V(X) = \nu.$$

THÉORÈME 4.3.9 Soit  $X \hookrightarrow \Gamma(b, \nu)$ . Alors  $X$  admet une espérance et une variance, et :

$$E(X) = b\nu \quad \text{et} \quad V(X) = b^2\nu.$$

#### 4.3.4 Stabilité de la loi Gamma

THÉORÈME 4.3.10 Soit  $(b, \nu, \nu') \in (\mathbb{R}_+^*)^3$ . Soit  $X \hookrightarrow \Gamma(b, \nu)$  et  $Y \hookrightarrow \Gamma(b, \nu')$ , telles que  $X$  et  $Y$  soient indépendantes. Alors  $X + Y \hookrightarrow \Gamma(b, \nu + \nu')$ .

COROLLAIRE 4.3.11 1. Soit  $X_1, \dots, X_n$  des variables mutuellement indépendantes, telles que pour tout  $i \in \llbracket 1, n \rrbracket$ ,  $X_i \hookrightarrow \Gamma(b, \nu_i)$ . Alors :

$$X_1 + \dots + X_n \hookrightarrow \Gamma(b, \nu_1 + \dots + \nu_n).$$

2. Soit  $X_1, \dots, X_n$  mutuellement indépendantes suivant toutes la même loi  $\mathcal{E}(\lambda)$ . Alors :

$$X_1 + \dots + X_n \hookrightarrow \Gamma\left(\frac{1}{\lambda}, n\right).$$

3. Soit  $X_1, \dots, X_n$  des variables mutuellement indépendantes, telles que pour tout  $i \in \llbracket 1, n \rrbracket$ ,  $X_i \hookrightarrow \gamma(\nu_i)$ . Alors :

$$X_1 + \dots + X_n \hookrightarrow \gamma(\nu_1 + \dots + \nu_n).$$

#### 4.3.5 Cas particuliers de la loi $\Gamma$

- $\Gamma(1, \nu) = \gamma(\nu)$
- $\Gamma(b, 1) = \mathcal{E}\left(\frac{1}{b}\right)$
- $\Gamma(b, n)$ ,  $n \in \mathbb{N}^*$  : loi d'Erlang de paramètres  $(b, n)$ , de densité :

$$\forall x \in \mathbb{R}, \quad f(x) = \begin{cases} \frac{x^{n-1}}{(n-1)! \cdot b^n} e^{-\frac{x}{b}} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0 \end{cases}.$$

- $\Gamma(2, \frac{r}{2})$ ,  $r \in \mathbb{N}^*$  : loi du chi-deux à  $r$  degrés de liberté.

### 4.4 Loi normale (ou loi gaussienne, ou loi de Laplace-Gauss)

#### 4.4.1 Description et transformations simples

PROPOSITION/DÉFINITION 4.4.1 Soit  $(m, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$ . Soit  $f_{m, \sigma}$  la fonction définie par :

$$\forall x \in \mathbb{R}, \quad f_{m, \sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Alors,  $f_{m, \sigma}$  est une densité de probabilité. Une variable aléatoire admettant  $f$  pour densité est dite suivre une loi normale (ou loi de Laplace-Gauss, ou loi gaussienne) de paramètres  $(m, \sigma^2)$  (ou  $(m, \sigma)$ ). On note  $X \hookrightarrow \mathcal{N}(m, \sigma^2)$  ou  $X \hookrightarrow \mathcal{N}(m, \sigma)$ .

Étant donné l'incertitude sur la notation (le programme stipule plutôt  $\mathcal{N}(m, \sigma^2)$ , qui est plutôt contraire à l'usage), la loi normale est souvent plutôt définie par son espérance et sa variance. On dira souvent que  $X$  suit une loi normale d'espérance  $m$  et de variance  $\sigma^2$  (ainsi qu'on le verra plus tard).

PROPOSITION 4.4.2 1. Soit  $(a, b) \in \mathbb{R}^* \times \mathbb{R}$ , alors

$$X \hookrightarrow \mathcal{N}(m, \sigma^2) \iff aX + b \hookrightarrow \mathcal{N}(am + b, a^2\sigma^2)$$

$$2. X \hookrightarrow \mathcal{N}(m, \sigma) \iff \frac{X-m}{\sigma} \hookrightarrow \mathcal{N}(0, 1).$$

$$3. X \hookrightarrow \mathcal{N}(0, 1) \iff \sigma X + m \hookrightarrow \mathcal{N}(m, \sigma^2).$$

DÉFINITION 4.4.3 La loi  $\mathcal{N}(0, 1)$  est appelée loi normale centrée réduite. Toute variable aléatoire suivant une loi normale peut donc se ramener par une transformation affine à une loi normale centrée réduite. Comme on le verra plus loin,  $m$  et  $\sigma$  sont l'espérance et l'écart-type de  $X$ , donc cette transformation affine consiste simplement à associer à  $X$  la variable  $X^*$ , variable centrée réduite associée à  $X$ .

#### 4.4.2 Étude de la densité et de la fonction de répartition de la loi normale centrée réduite

Densité d'une variable aléatoire suivant une loi normale centrée réduite :

$$\forall x \in \mathbb{R}, f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

PROPOSITION 4.4.4 La fonction  $f_{0,1}$  est paire, croissante sur  $\mathbb{R}_-$ , décroissante sur  $\mathbb{R}_+$ . La courbe de  $f_{0,1}$  admet une tangente horizontale en 0, est convexe sur  $]-\infty, -1[$  et sur  $]1, +\infty[$ , et concave sur  $[-1, 1]$ .

NOTATION 4.4.5 On note  $\Phi$  la fonction de répartition d'une variable aléatoire suivant une loi normale centrée réduite. Ainsi :

$$\forall x \in \mathbb{R}, \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

PROPOSITION 4.4.6 1.  $\Phi(0) = \frac{1}{2}$

2.  $\Phi - \frac{1}{2}$  est impaire, donc :

$$\forall x \in \mathbb{R}, \Phi(x) = 1 - \Phi(-x).$$

PROPOSITION 4.4.7 La courbe de  $\Phi$  est convexe sur  $\mathbb{R}_-$  et concave sur  $\mathbb{R}_+$ . Elle admet en 0 un point d'inflexion de pente  $\Phi'(0) = \frac{1}{\sqrt{2\pi}}$ . Ce point d'inflexion est aussi centre de symétrie de la courbe.

La densité et la fonction de répartition de la loi normale centrée réduite sont représentées dans la figure (4.4)

#### 4.4.3 Influence des paramètres

PROPOSITION 4.4.8 Soit  $(m, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$ . Alors :

$$\forall x \in \mathbb{R}, f_{m,\sigma}(x) = \frac{1}{\sigma} f_{0,1}\left(\frac{x-m}{\sigma}\right).$$

- Le terme  $\frac{1}{\sigma}$  à l'extérieur de  $f$  est un paramètre d'affinité verticale de rapport  $\frac{1}{\sigma}$

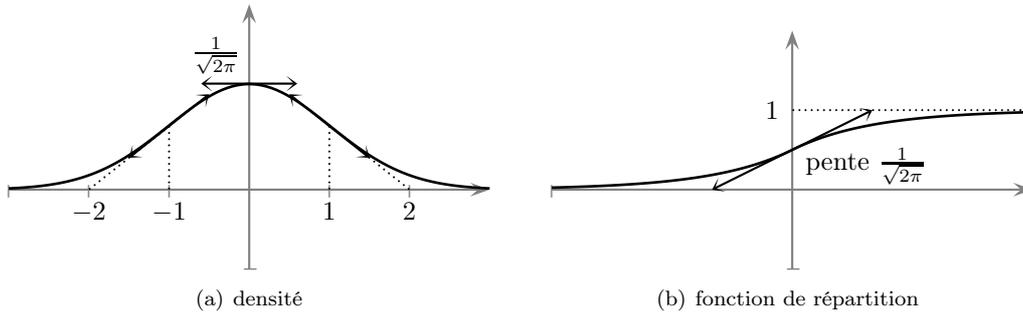
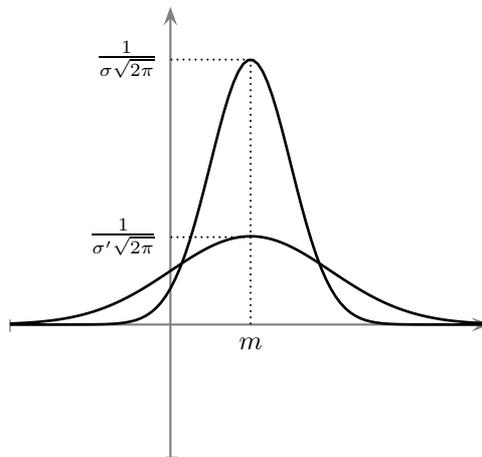


FIGURE 4.4 – Loi normale centrée réduite (échelles non respectées)

FIGURE 4.5 – Densité de  $\mathcal{N}(m, \sigma^2)$  et  $\mathcal{N}(m, (\sigma')^2)$ ,  $\sigma < \sigma'$ 

- Le terme  $\frac{1}{\sigma}$  à l'intérieur de  $f$  correspond à une affinité horizontale de rapport  $\sigma$  (compensant l'affinité verticale).
  - Le terme  $x - m$  correspond à une translation horizontale de  $m$ .
- Ainsi, la courbe est centrée en  $m$ , et plus  $\sigma$  est petit, plus la courbe de  $f_{m,\sigma}$  est concentrée autour de  $m$  (et moins elle est étalée).

NOTATION 4.4.9 On note  $\Phi_{m,\sigma}$  la fonction de répartition d'une variable aléatoire suivant une loi  $\mathcal{N}(m, \sigma^2)$ . Ainsi :

$$\forall x \in \mathbb{R}, \quad \Phi_{m,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^x e^{-\frac{(t-m)^2}{2\sigma^2}} dt.$$

PROPOSITION 4.4.10 Soit  $(m, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$ .

$$\forall x \in \mathbb{R}, \quad \Phi_{m,\sigma}(x) = \Phi\left(\frac{x-m}{\sigma}\right).$$

Ainsi, la courbe de  $\Phi_{m,\sigma}$  se déduit de celle de  $\Phi$  par une affinité horizontale de rapport  $\sigma$  suivie d'une translation de  $m$ .

