

Probabilités et statistiques

Pierre-Damien Olive

Mis à jour le 2019-10-17

Contents

Introduction

Le site avec la dernière version du cours est disponible sur : <https://olivepierre.github.io/probaStats/>

Part I

Probabilités

Chapter 1

Évènements et probabilités

1.1 Évènements

1.2 Probabilités

1.3 Probabilités conditionnelles

1.4 Indépendance

Chapter 2

Variables aléatoires

2.1 Variables aléatoires

2.2 Variables discrètes et continues

2.3 Vecteurs aléatoires

Chapter 3

Variables aléatoires discrètes

3.1 Fonctions de masse

Définition 3.1. La **fonction de masse** d'une variable aléatoire discrète X est la fonction $f : \mathbb{R} \mapsto [0, 1]$ définie par:

$$f(x) = \mathbb{P}(X = x).$$

Fonctions de répartition et fonctions de masse sont reliées par les relations:

$$F(x) = \sum_{x_i \leq x} f(x_i); f(x) = F(x) - F(x^-).$$

Lemme 3.1. Une fonction $f : \mathbb{R} \mapsto [0, 1]$ est une fonction de masse si et seulement si l'ensemble $\{x : f(x) > 0\}$ est dénombrable et $\sum_i f(x_i) = 1$, où les x_1, \dots, x_n, \dots sont les valeurs de x tel que $f(x) > 0$.

Exemple 3.1. On choisit un nombre entier entre 1 et n , de façon équiprobable. Un tel choix définit une variable aléatoire discrète, de loi uniforme.

Exemple 3.2. On lance une pièce n fois, et on tombe sur pile avec une probabilité p . L'ensemble des événements est $\Omega = \{\text{pile}, \text{face}\}$. Le nombre de fois que l'on obtient "pile" est une variable aléatoire discrète dont la fonction de masse vérifie:

$$f(x) = \begin{cases} 0 & \text{si } x \notin \{0, \dots, n\} \\ \binom{n}{k} p^k (1-p)^{n-k} & \text{sinon} \end{cases}$$

On dit que cette variable aléatoire suit une loi binômiale (notée $\mathcal{B}(n, p)$).

Lorsque $n = 1$, on dit que la variable aléatoire suit une loi de Bernoulli (notée $\mathcal{B}(p)$).

Exemple 3.3. On compte le nombre d'épreuves de Bernoulli nécessaires pour arriver sur "pile" pour la première fois. Ce nombre de lancer est une variables aléatoire discrète, à valeur dans \mathbb{N}^* et de fonction de masse:

$$f(k) = (1-p)^{k-1}p \forall k \in \mathbb{N}^*$$

La loi correspondante s'appelle la loi géométrique, de paramètre p , notée $\mathcal{G}(p)$.

Exemple 3.4. Une variable aléatoire prenant ses valeurs dans \mathbb{N} avec une fonction de masse

$$f(k) = \frac{\lambda^k}{k!} e^{-\lambda} \forall k \in \mathbb{N}$$

suit une loi de Poisson de paramètre $\lambda > 0$, notée $\mathcal{P}(\lambda)$.

Exemple 3.5. Soit $0 < p < 1$. On tire simultanément n boules dans une urne contenant A boules, dont pA gagnantes et $(1-p)A$ perdantes. On compte le nombre de boules gagnantes extraites. La loi de cette variable aléatoire discrète est la loi hypergéométrique, de fonction de masse:

$$f(k) = \frac{\binom{pA}{k} \binom{qA}{n-k}}{\binom{A}{n}}$$

3.2 Indépendance

Définition 3.2. Deux variables aléatoires discrètes X et Y sont **indépendantes** si les évènements $\{X = x\}$ et $\{Y = y\}$ sont indépendants pour tout x et y :

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

On peut étendre cette définition à un ensemble de variables aléatoires $(X_i)_{i \in I}$. Dans ce cas, l'égalité ci-dessus devient:

$$\mathbb{P}(X_i = x_i \forall i \in J) = \prod_{i \in J} \mathbb{P}(X_i = x_i)$$

pour toute partie finie J de I

Théorème 3.1. Soient X, Y deux variables aléatoires discrètes indépendantes, f, g deux fonctions réelles. Alors les variables aléatoires discrètes $f(X)$ et $g(Y)$ sont encore indépendantes.

3.3 Espérance

Définition 3.3. L'**espérance** d'une variables aléatoire discrète X avec une fonction de masse f est définie comme

$$\mathbb{E}X = \sum_{x: f(x) > 0} x f(x)$$

lorsque cette somme est *absolument* convergente; la convergence absolue est requise afin que cette somme ait une même valeur quelque soit l'ordre dans lequel on somme les x (cf. cours analyse sur les familles sommables).

En particulier, lorsque A est un évènement, on a

$$\mathbb{E}\mathbf{1}_A = \mathbb{P}(A).$$

Afin de calculer les espérances de fonctions de variables aléatoires, on peut envisager de calculer la fonction de masse de cette nouvelle variable aléatoire. Cette approche est délicate et peu pertinente grâce au lemme:

Lemme 3.2. *Soit X une variable aléatoire discrète, f sa fonction de masse. Soit g une fonction réelle. Alors l'espérance de $g(X)$ est définie par*

$$\mathbb{E}g(X) = \sum_x g(x)f(x)$$

lorsque cette somme est absolument convergente.