#### 4.4.4 Moments

THÉORÈME 4.4.11 Soit  $X \hookrightarrow \mathcal{N}(0, 1)$ . Alors  $X$  admet une espérance et une variance, et

$$E(X) = 0, \quad \text{et} \quad V(X) = 1.$$

THÉORÈME 4.4.12 Soit  $X \hookrightarrow \mathcal{N}(m, \sigma^2)$ . Alors  $X$  admet une espérance et une variance, et

$$E(X) = m \quad \text{et} \quad V(X) = \sigma^2, \quad \text{donc:} \quad \sigma(X) = \sigma$$

#### 4.4.5 Stabilité de la loi normale

THÉORÈME 4.4.13 Soit  $X \hookrightarrow \mathcal{N}(m, \sigma^2)$  et  $Y \hookrightarrow \mathcal{N}(m', \sigma'^2)$ . Si  $X$  et  $Y$  sont indépendantes, alors  $X + Y \hookrightarrow \mathcal{N}(m + m', \sigma^2 + \sigma'^2)$

COROLLAIRE 4.4.14 1. Soit  $X_1, \dots, X_n$  mutuellement indépendantes, telles que pour tout  $i \in \llbracket 1, n \rrbracket$ ,  $X_i \hookrightarrow \mathcal{N}(m_i, \sigma_i)$ . Alors

$$X_1 + \dots + X_n \hookrightarrow \mathcal{N}(m_1 + \dots + m_n, \sigma_1^2 + \dots + \sigma_n^2).$$

2. Soit  $X_1, \dots, X_n$  mutuellement indépendantes, suivant toutes la même loi  $\mathcal{N}(m, \sigma^2)$ . Alors

$$X_1 + \dots + X_n \hookrightarrow \mathcal{N}(nm, n\sigma^2).$$

3. En particulier, si  $X_1, \dots, X_n$  sont mutuellement indépendantes et suivent toutes la même loi  $\mathcal{N}(0, 1)$ , alors :

$$X_1 + \dots + X_n \hookrightarrow \mathcal{N}(0, n).$$

### 4.5 Tableaux récapitulatif

Nom	Par.	Not.	densité	fonction de répartition	$E(X)$	$V(X)$	Cas particuliers
<b>uniforme</b>	$[a, b]$	$\mathcal{U}([a, b])$	$f : x \mapsto \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon.} \end{cases}$	$F : x \mapsto \begin{cases} \frac{x-a}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{si } x < a \\ 1 & \text{si } x > b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	
<b>exponentielle</b>	$c$	$\mathcal{E}(c)$	$f : x \mapsto \begin{cases} ce^{-cx} & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases}$	$F : x \mapsto \begin{cases} 1 - e^{-cx} & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases}$	$\frac{1}{c}$	$\frac{1}{c^2}$	
<b>gamma-standard</b>	$\nu$	$\gamma(\nu)$	$f : x \mapsto \begin{cases} \frac{x^{\nu-1}}{\Gamma(\nu)} e^{-x} & \text{si } x > 0 \\ 0 & \text{sinon.} \end{cases}$	$F : x \mapsto \begin{cases} \frac{1}{\Gamma(\nu)} \int_0^x t^{\nu-1} e^{-t} dt & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases}$	$\nu$	$\nu$	$\gamma(1) = \mathcal{E}(1)$
<b>Gamma</b>	$(b, \nu)$	$\Gamma(b, \nu)$	$f : x \mapsto \begin{cases} \frac{x^{\nu-1}}{b^\nu \Gamma(\nu)} e^{-\frac{x}{b}} & \text{si } x > 0 \\ 0 & \text{sinon.} \end{cases}$	$F : x \mapsto \begin{cases} \frac{1}{b^\nu \Gamma(\nu)} \int_0^x t^{\nu-1} e^{-\frac{t}{b}} dt & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases}$	$b\nu$	$b^2\nu$	$\Gamma(1, \nu) = \gamma(\nu)$ $\Gamma(b, 1) = \mathcal{E}(\frac{1}{b})$ $\Gamma(2, \frac{r}{2}), r \in \mathbb{N}^*$ (loi du $\chi^2$ ) $\Gamma(b, n), n \in \mathbb{N}^*$ (loi d'Erlang)
<b>normale</b>	$(m, \sigma^2)$	$\mathcal{N}(m, \sigma^2)$	$f_{m,\sigma} : x \mapsto \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$	$\Phi_{m,\sigma} : x \mapsto \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-m)^2}{2\sigma^2}} dt$ $\Phi_{0,1} = \Phi : x \mapsto \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$	$m$	$\sigma^2$	$\mathcal{N}(0, 1)$ : loi normale centrée réduite.

$t$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7290	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9779	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

$u$	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.8	4.0	4.5
$\Phi(u)$	0.99865	0.99904	0.99931	0.99952	0.99966	0.99976	0.999841	0.999928	0.999968	0.999997

FIGURE 4.6 – Table des valeurs de  $\Phi = \Phi_{0,1}$ .



## Probabilités – Chapitre 5

# Convergences et approximations en probabilité

### 5.1 Convergence en probabilité

#### 5.1.1 Définition

DÉFINITION 5.1.1 Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.r. sur  $(\Omega, \mathcal{T}, P)$ , et  $X$  une autre v.a.r.. On dit que  $X_n$  converge en probabilité (ou converge stochastiquement) vers  $X$ , et on note  $X_n \xrightarrow{P} X$ , si :

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

#### 5.1.2 Inégalité de Bienaymé-Tchebychev

THÉORÈME 5.1.2 (Inégalité de Markov) Soit  $X$  une v.a.r. discrète ou à densité sur  $(\Omega, \mathcal{T}, P)$  telle que  $X(\Omega) \subset \mathbb{R}_+$ , admettant une espérance non nulle  $E(X)$ . Alors, pour tout  $\lambda \in \mathbb{R}_+^*$  :

$$P(X \geq \lambda E(X)) \leq \frac{1}{\lambda}.$$

COROLLAIRE 5.1.3 Sous les mêmes hypothèses, pour tout  $\varepsilon > 0$ ,

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon} \quad \text{et} \quad P(X > \varepsilon) \leq \frac{E(X)}{\varepsilon}$$

COROLLAIRE 5.1.4 Soit  $X$  une variable aléatoire (discrète ou à densité) admettant un moment d'ordre 2. Alors :

$$P(|X| \geq \varepsilon) \leq \frac{E(X^2)}{\varepsilon^2}.$$

THÉORÈME 5.1.5 (Inégalité de Bienaymé-Tchebychev) Soit  $Y$  une v.a.r. discrète ou à densité admettant une espérance  $m$  et une variance  $\sigma^2$ . Alors :

$$\forall \varepsilon > 0, P(|Y - m| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

### 5.1.3 Loi faible des grands nombres

THÉORÈME 5.1.6 (*Loi faible des grands nombres*) Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de v.a.r. (discrètes ou à densité) mutuellement indépendantes suivant une même loi, ayant une espérance  $m$  et une variance  $\sigma^2$ . Soit  $(Z_n)_{n \in \mathbb{N}^*}$  définie par :

$$\forall n \in \mathbb{N}, \quad Z_n = \frac{X_1 + \cdots + X_n}{n}.$$

Alors  $(Z_n)_{n \in \mathbb{N}^*}$  converge en probabilité vers la variable certaine égale à  $m$ . Plus précisément :

$$\forall \varepsilon > 0, \quad \forall n \in \mathbb{N}^*, \quad P(|Z_n - m| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

REMARQUE 5.1.7 L'hypothèse «  $X_n, n \in \mathbb{N}^*$ , mutuellement indépendantes » est plus forte que nécessaire. On peut la remplacer par l'hypothèse «  $X_n, n \in \mathbb{N}^*$ , deux à deux indépendantes » et même par «  $X_n, n \in \mathbb{N}^*$ , deux à deux non corrélées ».

THÉORÈME 5.1.8 (*Théorème d'or de Bernoulli*) Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables mutuellement indépendantes, suivant toutes une loi de Bernoulli de paramètre  $p$ . Soit :

$$\forall n \in \mathbb{N}^*, \quad Z_n = \frac{X_1 + \cdots + X_n}{n}.$$

Alors  $(Z_n)_{n \in \mathbb{N}^*}$  converge en probabilité vers la v.a.r. certaine égale à  $p$ . Plus précisément :

$$\forall \varepsilon > 0, \quad \forall n \in \mathbb{N}^*, \quad P(|Z_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}.$$

Conclusion : La fréquence statistique d'un événement tend vers la probabilité d'apparition de l'événement.

### 5.1.4 Un exemple classique (premier exemple d'estimation)

PROBLÈME 5.1.9 On tire 1000 fois à pile ou face avec une pièce déséquilibrée dont la probabilité d'obtention de Pile est  $p$ . On obtient 570 fois Pile. Donner un intervalle  $I$  tel que la probabilité que  $p \in I$  soit supérieure à 0.9.

Soit pour tout  $i \in \llbracket 1, 1000 \rrbracket$ ,  $X_i$  la variable de Bernoulli de paramètre  $p$  correspondant au résultat de la  $i$ -ème expérience, et soit  $Z_{1000} = \frac{1}{1000}(X_1 + \cdots + X_{1000})$ .

Soit  $\varepsilon > 0$ . D'après le théorème d'or de Bernoulli,  $P(|Z_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$ . Il suffit donc que  $\frac{1}{4n\varepsilon^2} \leq 0.1$ , soit  $\varepsilon \geq 0.05$ . Ainsi,  $P(0.52 \leq p \leq 0.62) \geq 1 - 0.1 = 0.9$ .

### 5.1.5 Une condition suffisante de convergence en probabilité

THÉORÈME 5.1.10 (*hors-programme*) Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.r.d., et  $X$  une autre v.a.r.d., ayant toutes une espérance et une variance. Si  $\lim_{n \rightarrow +\infty} E(X_n) = E(X)$  et si  $\lim_{n \rightarrow +\infty} V(X_n - X) = 0$ , alors  $(X_n)_{n \in \mathbb{N}}$  converge en probabilité vers  $X$ .

## 5.2 Convergence en loi

### 5.2.1 Définition

DÉFINITION 5.2.1 Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.r. et  $X$  une v.a.r.. On dit que  $(X_n)_{n \in \mathbb{N}}$  converge en loi vers  $X$ , et on note  $X_n \xrightarrow{L} X$  si pour tout  $x$  en lequel  $F_X$  est **continu**,

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x).$$

Dans le cas des variables à densité, l'hypothèse de continuité est vérifiée pour tout  $x$  de  $\mathbb{R}$ .

Dans le cas de variables discrètes à valeurs entières, on obtient la reformulation suivante équivalente :

DÉFINITION 5.2.2 Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.r.d. et  $X$  une v.a.r.d.. On suppose que pour tout  $n \in \mathbb{N}$ ,  $X_n(\Omega) \subset \mathbb{Z}$ . On dit que  $(X_n)_{n \in \mathbb{N}}$  converge en loi vers  $X$  si :

$$\forall x \in \mathbb{Z}, \quad \lim_{n \rightarrow +\infty} P(X_n = x) = P(X = x).$$

### 5.2.2 Théorème de la limite centrée

THÉORÈME 5.2.3 (**admis**) Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables aléatoires réelles définies sur un même espace probabilisé  $(\Omega, \mathcal{T}, P)$ , mutuellement indépendantes, de même loi, admettant une espérance  $m$  et un écart-type  $\sigma$ . Soit pour tout  $n \in \mathbb{N}^*$ ,  $S_n = \sum_{k=1}^n X_k$ , et  $S_n^*$  la variable centrée réduite associée. Alors  $S_n^* = \frac{S_n - nm}{\sigma\sqrt{n}}$  et  $S_n^* \xrightarrow{L} X$ , où  $X \hookrightarrow \mathcal{N}(0, 1)$ .

En d'autres termes, pour tous réels  $a < b$ ,  $\lim_{n \rightarrow +\infty} P(a < S_n^* \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt$ .

## 5.3 Approximations

### 5.3.1 Approximation d'une loi hypergéométrique par une loi binomiale

THÉORÈME 5.3.1 Soit  $n \in \mathbb{N}$ ,  $p \in ]0, 1[$  rationnel,  $p = \frac{a}{b}$ ,  $(a, b) \in \mathbb{N}^2$ . Soit, pour tout  $k \in \mathbb{N}^*$ ,  $X_k \hookrightarrow \mathcal{H}(kb, n, p)$ . Alors  $(X_k)_{k \in \mathbb{N}^*}$  converge en loi vers  $X \hookrightarrow \mathcal{B}(n, p)$ .