On peut ainsi calculer les moments pour tout ordre ($m_k = \mathbb{E}X^k$) ainsi que les moments centrés ($\sigma_k = \mathbb{E}(X - m_1)^k$). En particulier, on définit :

Définition 3.4. Le **moment d'ordre 1** s'appelle l'espérance de X , le **moment centré d'ordre 2** de X s'appelle la variance de X (notée $\mathbb{V}X$) : c'est une quantité positive! On appelle $\sigma = \sqrt{\sigma_2}$ l'écart-type de X .

Lemme 3.3. *On peut exprimer les moments centrés en fonction de moments d'ordres inférieurs ; en particulier,*

$$\sigma_2 = m_2 - m_1^2$$

Théorème 3.2. *L'opérateur \mathbb{E} est une forme linéaire (pour $\lambda \in \mathbb{R}$, X, Y deux variables aléatoires admettant une espérance, $\mathbb{E}(X + \lambda Y) = \mathbb{E}X + \lambda \mathbb{E}Y$), positive ($X \geq 0$ entraîne $\mathbb{E}X \geq 0$) vérifiant $\mathbb{E}\mathbf{1} = 1$.*

Par contre, il n'est pas vrai (en général) que

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

Lemme 3.4. *Si X, Y sont des variables aléatoires indépendantes, alors*

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

Définition 3.5. On dit que les variables aléatoires X, Y sont **non corrélées** si

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

Ainsi, les variables indépendantes sont non-corrélées, mais l'inverse n'est pas vrai !

Théorème 3.3. *Soient X, Y deux variables aléatoires, $\lambda \in \mathbb{R}$. Alors $\mathbb{V}\lambda X = \lambda^2 \mathbb{V}X$ et $\mathbb{V}(X + Y) = \mathbb{V}X + \mathbb{V}Y$ si ces variables sont non-corrélées.*

3.4 Cas particuliers de variables discrètes aléatoires

Théorème 3.4. *Soit X une variable aléatoire suivant une loi uniforme. Alors*

$$\mathbb{E}X = \frac{n+1}{2}, \mathbb{V}X = \frac{n^2-1}{12}$$

Théorème 3.5. *Soit X une variable aléatoire suivant une loi de Bernoulli. Alors*

$$\mathbb{E}X = p, \mathbb{V}X = p(1-p)$$

Théorème 3.6. *Soit X une variable aléatoire suivant une loi binômiale $\mathcal{B}(n, p)$. Alors*

$$\mathbb{E}X = np, \mathbb{V}X = np(1-p)$$

Théorème 3.7. *Soit X une variable aléatoire suivant une loi géométrique $\mathcal{G}(p)$. Alors*

$$\mathbb{E}X = \frac{1}{p}, \mathbb{V}X = \frac{1-p}{p^2}$$

Théorème 3.8. *Soit X une variable aléatoire suivant une loi de Poisson $\mathcal{P}(\lambda)$. Alors*

$$\mathbb{E}X = \lambda, \mathbb{V}X = \lambda$$

3.5 Dépendance

Définition 3.6. Soient X, Y deux variables aléatoires discrètes. La **fonction de répartition jointe** est définie par

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$

La **fonction de masse jointe** est donnée par

$$f(x, y) = \mathbb{P}(X = x, Y = y).$$

Lemme 3.5. *Les variables aléatoires discrètes X, Y sont indépendantes si et seulement si*

$$f_{X,Y}(x, y) = f_X(x)f_Y(y).$$

Plus généralement, X, Y sont indépendantes si et seulement si il existe des fonctions réelles f, g telles que

$$f_{X,Y}(x, y) = f(x)g(y).$$

A partir de la connaissance de $f_{X,Y}$, il est possible de calculer f_X ou f_Y grâce au théorème suivant :

Théorème 3.9. On a :

$$f_X(x) = \sum_y f_{X,Y}(x, y)$$

$$f_Y(y) = \sum_x f_{X,Y}(x, y)$$

Le calcul de l'espérance d'une fonction de variables aléatoires est rendu facile par le

Lemme 3.6. Soit g une fonction telle que $g(X, Y)$ soit une variable aléatoire. Alors

$$\mathbb{E}g(X, Y) = \sum_{x,y} g(x, y) f_{X,Y}(x, y)$$

si cette somme existe.

Définition 3.7. Soient X, Y deux variables aléatoires. Leur **covariance** est définie par :

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y.$$

Le **coefficient de corrélation** est :

$$\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\mathbb{V}X\mathbb{V}Y}}.$$

Lemme 3.7. Le coefficient de corrélation est en module inférieur à 1 et égal à ± 1 si et seulement si il existe des réels a, b, c tels que

$$\mathbb{P}(aX + bY + c = 0) = 1.$$

Théorème 3.10. Soit X, Y deux variables aléatoires discrètes. Alors

$$(\mathbb{E}XY)^2 \leq (\mathbb{E}X^2)(\mathbb{E}Y^2)$$

avec égalité si et seulement si il existe des réels λ, μ non tous nuls tels que

$$\mathbb{P}(\lambda X + \mu Y = 0) = 1.$$

3.6 Loi conditionnelle et espérance conditionnelle

Soient X, Y deux variables aléatoires discrètes.