La restriction aux valeurs  $kb$  du premier paramètre (multiples de  $b$ ) est motivée par le fait que la loi hypergéométrique  $\mathcal{H}(N, n, p)$  n'est bien définie que lorsque  $Np$  est entier (pour pouvoir considérer les coefficients binomiaux). Ainsi, lorsque le premier paramètre devient grand (tout en assurant cette condition), on se rapproche d'une loi binomiale.

**En pratique, on considère qu'on peut approcher une loi hypergéométrique  $\mathcal{H}(N, n, p)$  par une loi binomiale  $\mathcal{B}(n, p)$  si  $N > 10n$ .**

Signification : Lorsque le nombre de boules de l'urne est très grand, le fait de retirer une boule de l'urne ne change pas beaucoup la proportion de boules à succès. Ainsi, on est presque dans le cas d'un tirage avec remise.

### 5.3.2 Approximation d'une loi binomiale par une loi de Poisson

THÉORÈME 5.3.2 Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.r.d. telles que pour tout  $n \in \mathbb{N}$ ,  $X_n \hookrightarrow \mathcal{B}(n, p_n)$ , où  $(p_n)_{n \in \mathbb{N}^*}$  est une suite telle que  $\lim_{n \rightarrow +\infty} np_n = \lambda \in \mathbb{R}_+^*$ . Alors  $(X_n)_{n \in \mathbb{N}^*}$  converge en probabilité vers une v.a.r.d.  $X \hookrightarrow \mathcal{P}(\lambda)$ . Cas particulier :  $\forall n \in \mathbb{N}$ ,  $X_n \hookrightarrow \mathcal{B}(n, \frac{\lambda}{n})$ .

**En pratique, on considère qu'on peut approcher une loi binomiale  $\mathcal{B}(n, p)$  par une loi de Poisson  $\mathcal{P}(np)$  si  $n \geq 30$ ,  $p \leq 0,1$  et  $np \leq 10$ .**

### 5.3.3 Approximation d'une loi binomiale par une loi normale

Soit  $p \in ]0, 1[$  et  $q = 1 - p$ . Soit pour tout  $n \in \mathbb{N}^*$ ,  $X_n \hookrightarrow \mathcal{B}(n, p)$ . Alors  $(X_n^*)_{n \in \mathbb{N}^*}$  converge en loi vers une variable suivant une loi normale centrée réduite.

Ainsi, lorsque  $n$  est grand,  $X_n$  s'approche d'une loi normale, d'espérance  $np$  et de variance  $npq$ .

**En pratique, on considère qu'on peut approcher une loi binomiale  $\mathcal{B}(n, p)$  par une loi normale  $\mathcal{N}(np, npq)$  si  $n \geq 30$ ,  $np \geq 5$  et  $nq \geq 5$ .**

### 5.3.4 Approximation d'une loi de Poisson par une loi normale

Soit  $\mu > 0$ , et soit pour tout  $n \in \mathbb{N}^*$ ,  $X_n \hookrightarrow \mathcal{P}(n\mu)$ . Alors  $(X_n^*)_{n \in \mathbb{N}^*}$  converge en loi vers une variable suivant une loi normale centrée réduite.

Ainsi, lorsque  $n$  est grand,  $X_n$  s'approche d'une loi normale, d'espérance et de variance  $n\mu$ . Ainsi, pour des paramètres  $\lambda = n\mu$  assez grands, une loi  $\mathcal{P}(\lambda)$  peut être approchée par une loi  $\mathcal{N}(\lambda, \lambda)$ .

**En pratique, on considère qu'on peut approcher une loi de Poisson  $\mathcal{P}(\lambda)$  par une loi normale  $\mathcal{N}(\lambda, \lambda)$  si  $\lambda \geq 18$ .**

# Probabilités – Chapitre 6

## Estimateurs

### Position du problème

Étant donné une variable  $X$  suivant une loi appartenant à une famille  $(\mu_\theta)_\theta$  de lois dépendant d'un paramètre réel ou vectoriel  $\theta$ , mais ne connaissant pas le paramètre  $\theta$  associé à la variable  $X$ , déterminer une valeur approchée de  $\theta$ , à l'aide de réalisations successives d'une expérience dont le résultat est la variable  $X$ .

Parfois, il est plus simple de donner une estimation de  $g(\theta)$ , où  $g$  est une fonction dépendant de  $\theta$ , par exemple  $E(X)$ , ou  $V(X)$ .

EXEMPLE 6.0.3 Cas d'une pièce déséquilibrée. Recherche d'une valeur approchée de  $p$ , probabilité d'obtenir  $P$ .

## 6.1 Échantillons d'une loi de probabilité

Soit  $(\Omega, \mathcal{T}, P)$  un espace probablisé, et  $\mathcal{L}$  une loi de probabilité.

DÉFINITION 6.1.1 1. Un échantillon de taille  $n$  (ou  $n$ -échantillon) de la loi  $\mathcal{L}$  est une famille  $\varepsilon_n = (X_1, \dots, X_n)$  de  $n$  variables aléatoires suivant toutes la loi  $\mathcal{L}$ .

2. La loi  $\mathcal{L}$  est appelée loi parente de l'échantillon.

On parle parfois aussi d'*échantillon aléatoire*, à distinguer de l'*échantillon observé*, (correspondant à la réalisation de ce vecteur aléatoire lors d'une expérience donnée)

DÉFINITION 6.1.2 Soit  $(X_1, \dots, X_n)$  un  $n$ -échantillon de loi parente  $\mathcal{L}$ . On dit que l'échantillon  $(X_1, \dots, X_n)$  est indépendant identiquement distribué (en abrégé : i.i.d.) si les  $X_i$  sont mutuellement indépendants.

DÉFINITION 6.1.3 On appelle statistique sur un  $n$ -échantillon i.i.d.  $\varepsilon_n = (X_1, \dots, X_n)$ , une variable aléatoire  $Y = \varphi(X_1, \dots, X_n)$ , où  $\varphi$  est une application de  $\mathbb{R}^n$  dans  $\mathbb{R}$ .

Attention, toutes les fonctions  $\varphi$  ne conviennent pas dans cette définition, seules les fonctions  $\varphi$  telles que  $\varphi(X_1, \dots, X_n)$  soit une variable aléatoire conviennent. C'est le cas notamment si  $\varphi$  est continue.

- EXEMPLES 6.1.4
1. La moyenne empirique :  $\varphi(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$
  2. La variance empirique :  $\varphi(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{1}{n^2} \left( \sum_{i=1}^n X_i \right)^2$
  3.  $\sup(X_1, \dots, X_n)$  et  $\inf(X_1, \dots, X_n)$ .

## 6.2 Estimateurs, estimation ponctuelle

### 6.2.1 Définitions

**But :** Une expérience aléatoire donne un résultat, exprimé sous forme d'une variable aléatoire  $X$ . On cherche, à l'aide de réalisations successives de cette expérience, de déterminer la loi de  $X$ , sachant qu'on sait que  $X$  est dans une certaine famille de lois ( $\mu_\theta$ ), dépendant d'un paramètre  $\theta \in \mathbb{R}^n$  (par exemple la famille des lois de Bernoulli). Il reste alors à déterminer, ou à évaluer  $\theta$ , ou plus généralement,  $g(\theta)$ , où  $g$  est une fonction de  $\mathbb{R}^n$  dans  $\mathbb{R}$ .

On note dans la suite  $\Theta$  l'ensemble des valeurs possibles du paramètre  $\theta$ .

DÉFINITION 6.2.1 Un estimateur de  $g(\theta)$  est une statistique  $\varphi$  sur un échantillon  $\varepsilon_n = (X_1, \dots, X_n)$  de loi parente  $\mu_\theta$ , et à valeurs dans  $g(\Theta)$ .

DÉFINITION 6.2.2 Une estimation de  $g(\theta)$  est une réalisation d'un estimateur  $T_n$  de  $g(\theta)$ . C'est donc une valeur  $T_n(\omega)$ , c'est-à-dire  $\varphi(x_1, \dots, x_n)$ , où  $(x_1, \dots, x_n) = (X_1(\omega), \dots, X_n(\omega))$  est une réalisation de  $(X_1, \dots, X_n)$ .

EXEMPLES 6.2.3

1.  $(X_1, \dots, X_n)$  échantillon i.i.d. de loi parente  $\mathcal{B}(p)$ .

(a)  $S_n = \frac{X_1 + \dots + X_n}{n}$  est un estimateur de  $p$

(b)  $T_n = \frac{2}{n(n+1)} \sum_{k=1}^n kX_k$  est un estimateur de  $p$

(c)  $U_n = 1 - S_n$  est un estimateur de  $p$ , mais sans pertinence !

2.  $X_i \leftrightarrow \mathcal{U}([a, b])$ , i.i.d., alors  $V_n = \sup X_i - \inf X_i$  est un estimateur de  $b - a$ .

### 6.2.2 Biais d'un estimateur

Comme on l'a vu dans l'exemple précédent, la notion d'estimateur est définie indépendamment de la pertinence de l'estimation. Les notions qui suivent introduisent des critères de qualité d'un estimateur, à l'ordre 1 (espérance) et à l'ordre 2 (variance).

DÉFINITION 6.2.4 Soit  $T_n$  un estimateur de  $g(\theta)$ . Le biais de  $T_n$  est :

$$b_{T_n}(\theta) = E(T_n) - g(\theta).$$

On note souvent  $E_\theta(T_n)$  pour  $E(T_n)$ , afin d'indiquer la dépendance vis-à-vis de  $\theta$ . Ne soyez pas effrayés par cette notation, dans laquelle vous pouvez considérer que  $\theta$  n'a qu'un rôle figuratif.

DÉFINITION 6.2.5 On dit qu'un estimateur  $T_n$  est sans biais (ou non biaisé) si  $b_{T_n}(\theta) = 0$ , c'est-à-dire si  $E(T_n) = g(\theta)$ . Dans le cas contraire, on dit que  $T_n$  est un estimateur biaisé.

EXEMPLES 6.2.6 Dans les exemples précédents,  $S_n$  et  $T_n$  sont des estimateurs sans biais, alors que  $U_n$  et  $V_n$  sont biaisés.

Le fait d'être un estimateur sans biais n'est ni une condition nécessaire, ni une condition suffisante pour être un bon estimateur de bonne qualité :

- Si  $T_n$  est sans biais, mais que sa variance est forte, il y a peu de chance qu'une réalisation de  $T_n$  donne une valeur proche de  $E(T_n)$  donc de  $g(\theta)$ . Ainsi, un estimateur sera de meilleure qualité si sa variance est faible (variable concentrée autour de son espérance)
- Un estimateur peut être biaisé mais de faible variance. Si le biais est petit, cela peut être un meilleur estimateur qu'un estimateur non biaisé mais de forte variance.

Il nous faut donc un critère de qualité portant sur la variance.

### 6.2.3 Risque quadratique

DÉFINITION 6.2.7 Soit  $T_n$  un estimateur de  $g(\theta)$ . Si  $T_n$  admet un moment d'ordre 2, on définit le risque quadratique de  $T_n$  (ou erreur quadratique moyenne) par :

$$r_{T_n}(\theta) = E_\theta((T_n - g(\theta))^2).$$

EXEMPLES 6.2.8 1.  $r_{S_n}(\theta) = \frac{1}{n}p(1-p)$

2.  $r_{T_n}(\theta) \underset{+\infty}{\sim} \frac{4}{3n}p(1-p)$

THÉORÈME 6.2.9 1. Si  $T_n$  est un estimateur sans biais, alors  $r_{T_n}(\theta) = V_\theta(T_n)$ .

2. Plus généralement,  $r_{T_n}(\theta) = V_\theta(T_n) + b_{T_n}^2(\theta) = \sigma^2(T_n) + b_{T_n}^2(\theta)$ .

REMARQUE 6.2.10 Si  $r_{T_n}(\theta)$  est faible, à la fois le biais est faible, et la variance de  $T_n$  est faible, donc, même si  $T_n$  est biaisé, sa moyenne est proche de  $g(\theta)$ , et la variable étant concentrée, lors d'une réalisation, la probabilité d'obtenir une valeur proche de cette moyenne est importante. Ce risque quadratique est donc un bon critère de qualité.

### 6.2.4 Convergence d'une suite d'estimateurs

DÉFINITION 6.2.11 On dit que  $(T_n)_{n \in \mathbb{N}^*}$  est une suite d'estimateurs de  $g(\theta)$  si et seulement si :

- pour tout  $n \in \mathbb{N}^*$ ,  $(X_1, \dots, X_n)$  est un  $n$ -échantillon i.i.d.
- pour tout  $n \in \mathbb{N}^*$ ,  $T_n$  est un estimateur de  $g(\theta)$ , défini à l'aide de  $X_1, \dots, X_n$ , autrement dit, il existe une application  $\varphi_n$  de  $\mathbb{R}^n$  dans  $g(\Theta)$  telle que  $T_n = \varphi_n(X_1, \dots, X_n)$ .

Souvent, l'expression de  $\varphi_n$  est générale, et la donnée de  $\varphi_n(X_1, \dots, X_n)$  pour un  $n$  particulier se généralise pour toute valeur de  $n$ .

Par abus de notation, on note souvent  $T_n$  au lieu de  $(T_n)_{n \in \mathbb{N}^*}$ , et par abus de langage, on dit simplement que  $T_n$  est un estimateur, au lieu de dire que c'est une suite d'estimateurs.

EXEMPLE 6.2.12 En définissant l'estimateur  $S_n = \frac{X_1 + \dots + X_n}{n}$ , on définit en fait la suite d'estimateurs  $(S_n)_{n \in \mathbb{N}^*}$ .

DÉFINITION 6.2.13 On dit qu'une suite d'estimateurs  $(T_n)_{n \in \mathbb{N}^*}$  est asymptotiquement sans biais si  $\lim_{n \rightarrow +\infty} E(T_n) = g(\theta)$ , c'est-à-dire si  $\lim_{n \rightarrow +\infty} b_{T_n}(\theta) = 0$ .

Par abus de langage, on dit plus simplement que l'estimateur  $T_n$  est asymptotiquement sans biais.

EXEMPLE 6.2.14 Soit  $(X_1, \dots, X_n)$  i.i.d. de loi parente  $\mathcal{U}[a-1, a]$ , et  $T_n = \max(U_1, \dots, U_n)$ . Alors  $T_n$  est un estimateur asymptotiquement sans biais de  $a$ .

PROPOSITION 6.2.15 Soit  $T_n$  un estimateur ? Alors le risque quadratique de  $T_n$  tend vers 0 lorsque  $n$  tend vers  $+\infty$  si et seulement si  $V(T_n)$  tend vers 0 et  $T_n$  est asymptotiquement sans biais.

DÉFINITION 6.2.16 On dit que la suite d'estimateur  $(T_n)_{n \in \mathbb{N}^*}$  (ou l'estimateur  $T_n$ ) est convergent (ou consistant) si la suite  $(T_n)_{n \in \mathbb{N}^*}$  converge en probabilité vers la variable certaine  $g(\theta)$ , c'est-à-dire si :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} P(|T_n - g(\theta)| > \varepsilon) = 0.$$

PROPOSITION 6.2.17 Soit  $(X_1, \dots, X_n)$  un  $n$ -échantillon i.i.d., de loi parente  $\mu_\theta$ , et d'espérance  $E_\theta(X_i) = g(\theta)$ , dépendant de  $\theta$ . Alors  $S_n = \frac{X_1 + \dots + X_n}{n}$  est un estimateur sans biais et convergent de  $g(\theta)$ .

REMARQUE 6.2.18 La convergence d'un estimateur est souvent obtenue grâce à l'inégalité de Bienaymé-Tchebychev, ou grâce à la loi faible des grand nombres qui en découle.

Les deux propositions qui suivent sont bien pratiques, mais pas au programme : refaire la démonstration si vous voulez les utiliser.

PROPOSITION 6.2.19 Soit  $T_n$  un estimateur sans biais tel que  $\lim_{n \rightarrow +\infty} V(T_n) = 0$ . Alors  $T_n$  est un estimateur convergent.

PROPOSITION 6.2.20 Soit  $T_n$  un estimateur asymptotiquement sans biais tel que  $\lim_{n \rightarrow +\infty} V(T_n) = 0$ . Alors  $T_n$  est convergent.

EXEMPLE 6.2.21  $X_i \hookrightarrow \mathcal{U}[a-1, a]$ , mutuellement indépendantes,  $T_n = \max(X_1, \dots, X_n)$ . Alors  $T_n$  est un estimateur convergent de  $a$ .

## 6.3 Estimation par intervalles de confiance

### 6.3.1 Notion d'intervalle de confiance

Soit  $T_n$  un estimateur de  $g(\theta)$ , même de bonne qualité. On ne peut pas être assuré que  $T_n(\omega)$  sera une bonne approximation de  $g(\theta)$ . Si l'estimateur est de bonne qualité, on pourra simplement dire qu'il y a de fortes chances que  $T_n(\omega)$  soit une bonne approximation de  $g(\theta)$ . De plus, la notion de « bonne approximation » est toute relative, et nécessite d'être quantifiée.

Étant donné  $\varepsilon > 0$ , la probabilité que  $T_n(\omega)$  appartienne à  $]g(\theta) - \varepsilon, g(\theta) + \varepsilon[$  est d'autant plus proche de 1 que l'estimateur  $T_n$  est de bonne qualité. C'est donc, pour une valeur de  $\varepsilon$  donnée, cette probabilité d'appartenance (ou taux de confiance) qui jugera de la fiabilité de l'approximation.

Par exemple, un estimateur peut donner une valeur approchée à 0.1 près de  $g(\theta)$ , avec un taux de

confiance de 0.9, ce qui signifie :

$$P(|T_n - g(\theta)| < 0.1) > 0.9 \quad \text{donc:} \quad P(g(\theta) - \varepsilon < T_n < g(\theta) + \varepsilon) > 0.9$$

C'est cette notion d'intervalle et de taux de confiance qui fait l'objet de cette section. Dans sa définition générale, l'intervalle de confiance a des bornes aléatoires, comme dans l'exemple étudié ci-dessus (les bornes aléatoires étant ici  $T_n - \varepsilon$  et  $T_n + \varepsilon$ )

**DÉFINITION 6.3.1** Soit  $(X_1, \dots, X_n)$  un  $n$ -échantillon i.i.d. et  $U_n$  et  $V_n$  deux statistiques sur  $(X_1, \dots, X_n)$ , par exemple des estimateurs. Soit  $\alpha \in ]0, 1[$ .

- On dit que l'intervalle  $[U_n, V_n]$  est un intervalle de confiance de  $g(\theta)$  au niveau de confiance  $1 - \alpha$  si :

$$P_\theta(U_n \leq g(\theta) \leq V_n) = 1 - \alpha$$

- On dit que l'intervalle  $[U_n, V_n]$  est un intervalle de confiance de  $g(\theta)$  au niveau de confiance au moins égal à  $1 - \alpha$  si :

$$P_\theta(U_n \leq g(\theta) \leq V_n) \geq 1 - \alpha.$$

Le réel  $\alpha$  est appelé taux de risque

**DÉFINITION 6.3.2** Une estimation de l'intervalle de confiance est la réalisation de l'intervalle de confiance :  $[U_n(\omega), V_n(\omega)]$ .

### 6.3.2 Exemple important : espérance d'une loi normale de $\sigma^2$ donné

Soit  $X_i \hookrightarrow \mathcal{N}(m, \sigma^2)$ ,  $\sigma^2$  étant donné, et les  $X_i$  étant mutuellement indépendants. Soit :

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n}$$

la moyenne empirique.

**PROPOSITION 6.3.3** L'intervalle  $[\overline{X}_n - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \frac{\alpha}{2}), \overline{X}_n + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \frac{\alpha}{2})]$  est un intervalle de confiance de  $m$  au niveau de confiance  $1 - \alpha$ .

**Ce résultat ne doit pas être utilisé tel quel, il faut refaire le raisonnement à chaque fois**

La valeur  $\Phi^{-1}(1 - \frac{\alpha}{2})$  se trouve par lecture inverse de la table de valeurs de  $\Phi$ .

**REMARQUE 6.3.4** Pour trouver cet intervalle de confiance, on a utilisé la symétrie de  $\Phi$ , en répartissant de manière égale la marge d'erreur  $\alpha$  aux deux infinis. On obtient de la sorte l'intervalle  $[-\Phi^{-1}(1 - \frac{\alpha}{2}), \Phi^{-1}(1 - \frac{\alpha}{2})]$ , qu'on appelle intervalle de dispersion symétrique de niveau  $1 - \alpha$  de  $\overline{X}_n^*$ . On n'était pas obligé de faire ce choix : on aura pu aussi répartir la marge d'erreur en  $\frac{3\alpha}{4}$  au voisinage de  $+\infty$  et  $\frac{\alpha}{4}$  au voisinage de  $-\infty$ .

**REMARQUE 6.3.5** Cette méthode peut être généralisée au cas d'un estimateur  $T_n$  tel qu'il existe une application injective  $\varphi$  telle que  $\varphi(T_n)$  suivent une loi ne dépendant pas de  $\theta$  (ici,  $\varphi : x \mapsto \frac{(x-m)\sqrt{n}}{\sigma^2}$  nous ramène à la loi normale centrée réduite).

### 6.3.3 Cas typique : estimation d'une espérance à l'aide de la moyenne empirique

L'exemple précédent est un exemple générique, à approximation près. En effet, dans le cas général,  $\overline{X}_n^*$  converge en loi vers la loi normale centrée réduite, et pour  $n$  assez grand, on peut supposer que  $\overline{X}_n^*$  suit une loi normale centrée réduite. Le raisonnement précédente s'applique alors, si  $X_n$  est de variance connue.

### 6.3.4 Exemple : estimation par intervalle de confiance du paramètre d'une loi de Bernoulli

Soit  $X_i \hookrightarrow \mathcal{B}(p)$ , mutuellement indépendants. Alors, par la démarche précédente,

$$\left[ \overline{X}_n - \frac{1}{2\sqrt{n}} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right), \overline{X}_n + \frac{1}{2\sqrt{n}} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \right]$$

est un intervalle de confiance de  $p$  de niveau de confiance supérieur à  $1 - \alpha$ .

- À taux de confiance  $1 - \alpha$  et précision  $\beta$  imposés, trouver  $n$  nécessaire.  
Exemple :  $\alpha = \beta = 0.01$ , on trouve  $\sqrt{n} \geq 128.7$  donc  $n \geq 16577$ .
- À taux de confiance imposé, et  $n$  imposé, trouver un intervalle de confiance  
Exemple :  $\alpha = 0.01$ ,  $n = 1000$ , on trouve  $\beta = 0.0407$ .
- À précision imposée, et  $n$  imposé, trouver le taux de confiance.  
Exemple :  $n = 1000$ ,  $\beta = 0.01$ , on trouve  $\alpha = 0.52$ .

Comparaison avec la méthode déjà vue utilisant la formule de Bienaymé-Tchebychev.

Pour obtenir par cette méthode un intervalle de confiance à la précision 0.01, avec un niveau de risque de 0.01, il faut  $n \geq 250000$ . Cette méthode est donc beaucoup moins bonne !

# Probabilités – Chapitre 7

## Éléments de statistiques

**But :** À partir de données brutes, qualitatives (non chiffrées) ou quantitatives (chiffrées), définir des paramètres permettant de cerner du premier coup d’œil les tendances générales de la population étudiée : représentations graphiques, ordre de grandeur (moyenne, mode, médiane...), estimation de l’étalement (écart-type, quartiles...).

### 7.1 Définitions – Terminologie élémentaire

#### 7.1.1 Statistiques

DÉFINITION 7.1.1 Une *statistique* est une application  $X : \Omega \rightarrow C$ , où  $\Omega$  est un ensemble fini, et  $C$  un ensemble quelconque.

TERMINOLOGIE 7.1.2 Les objets en jeu portent des noms très liés à l’étude de la population d’un état, qui est l’origine des statistiques :

- $\Omega$  est appelé *population*,
- un élément  $x$  de  $\Omega$  est appelé *individu*,
- $C$  est appelé *ensemble des caractères*,
- un élément  $c$  de  $C$  est un *caractère*,
- un *échantillon* est un sous-ensemble de  $\Omega$ . On étudie souvent une statistique en analysant sa restriction sur un échantillon le plus représentatif possible de la population.

Ainsi, une statistique est une fonction qui à chaque individu associe un caractère. Ce caractère peut être chiffré (la taille, le poids, l’âge...) ou non (une couleur, le fait de porter des lunettes...)

TERMINOLOGIE 7.1.3 On dit qu’un caractère est *qualitatif* s’il ne peut pas être chiffré, et *quantitatif* s’il peut l’être.

Pour l’étude mathématique des statistiques, on s’intéressera essentiellement à des caractères quantitatifs. Ainsi,  $C$  est égal à un sous-ensemble de  $\mathbb{R}$ . Quitte à prolonger  $X$  sur son ensemble image, on peut supposer que  $C = \mathbb{R}$ .

### 7.1.2 Effectif, fréquence

Soit  $X : \Omega \rightarrow \mathbb{R}$  une statistique quantitative. La population  $\Omega$  étant finie (hypothèse de la définition d'une statistique), son image  $X(\Omega)$  par  $X$  est également finie. On note  $\{x_1, \dots, x_p\}$  les éléments de  $X(\Omega)$ . Il s'agit d'un sous-ensemble fini de  $\mathbb{R}$ , qu'on peut donc ordonner. On supposera toujours dans la suite que les indices sont choisis de sorte que  $x_1 < \dots < x_p$ .

DÉFINITION 7.1.4 Soit  $x_i \in X(\Omega)$ ,  $(a, b) \in \mathbb{R}^2$ .

- L'effectif  $n_i$  de la valeur  $x_i$  est le nombre d'individus de  $\Omega$  ayant le caractère  $x_i$ . Ainsi,  $n_i = |X^{-1}(x_i)|$ .
- L'effectif cumulé en  $x_i$  est le nombre d'individu dont le caractère est au plus  $x_i$  : il s'agit donc de  $|X^{-1}(]-\infty, x_i])|$ , soit encore de  $\sum_{j=1}^i n_j$ .
- L'effectif de l'intervalle  $]a, b]$  est  $|X^{-1}(]a, b])|$  : c'est le nombre d'individus dont le caractère est dans l'intervalle  $]a, b]$ .
- Plus généralement, l'effectif d'un intervalle  $I$  est  $|X^{-1}(I)|$ .
- L'effectif cumulé en  $b$  (non plus forcément une valeur de  $X(\Omega)$ ) est l'effectif de  $]-\infty, b]$ , soit  $|X^{-1}(]-\infty, b])|$ .

DÉFINITION 7.1.5 Soit  $x_i \in X(\Omega)$ ,  $(a, b) \in \mathbb{R}^2$ . Soit  $n = |\Omega|$ .

- La fréquence  $f_i$  de la valeur  $x_i$  est la proportion d'individus ayant le caractère  $x_i$ . Ainsi,  $f_i = \frac{n_i}{n}$ .
- La fréquence cumulée en  $x_i$  est  $\sum_{j=1}^i f_j = \frac{1}{n} \sum_{j=1}^i n_j$ .
- La fréquence d'un intervalle  $I$  est  $\frac{1}{n} \cdot |X^{-1}(I)|$ .

### 7.1.3 Séries statistiques

La donnée de la fonction statistique  $X$  est souvent inutile, peu maniable, voire non souhaitable (non anonyme). Pour l'étude globale de la population, il est important de savoir le nombre d'individus ayant tel caractère, mais leur identité importe peu. On crée donc un anonymat en ne retenant d'une série statistique que les fréquences ou les effectifs. On obtient ainsi ce qu'on appelle une série statistique. On distingue suivant que l'ensemble des caractères possibles est fini ou non.

DÉFINITION 7.1.6 Une *série statistique discrète* est une famille  $\{(x_i, n_i), i \in I\}$ , où  $I$  est fini ou dénombrable ; les  $x_i$  représentent les valeurs possibles d'un caractère et  $n_i$  est l'effectif correspondant. On représente également parfois une série statistique par les fréquences :  $\{(x_i, f_i), i \in I\}$

Dans le cas où l'ensemble des valeurs possibles du caractère n'est pas énumérable, ou même dans le cas où il est énumérable mais grand (dénombrable, ou même fini, mais de grand cardinal, donc peu maniable), on effectue des regroupements de valeurs, pour se ramener à un nombre plus restreint (généralement fini) de valeurs.

DÉFINITION 7.1.7 Un *classe* de caractère est un regroupement de valeurs possibles du caractère.

Souvent, si  $C = \mathbb{R}$ , on effectue le découpage suivant :

$$]-\infty, a_1], ]a_1, a_2], \dots, ]a_{p-1}, a_p], ]a_p, +\infty[,$$

où  $a_1 < a_2 < \dots < a_p$ . On parlera des classes définies par les bornes  $(a_1, \dots, a_p)$ .

On remarquera qu'un découpage admissible de  $C$  doit consister en une *partition* de  $C$  afin que chaque individu soit compté une et une seule fois !

**DÉFINITION 7.1.8** Une *série statistique continue*, ou *groupée* est la donnée d'une partition  $\{C_i, i \in I\}$  de  $C$  en intervalles, et d'une famille  $\{(C_i, n_i), i \in I\}$ ,  $n_i$  correspondant à l'effectif de la classe  $C_i$ .

**DÉFINITION 7.1.9** Soit  $\{(C_i, n_i), i \in \llbracket 1, p \rrbracket\}$  une série statistique continue, où pour tout  $i \in \llbracket 1, p \rrbracket$ ,  $C_i = ]a_i, a_{i+1}]$ ,  $(a_i)_{i \in \llbracket 1, p+1 \rrbracket}$  étant une suite croissante de réels. Sa *discrétisation* est la série statistique discrète (finie)  $\{(x_i, n_i), i \in \llbracket 1, p \rrbracket\}$ , où pour tout  $i \in \llbracket 1, p \rrbracket$ ,  $x_i = \frac{a_i + a_{i+1}}{2}$ . Ainsi, on concentre chaque classe en son milieu.

**REMARQUE 7.1.10** Si les intervalles extrémaux sont non bornés, ou peu représentatifs, il arrive souvent qu'on nous donne les valeurs moyennes au sein de ces classes. On peut alors encore définir une discrétisation en concentrant l'effectif de ces classes en leur valeur moyenne.

## 7.1.4 Représentations graphiques

### 7.1.4.1 Séries discrètes

#### 1. Diagrammes en bâton

On met en abscisse les valeurs du caractère, et en ordonnée les effectifs (pour le diagramme des effectifs) ou les fréquences (pour le diagramme des fréquences).

L'effectif d'une valeur  $x_i$  est représenté par un bâton d'abscisse  $x_i$  de hauteur  $n_i$ .

#### 2. Polygone des effectifs

Il est obtenu en reliant le sommet des bâtons du diagramme en bâton.

#### 3. Diagramme et polygone des effectifs/fréquences cumulés

Pareil que pour les effectifs, mais en considérant les effectifs cumulés en chacune des valeurs  $x_i$ . En particulier, la taille des bâtons est croissante.

#### 4. Fonction de répartition

La fonction de répartition d'une série statistique  $X$  est la fonction  $F_X$  de  $\mathbb{R}$  dans  $[0, 1]$  qui à tout  $x \in \mathbb{R}$  associe la fréquence cumulée  $F_X(x) = \sum_{i|x_i \leq x} f_i$ .

Le graphe de la fonction de répartition est formé de plateaux reliant les sommets des bâtons au bâton suivant (inclus du côté gauche, exclus du côté droit) ; il est complété par un plateau égal à 1 allant à  $+\infty$  et un plateau égal à 0 allant à  $-\infty$ .

### 7.1.4.2 Séries continues

#### 1. Histogramme

On représente en abscisse les bords des intervalles représentant les classes de la série statistique. Une classe  $]a_i, a_{i+1}]$  est représentée par un rectangle de base le segment d'abscisse  $[a_i, a_{i+1}]$ , et dont l'aire est égale à l'effectif  $n_i$  de la classe (donc de hauteur  $\frac{n_i}{a_{i+1} - a_i}$ .)

#### 2. Polygone des effectifs/fréquences

Il est obtenu en reliant le milieu des cotés hauts des rectangles de l'histogramme.

### 3. Diagramme et polygone des fréquences cumulés

C'est un diagramme en bâtons représentant en chacune des coordonnées  $a_i$  séparant les classes la fréquence cumulée en ces valeurs.

### 4. Fonction de répartition

La fonction de répartition est la fonction qui à  $x$  associe la fréquence cumulée  $F_X(x)$  en  $x$ , avec l'hypothèse que l'effectif des classes est uniformément réparti dans chaque classe. Ainsi, elle est affine sur chaque intervalle  $]a_i, a_{i+1}]$ . Son expression est donc :

$$\forall i \in \llbracket 1, p \rrbracket, \forall x \in ]a_i, a_{i+1}], f(x) = \sum_{j=1}^{i-1} f_j + \frac{x - a_i}{a_{i+1} - a_i} \cdot f_i.$$

Ainsi, le graphe de la fonction de répartition est le polygone reliant le sommet des bâtons du diagramme des fréquences cumulées, complété éventuellement par 0 et 1 à gauche et à droite.

## 7.2 Paramètres de position

**But :** Avoir un ordre de grandeur des valeurs prises par le caractère.

### 7.2.1 Modes, moyennes

DÉFINITION 7.2.1 • Un *mode* d'une série statistique discrète est une valeur  $x$  du caractère dont l'effectif est maximal.

- Une *classe modale* d'une série statistique groupée est une classe pour laquelle la concentration de l'effectif (effectif/taille de la classe) est maximale, donc correspondant à un rectangle de hauteur maximale dans l'histogramme.

REMARQUE 7.2.2 Le mode (ou la classe modale) n'est pas forcément unique.

DÉFINITION 7.2.3 Voici différents types de moyennes d'une série discrète  $\{(x_i, n_i), i \in \llbracket 1, p \rrbracket\}$  :

- Moyenne arithmétique  $\bar{x} : x = \frac{1}{n} \sum_{i=1}^p n_i x_i = \sum_{i=1}^p f_i x_i$ .
- Moyenne géométrique  $g : g = \sqrt[n]{\prod_{i=1}^p x_i^{n_i}} = \prod_{i=1}^p x_i^{f_i}$ .
- Moyenne harmonique  $h : \frac{1}{h} = \frac{1}{n} \sum_{i=1}^p \frac{n_i}{x_i} = \sum_{i=1}^p \frac{f_i}{x_i}$ .
- Moyenne quadratique  $q : q = \sqrt{\frac{1}{n} \sum_{i=1}^p n_i x_i^2} = \sqrt{\sum_{i=1}^p f_i x_i^2}$ .

DÉFINITION 7.2.4 Soit  $\{(x_i, n_i), i \in \llbracket 1, p \rrbracket\}$  une série statistique discrète, et soit  $k \in \mathbb{N}$ .

Le moment d'ordre  $k$  est  $m_k = \sum_{i=1}^p f_i x_i^k$ .

EXEMPLE 7.2.5  $m_0 = 1, m_1 = \bar{x}, m_2 = q^2$ .

DÉFINITION 7.2.6 (Cas des séries statistiques groupées). On définit les différentes moyennes et les moments comme étant les moyennes et moments de leur discrétisation.

## 7.2.2 Médiane

La médiane est intuitivement la valeur du caractère séparant l'effectif en deux parts égales. Cette notion va dépendre de la parité de  $|\Omega|$ .

### Cas d'une série statistique discrète

Soit  $\{(x_i, n_i), i \in \llbracket 1, p \rrbracket\}$  une série statistique discrète.

On ordonne les individus de  $\Omega$  par ordre croissant (au sens large) de caractère :  $\Omega = \{\omega_1, \dots, \omega_n\}$ , où  $X(\omega_1) \leq \dots \leq X(\omega_n)$ . (Notez que puisque dans une série statistique, on perd les données concernant  $\Omega$  à part son cardinal, cela revient à recréer fictivement  $\Omega$ , c'est-à-dire à remettre des étiquettes, mais anonymes sur les différentes personnes, un peu comme les codes d'anonymat lors d'examens ou concours.)

DÉFINITION 7.2.7 La médiane  $m$  de la série  $\{(x_i, n_i), i \in \llbracket 1, p \rrbracket\}$  est :

- $m = X(\omega_{\frac{n+1}{2}})$  si  $n = |\Omega|$  est impair ;
- $m = \frac{1}{2} (X(\omega_{\frac{n}{2}}) + X(\omega_{\frac{n}{2}+1}))$  si  $n$  est pair.

Remarquez que  $m$  peut prendre une valeur hors de  $C$ .

### Cas d'une série statistique continue

Soit  $\{(C_i, n_i), i \in \llbracket 1, p \rrbracket\}$  une série statistique groupée (dont on a borné les extrêmes si besoin), dont l'effectif de chaque classe est non nul. Alors la fonction de répartition  $F_X$  est continue et strictement croissante, et définit donc une bijection de  $\mathbb{R}$  sur  $[0, 1]$ . Ainsi, on peut définir :

DÉFINITION 7.2.8 La médiane  $m$  de la série  $\{(C_i, n_i), i \in \llbracket 1, p \rrbracket\}$  est l'unique antécédant  $F_X^{-1}(\frac{1}{2})$  de  $\frac{1}{2}$  par  $F_X$ . Ainsi,  $m$  est la seule valeur telle que  $F_X(m) = \frac{1}{2}$ .

REMARQUE 7.2.9 Si une classe est d'effectif nul, la fonction de répartition présente un plateau. Si un tel plateau prend la valeur  $\frac{1}{2}$ , on ne peut pas définir la médiane de la manière précédente. La médiane sera alors le milieu du plateau.

Calcul pratique :

On repère la classe qui fait passer l'effectif (cumulé) au dessus de la moitié (ou la fréquence cumulée au dessus de  $\frac{1}{2}$ ). On effectue ensuite une règle de 3 sur cette classe. Soit  $C_i = ]a_i, a_{i+1}]$  cette classe, de fréquences cumulées  $g_i$  en  $a_i$  et  $g_{i+1}$  en  $a_{i+1}$ . Alors :

$$\frac{m - a_i}{a_{i+1} - a_i} = \frac{\frac{1}{2} - g_i}{g_{i+1} - g_i}.$$

On peut bien sûr le faire avec la fréquence cumulée, en remplaçant  $\frac{1}{2}$  par  $\frac{n}{2}$ .

## 7.3 Paramètres de dispersion

### 7.3.1 Écart-type

#### Cas d'une série discrète

Soit  $\{(x_i, n_i)_{i \in \llbracket 1, p \rrbracket}\}$  une série statistique discrète.

DÉFINITION 7.3.1 • L'écart arithmétique moyen est  $E = \frac{1}{n} \sum_{i=1}^p n_i |x_i - \bar{x}| = \sum_{i=1}^p f_i |x_i - \bar{x}|$ .

- L'écart quadratique moyen (ou écart-type) est  $\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^p n_i (x_i - \bar{x})^2} = \sqrt{\sum_{i=1}^p f_i (x_i - \bar{x})^2}$ .
- La variance est  $V = \sigma^2$ .

DÉFINITION 7.3.2 On appelle plus généralement moment centré d'ordre  $k$  ( $k \in \mathbb{N}$ ) la quantité :

$$\mu_k = \sum_{i=1}^p f_i (x_i - \bar{x})^k.$$

EXEMPLE 7.3.3  $\mu_0 = 1$ ,  $\mu_1 = 0$ ,  $\mu_2 = V$ .

### Cas d'une série continue

On définit de même l'écart arithmétique moyen, l'écart-type, la variance et les moments centrés d'une série statistique continue, en considérant sa discrétisation.

PROPOSITION 7.3.4 (Relation de Koenig-Huyghens)  $\sigma^2 = q^2 - \bar{x}^2$ , soit :  $\mu_2 = m_2 - m_1^2$ .

## 7.3.2 Quantiles

Nous avons déjà vu la médiane, qui peut aussi être vu comme un paramètre de dispersion, par comparaison à la moyenne. Intuitivement, si  $m < \bar{x}$ , l'effectif est en majorité sous la moyenne. Par compensation, l'écart par rapport à la moyenne des individus au dessus de la moyenne doit être plus important que l'écart par rapport à la moyenne des individus se situant en dessous. Ainsi, la série est plus étalée sur les valeurs supérieures, et plus concentrée sur les valeurs inférieures. Inversement si  $m > \bar{x}$ .

Nous voyons d'autres paramètres similaires à la médiane, qui permettent de situer le quart, le dixième etc. de l'effectif. **Nous nous limitons au cas d'une série groupée**  $\{(C_i, n_i), i \in \llbracket 1, p \rrbracket\}$  dont les effectifs de chaque classe sont non nuls, et qu'on a borné aux extrêmes

DÉFINITION 7.3.5 • Le *premier quartile* est l'unique valeur  $q_1$  telle que  $F_X(q_1) = \frac{1}{4}$ .

- Le *troisième quartile* est l'unique valeur  $q_3$  telle que  $F_X(q_3) = \frac{3}{4}$ .
- L'*intervalle interquartile* est l'intervalle  $[q_1, q_3]$ .
- L'*espace interquartile* est la longueur de cet intervalle, à savoir  $q_3 - q_1$ .

Ainsi, l'intervalle interquartile est l'intervalle regroupant la moitié centrale des individus. L'espace interquartile est d'autant plus petit que la série est concentrée en son centre.

On définit de même :

DÉFINITION 7.3.6 • Soit  $k \in \llbracket 1, 9 \rrbracket$ . Le  $k$ -ième *décile* est  $d_k$  tel que  $F_X(d_k) = \frac{k}{10}$ .

- Soit  $k \in \llbracket 1, 99 \rrbracket$ . Le  $k$ -ième *centile* (ou *percentile*) est  $c_k$  tel que  $F_X(c_k) = \frac{k}{100}$ .
- Soit  $k \in \llbracket 1, 999 \rrbracket$ . Le  $k$ -ième *millile* est  $mil_k$  tel que  $F_X(mil_k) = \frac{k}{1000}$ .

Principe de calcul :

C'est le même que pour la médiane : en considérant les fréquences ou les effectifs cumulés, on repère la classe dans laquelle se situe le quantile à calculer, et on effectue une règle de 3.

## 7.4 Statistiques descriptives bivariées

Le but est d'observer deux caractères d'un même individu simultanément, et éventuellement d'établir une relation (moyenne) entre ces deux caractères. Par exemple, on peut chercher à établir un lien entre poids et taille d'un individu (relation non linéaire, voir l'indice de masse corporel). Nous développons ici un exemple assez simple, concernant les notes obtenues à deux devoirs de mathématiques par des mêmes élèves. La corrélation est ici évidente : si les élèves sont réguliers, les notes obtenues aux deux devoirs devraient sensiblement être les mêmes, donc on devrait avoir une relation affine moyenne du type  $y = x$ . Ce relevé de notes est bien réel, mais ne correspond pas à un relevé de votre classe.

On se limite ici au cas de l'étude de deux caractères quantitatifs, le cas de caractères qualitatifs ou d'une association d'un caractère quantitatif et d'un caractère qualitatif (par exemple le salaire par répartition par classe socio-professionnelle d'une population) nécessitant d'autres types d'étude.

### 7.4.1 Définitions et représentations

**DÉFINITION 7.4.1** Une série statistique double sur un échantillon  $\Omega' = \{\omega_1, \dots, \omega_n\}$  de la population est la donnée de  $n$  couples  $(x_i, y_i)$ ,  $i \in \llbracket 1, n \rrbracket$ ,  $x_i$  correspondant à la mesure du premier caractère ( $A$ ) de l'individu  $\omega_i$ , et  $y_i$  à la mesure du second caractère ( $B$ ) du même individu.

La série statistique double est aussi parfois par le couple  $(\mathbf{X}, \mathbf{Y})$  de vecteurs de  $\mathbb{R}^n$ , où

$$\mathbf{X} = (x_1, \dots, x_n) \quad \text{et} \quad \mathbf{Y} = (y_1, \dots, y_n)$$

représentent respectivement les séries statistiques (simples) liées à chacun des caractères  $A$  et  $B$ .

**EXEMPLE 7.4.2** Les notes obtenues par une classe de 38 élèves lors de la première épreuve de mathématiques d'un concours blanc sont données par le vecteur  $\mathbf{X}$  suivant :

$$\mathbf{X} = (2.6, 20, 15, 8.9, 8.1, 15.7, 14.3, 8.9, 10.3, 11.3, 4.1, 10.3, 2.8, 9.3, 8.8, 10.9, 7.3, 7.7, \\ 8.8, 4.3, 8.9, 6.7, 6.8, 6.1, 4.3, 2, 4.4, 1.5, 8.2, 7, 4.9, 2.4, 6, 6.3, 6.4, 5, 3.5, 0.9)$$

Les notes de la deuxième épreuve sont données, dans le même ordre sur les élèves, par le vecteur

$$\mathbf{Y} = (0.3, 20, 15, 14.9, 7.3, 16.4, 10.7, 10.4, 7.9, 9.3, 8.7, 14, 5, 9.4, 8.9, 5.7, 8.1, 6.3, 10.9, 5.9, 8.4, 6.9, \\ 7.9, 7.7, 3, 8.9, 4.3, 5, 6.9, 3.4, 5.4, 5.1, 3.9, 2.3, 3.3, 3.6, 3.4, 3.1)$$

Vous remarquerez que cette présentation n'est pas très lisible. Notamment, les données sont données de manière trop compactes, et on a du mal à associer les valeurs des deux séries statistiques. La présentation sous forme de couples est déjà plus satisfaisante :

$$(2.6, 0.3), (20, 20), (15, 15), (8.9, 14.9), (8.1, 7.3), (15.7, 16.4), (14.3, 10.7), (8.9, 10.4), (10.3, 7.9), \\ (11.3, 9.3), (4.1, 8.7), (10.3, 14), (2.8, 5), (9.3, 9.4), (8.8, 8.9), (10.9, 5.7), (7.3, 8.1), (7.7, 6.3), \\ (8.8, 10.9), (4.3, 5.9), (8.9, 8.4), (6.7, 6.9), (6.8, 7.9), (6.1, 7.7), (4.3, 3), (2, 8.9), (4.4, 4.3), (1.5, 5), \\ (8.2, 6.9), (7, 3.4), (4.9, 5.4), (2.4, 5.1), (6, 3.9), (6.3, 2.3), (6.4, 3.3), (5, 3.6), (3.5, 3.4), (0.9, 3.1)$$

La représentation suivante, sous forme de tableau, est encore plus lisible. On a numéroté les élèves de 1 à 38 dans la figure 7.1

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$x_i$	2.6	20	15	8.9	8.1	15.7	14.3	8.9	10.3	11.3	4.1	10.3	2.8	9.3
$y_i$	0.3	20	15	14.9	7.3	16.4	10.7	10.4	7.9	9.3	8.7	14	5	9.4
i	15	16	17	18	19	20	21	22	23	24	25	26	27	28
$x_i$	8.8	10.9	7.3	7.7	8.8	4.3	8.9	6.7	6.8	6.1	4.3	2	4.4	1.5
$y_i$	8.9	5.7	8.1	6.3	10.9	5.9	8.4	6.9	7.9	7.7	3	8.9	4.3	5
i	29	30	31	32	33	34	35	36	37	38				
$x_i$	8.2	7	4.9	2.4	6	6.3	6.4	5	3.5	0.9				
$y_i$	6.9	3.4	5.4	5.1	3.9	2.3	3.3	3.6	3.4	3.1				

FIGURE 7.1 – tableau des notes aux deux épreuves

DÉFINITION 7.4.3 Soit une série statistique  $(x_i, y_i)_{i \in [1, n]}$ . Le nuage de points associée à cette série statistique est l'ensemble des points de  $\mathbb{R}^2$  de coordonnées  $(x_i, y_i)$ .

EXEMPLE 7.4.4 Reprenant l'exemple précédent, on obtient le nuage de points représenté dans la figure 7.2. La croix (+) représente le point moyen du nuage (voir plus loin)

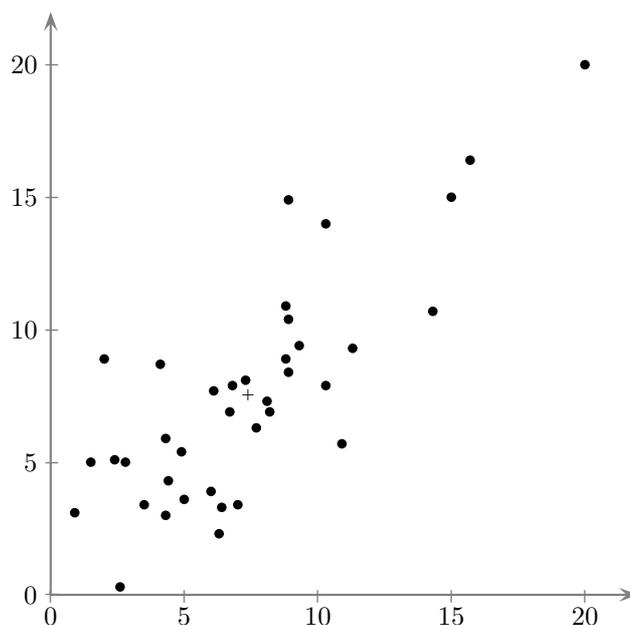


FIGURE 7.2 – Nuage de points

Comme pour les séries statistiques simples, on peut regrouper les valeurs de la série statistique en classes, pour chacun des caractères. Pour simplification, on prend dans l'exemple considéré des classes assez grandes, en groupant par tranches de 4 points. Dans un tableau à deux entrées, on indique alors le nombre (ou la fréquence) d'individus de l'échantillon considéré entrant dans les deux classes pour chacun des caractères  $A$  et  $B$ .

Notons dans la suite  $A_1, \dots, A_k$  les classes du caractère  $A$ , et  $B_1, \dots, B_\ell$  les classes pour le caractère  $B$ . On note alors  $n_{i,j}$  le nombre d'individus rentrant dans la classe  $A_i$  et dans la classe  $B_j$ , et  $f_{i,j}$  la fréquence correspondante.

Graphiquement cela revient à quadriller le plan  $\mathbb{R}^2$ , et à relever le nombre de points du nuage dans chaque carré (ou rectangle) du quadrillage. Voir figure 7.3

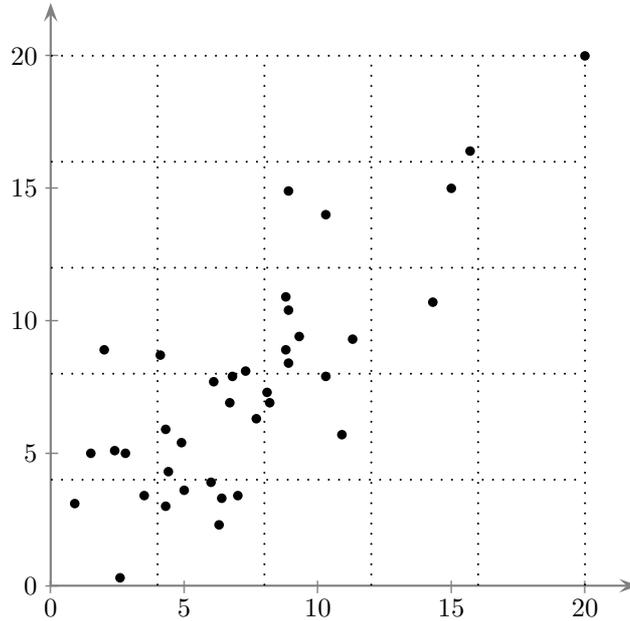


FIGURE 7.3 – Répartition en classes

On obtient le tableau des effectifs et le tableau des fréquences de la figure 7.4

$B \setminus A$	$[0,4]$	$]4,8]$	$]8,12]$	$]12,16]$	$]16,20]$
$[0,4]$	3	6	0	0	0
$]4,8]$	3	7	4	0	0
$]8,12]$	1	2	6	1	0
$]12,16]$	0	0	2	1	0
$]16,20]$	0	0	0	1	1
$B \setminus A$	$[0,4]$	$]4,8]$	$]8,12]$	$]12,16]$	$]16,20]$
$[0,4]$	0.080	0.158	0	0	0
$]4,8]$	0.080	0.184	0.105	0	0
$]8,12]$	0.026	0.053	0.158	0.026	0
$]12,16]$	0	0	0.053	0.026	0
$]16,20]$	0	0	0	0.026	0.026

FIGURE 7.4 – Tableau des effectifs et des fréquences par classes de notes

### 7.4.2 Fréquences marginales, fréquences conditionnelles

DÉFINITION 7.4.5 • Les fréquences marginales des classes  $A_i$  du caractère  $A$  sont définies par :

$$f_{i,\bullet} = \sum_{j=1}^{\ell} f_{i,j}.$$

Il s'agit donc de la somme des fréquences de la  $i$ -ième colonne.

- Les fréquences marginales des classes  $B_i$  du caractère  $B$  sont définies par :

$$f_{\bullet,j} = \sum_{i=1}^{\ell} f_{i,j}.$$

Il s'agit donc de la somme des fréquences de la  $j$ -ième ligne.

Tout comme leur analogue pour les couples de variables aléatoires, les fréquences marginales correspondent aux fréquences obtenues pour l'étude d'un seul des deux caractères (en oubliant les résultats obtenus pour l'autre)

On obtient par exemple, pour l'exemple précédent, les tableaux de la figure 7.5. Dans cette figure, la dernière ligne indique les fréquences marginales pour le caractère  $A$ , dans la répartition par tranches de 4 points. La dernière colonne indique les fréquences marginales pour le caractère  $B$ . La dernière case indique l'effectif ou la fréquence totale.

$B \setminus A$	[0,4]	]4,8]	]8,12]	]12,16]	]16,20]	
[0,4]	3	6	0	0	0	9
]4,8]	3	7	4	0	0	14
]8,12]	1	2	6	1	0	10
]12,16]	0	0	2	1	0	3
]16,20]	0	0	0	1	1	2
	7	15	12	3	1	38

$B \setminus A$	[0,4]	]4,8]	]8,12]	]12,16]	]16,20]	
[0,4]	0.080	0.158	0	0	0	0.237
]4,8]	0.080	0.184	0.105	0	0	0.368
]8,12]	0.026	0.053	0.158	0.026	0	0.263
]12,16]	0	0	0.053	0.026	0	0.080
]16,20]	0	0	0	0.026	0.026	0.053
	0.184	0.395	0.317	0.080	0.026	1

FIGURE 7.5 – Tableau des effectifs et des fréquences marginales par classes de notes

Évidemment, une telle étude peut se faire sans grouper les valeurs en classes, si le nombre de valeurs possibles pour chaque caractère est assez petit.

**DÉFINITION 7.4.6** Soit  $(i, j) \in \llbracket 1, k \rrbracket \times \llbracket 1, \ell \rrbracket$ . On appelle fréquence conditionnelle de la classe  $A_i$  sachant  $B_j$ , fréquence de la classe  $A_i$  conditionnée à  $B_j$ , la quantité  $\frac{f_{i,j}}{f_{\bullet,j}}$ . C'est donc, dans le sous-ensemble de l'échantillon constitué des individus tels que le caractère  $B$  se trouve dans la classe  $B_j$  la proportion de ceux pour lesquels le caractère  $A$  se trouve dans la classe  $A_i$ .

On définit de même la fréquence conditionnelle de la classe  $B_j$  sachant  $A_i$ , par  $\frac{f_{i,j}}{f_{i,\bullet}}$ .

Dans le premier cas, la somme sur chaque ligne est égale à 1 (fréquence de l'effectif total de  $B_j$  dans lui-même). La somme sur les colonnes n'a pas de sens.

Dans le deuxième cas, la somme sur chaque colonne est égale à 1. La somme sur les lignes n'a pas de sens.

On donne les fréquences conditionnées au caractère  $B$  de l'exemple précédent dans la figure 7.6 ; les fréquences conditionnées au caractère  $A$  dans la figure 7.7

$B \setminus A$	[0,4]	]4,8]	]8,12]	]12,16]	]16,20]	
[0,4]	0.333	0.667	0	0	0	1
]4,8]	0.214	0.5	0.286	0	0	1
]8,12]	0.1	0.2	0.6	0.1	0	1
]12,16]	0	0	0.667	0.333	0	1
]16,20]	0	0	0	0.5	0.5	1

FIGURE 7.6 – Tableau des fréquences conditionnées au caractère  $B$ 

$B \setminus A$	[0,4]	]4,8]	]8,12]	]12,16]	]16,20]	
[0,4]	0.429	0.4	0	0	0	
]4,8]	0.429	0.467	0.333	0	0	
]8,12]	0.143	0.133	0.5	0.333	0	
]12,16]	0	0	0.167	0.333	0	
]16,20]	0	0	0	0.333	1	
	1	1	1	1	1	

FIGURE 7.7 – Tableau des fréquences conditionnées au caractère  $A$ 

### 7.4.3 Paramètres associés à une série statistique bivariée

DÉFINITION 7.4.7 Soit  $\bar{x}$  la moyenne de la série statistique  $(x_i)_{i \in \llbracket 1, n \rrbracket}$ , et  $\bar{y}$  celle de  $(y_i)_{i \in \llbracket 1, n \rrbracket}$ , autrement dit :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{et} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

On définit le point moyen du nuage de points  $(x_i, y_i)_{i \in \llbracket 1, n \rrbracket}$  comme étant le point  $(\overline{lx}, \overline{ly})$ .

Un calcul rapide (avec un outil informatique par exemple, du type tableur) donne, dans l'exemple précédent, le point moyen  $(7.39, 7.57)$  représenté par une croix (+) sur la figure 7.2.

DÉFINITION 7.4.8 On définit  $\sigma_x$  et  $\sigma_y$  les écarts-types empiriques (positifs) des séries marginales  $(x_i)_{i \in \llbracket 1, n \rrbracket}$  et  $(y_i)_{i \in \llbracket 1, n \rrbracket}$ , comme pour les séries statistiques simples, à savoir :

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{et} \quad \sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2.$$

Les quantités  $\sigma_x^2$  et  $\sigma_y^2$  sont les variances empiriques des séries marginales.

La formule de König-Huygens s'applique bien entendu pour les variances marginales :

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left( \frac{1}{n} \sum_{i=1}^n x_i \right)^2 \quad \text{et} \quad \sigma_y^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - \left( \frac{1}{n} \sum_{i=1}^n y_i \right)^2$$

DÉFINITION 7.4.9 La covariance empirique de la série statistique  $(x_i, y_i)_{i \in \llbracket 1, n \rrbracket}$  est le réel défini par :

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n x_k y_k - \bar{x} \bar{y}.$$

D'après la relation de König-Huygens, on obtient immédiatement :

$$\sigma_x^2 = \text{cov}(x, x)$$

PROPOSITION 7.4.10 On a :  $\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_k - \bar{x})(y_k - \bar{y})$ .

◁ *Démonstration.*

Immédiat en développant le produit et en utilisant la définition de la moyenne. ▷

DÉFINITION 7.4.11 On définit le coefficient de corrélation, généralement noté  $r_{x,y}$  ou  $\rho_{x,y}$ , par :

$$\rho_{x,y} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}.$$

THÉORÈME 7.4.12 On a toujours  $|\rho_{x,y}| \leq 1$ . De plus,  $|\rho_{x,y}| = 1$  si et seulement s'il existe  $\lambda$  tel que pour tout  $i \in \llbracket 1, n \rrbracket$ , on ait  $y_i = \lambda x_i$  (il existe donc une relation affine entre  $x$  et  $y$ )

◁ *Démonstration.*

D'après l'inégalité de Cauchy-Schwarz numérique,

$$\sum_{k=1}^n (x_k - \bar{x})^2 \sum_{k=1}^n (y_k - \bar{y})^2 \geq \left( \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y}) \right)^2.$$

Ainsi, en divisant par  $n^2$ , et en prenant la racine, il vient

$$|\text{cov}(x, y)| \leq \sigma_x \sigma_y,$$

d'où  $|\rho_{x,y}| \leq 1$ .

De plus, le cas d'égalité se produit si et seulement si les vecteurs  $\begin{pmatrix} x_1 \\ \vdots \\ x_i \end{pmatrix}$  et  $\begin{pmatrix} y_1 \\ \vdots \\ y_i \end{pmatrix}$  sont colinéaires, d'où l'existence du coefficient  $\lambda$ . ▷

#### 7.4.4 Droites de régression (ou droite des moindres carrés)

Dans l'exemple traité dans ce paragraphe, on voit (et on a expliqué pourquoi on s'y attendait...) que le nuage de point est concentré autour d'une droite particulière, dont l'équation est  $y = x$ . Dans le cas particulier où les nuages de points ont une direction de concentration comme ici, on cherche à définir de manière canonique une droite qui représenterait au mieux le nuage de point, donc qui « approcherait » au mieux le nuage de point, suivant un certain critère d'approximation. On définit souvent cette meilleure approximation comme suit :

DÉFINITION 7.4.13 La droite des moindres carrés de  $y$  en  $x$  de la série (ou du nuage), aussi appelée droite de régression de  $y$  en  $x$ , est la droite d'équation  $y = m_x x + p$ , où  $m_x$  et  $p$  sont tels que la quantité

$$\sum_{k=1}^n (y_k - m_x x_k - p)^2$$

soit minimum.

Il s'agit donc de la droite minimisant l'écart quadratique des ordonnées à celles de la droite (à abscisse fixée).

DÉFINITION 7.4.14 De même, la droite des moindres carrés de  $x$  en  $y$  de la série (ou du nuage), aussi appelée droite de régression de  $x$  en  $y$ , est la droite d'équation  $x = m_y y + p$ , où  $m_y$  et  $p$  sont tels que la quantité

$$\sum_{k=1}^n (x_k - m_x y_k - p)^2$$

soit minimum.

Il s'agit donc de la droite minimisant l'écart quadratique des abscisses à celles de la droite (à ordonnée fixée).

**Méthode 1 :** Recherche des extremums d'une fonction de plusieurs variables.

Cela nous ramène à la résolution du système

$$\begin{cases} \sum_{k=1}^n (y_k - m_x x_k - p) = \bar{y} - m_x \bar{x} - p = 0 \\ \sum_{k=1}^n x_k (y_k - m_x x_k - p) = 0 \end{cases}$$

ce qui nous amène à :

THÉORÈME 7.4.15 La droite de régression de  $y$  en  $x$  est la droite d'équation  $y = m_x x + p$ , de coefficient directeur

$$m_x = \frac{\sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})}{\sum_{k=1}^n (x_k - \bar{x})^2} = \frac{\text{COV}(x, y)}{\sigma_x^2} = \rho_{x, y} \frac{\sigma_y}{\sigma_x},$$

et passant par le point moyen du nuage.

**Méthode 2 :** Méthode algébrique des moindres carrés.

Avec les notations précédentes, et en posant  $U$  le vecteur colonne constitué uniquement de 1, on veut minimiser la quantité  $\|\mathbf{Y} - m_x \mathbf{X} - pU\|$ , donc la quantité  $\left\| \mathbf{Y} - A \begin{pmatrix} m_x \\ p \end{pmatrix} \right\|$ , où  $A$  est la matrice

$$A = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}$$

Ainsi, cette quantité est minimum lorsque  $A \begin{pmatrix} m_x \\ p \end{pmatrix}$  est égal au projeté orthogonal sur  $\text{Im } A$  du vecteur  $Y$ . Or,  $\text{Im } A$  est l'espace engendré par  $A$  et  $U$ . Un résultat du cours d'algèbre bilinéaire nous affirme qu'on a alors

$${}^t A A \begin{pmatrix} m_x \\ p \end{pmatrix} = {}^t A \mathbf{Y}$$

Cela fournit un système de deux équations à deux inconnues, et on trouve :

THÉORÈME 7.4.16 *La droite de régression de  $y$  en  $x$  est la droite d'équation  $y = m_x x + p_x$ , avec*

$$m_x = \frac{\text{cov}(x, y)}{\sigma_x^2} = \rho_{x,y} \frac{\sigma_y}{\sigma_x}, \quad \text{et} \quad p_x = \bar{y} - m_x \bar{x}.$$

*Autrement dit, l'équation se réécrit :  $y - \bar{y} = \rho_{x,y} \frac{\sigma_y}{\sigma_x} (x - \bar{x}) = \frac{\text{cov}(x, y)}{\sigma_x^2} (x - \bar{x})$ .*

THÉORÈME 7.4.17 *La droite de régression de  $x$  en  $y$  est la droite d'équation  $x = m_y y + p_y$ , avec*

$$m_y = \frac{\text{cov}(x, y)}{\sigma_y^2} = \rho_{x,y} \frac{\sigma_x}{\sigma_y}, \quad \text{et} \quad p_y = \bar{x} - m_y \bar{y}$$

*Autrement dit, l'équation se réécrit :  $x - \bar{x} = \rho_{x,y} \frac{\sigma_x}{\sigma_y} (y - \bar{y}) = \frac{\text{cov}(x, y)}{\sigma_y^2} (y - \bar{y})$ .*

Dans l'exemple qui nous suit, nous trouvons :

$$\sigma_x = 4.16, \quad \sigma_y = 4.26 \quad \text{cov}(x, y) = 13.76 \quad \rho_{x,y} = 0.776.$$

On obtient donc :

$$m_x = 0.795 \quad p_x = 1.695 \quad m_y = 0.758 \quad p_y = 1.65.$$

Ainsi, la droite de régression linéaire de  $y$  en  $x$  est la droite d'équation  $y = 0.795x + 1.695$ , et la droite de régression linéaire de  $x$  en  $y$  est la droite d'équation  $x = 0.758y + 1.65$ , donc  $y = 1.319x - 2.177$ .

On a représenté ces deux droites sur le nuage de points, dans la figure 7.8

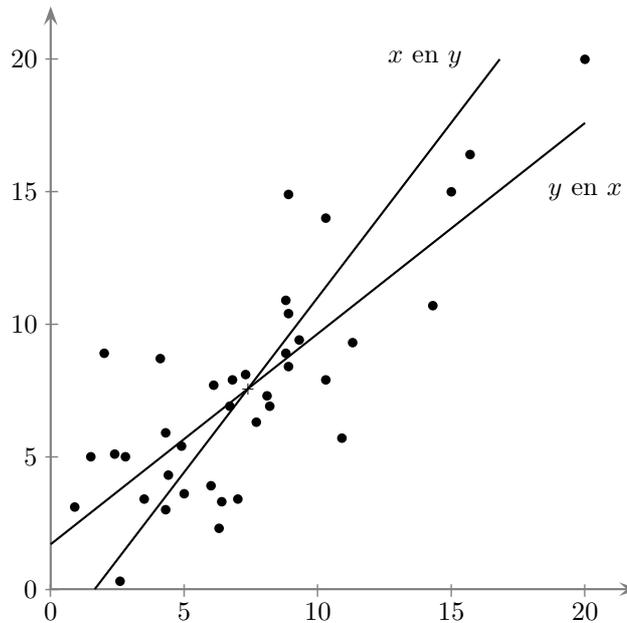


FIGURE 7.8 – Droites de régression