Définition 3.8. La **fonction de répartition conditionnelle** de Y sachant $X = x$, notée $F_{Y|X}(\bullet|x)$ est définie par

$$F_{Y|X}(y|x) = \mathbb{P}(Y \leq y | X = x)$$

pour tout x tel que $\mathbb{P}(X = x) > 0$.

La **fonction de masse conditionnelle** de Y sachant $X = x$, notée $f_{Y|X}(\bullet|x)$ est définie par

$$f_{Y|X}(y|x) = \mathbb{P}(Y = y|X = x)$$

pour tout x tel que $\mathbb{P}(X = x) > 0$.

On a donc

$$f_{Y|X} = f_{X,Y}/f_X$$

lorsque cette expression est définie. Encore, $f_{Y|X} = f_Y$ si et seulement si X et Y sont indépendantes.

Définition 3.9. L'espérance conditionnelle de Y par rapport à X est définie comme

$$\mathbb{E}(Y|X = x) = \sum_y y f_{Y|X}(y|x)$$

Théorème 3.11. On a :

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

Théorème 3.12. Pour toute fonction g pour laquelle les sommes sont définies, on a :

$$\mathbb{E}(Yg(X)) = \mathbb{E}(g(X)\mathbb{E}(Y|X))$$

Définition 3.10. La **variance conditionnelle** de Y sachant X est définie par

$$\mathbb{V}(Y|X) = \mathbb{E}([Y - \mathbb{E}(Y|X)]^2|X).$$

La variance conditionnelle est liée à l'espérance conditionnelle par le théorème :

Théorème 3.13.

$$\mathbb{V}(Y) = \mathbb{E}(\mathbb{V}[Y|X]) + \mathbb{V}(\mathbb{E}[Y|X])$$

3.7 Somme de variables aléatoires discrètes

Soient X, Y deux variables aléatoires discrètes.

Théorème 3.14. La loi de $X + Y$ est définie par

$$\mathbb{P}(X + Y = z) = \sum_x f_{X,Y}(x, z - x)$$

Si, en outre, X et Y sont indépendantes, alors

$$\mathbb{P}(X + Y = z) = \sum_x f_X(x)f_Y(z - x) = \sum_y f_X(z - y)f_Y(y)$$

La fonction masse de $X + Y$ s'appelle alors le **produit de convolution** des fonctions masses de X et Y , et est noté $f_X \star f_Y$.

Chapter 4

Variables aléatoires continues

4.1 densités de probabilités

Définition 4.1. Une variable aléatoire réelle X est dite **continue** s'il existe une fonction f positive telle que

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

Une telle fonction est une **densité** de X . Cette densité n'est pas unique. Une telle densité est généralement notée f_X .

Lorsque F_X est dérivable, F'_X est une densité de X .

Lemme 4.1. *Si X admet f comme densité, alors:*

- $\int_{-\infty}^{+\infty} f(x)dx = 1$;
- $\mathbb{P}(X = x) = 0, \forall x \in \mathbb{R}$;
- $\mathbb{P}(X \in \mathcal{E}) = \int_{\mathcal{E}} f(x)dx$ pour tout événement \mathcal{E} ;
- X et $-X$ ont les mêmes distributions si et seulement si f_X est paire;
- Si f et g sont des densités, il en est de même pour $\lambda f + (1 - \lambda)g$ pour $0 \leq \lambda \leq 1$.

Exemple 4.1. Soient a, b deux réels avec $a < b$. La **loi uniforme** sur l'intervalle $[a, b]$ est définie par sa densité de probabilité

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{sinon} \end{cases}$$

On note cette loi $\mathcal{U}(a, b)$.

Exemple 4.2. La **loi exponentielle** modélise la durée de vie d'un phénomène sans mémoire : la probabilité que le phénomène dure au moins $s + t$ sachant qu'il a déjà duré t est la même que la probabilité de durer s à partir de la mise en fonction initiale.

Soit $\lambda > 0$. La densité associée à cette loi est :

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{si } t > 0 \\ 0 & \text{sinon} \end{cases}$$

On note la loi exponentielle de paramètre $\lambda : \mathcal{E}(\lambda)$.

Lemme 4.2. Soient X, Y deux variables aléatoires qui suivent deux lois exponentielles, de paramètres λ, μ , alors $\min(X, Y)$ suit encore une loi exponentielle, de paramètre $\lambda + \mu$.

Exemple 4.3. La loi **Gamma** (qui généralise la loi exponentielle) modélise également des durées; elle est caractérisée par deux paramètres k, θ positifs.

Sa densité est

$$f(t) = \begin{cases} \frac{t^{k-1} \exp(-\frac{t}{\theta})}{\theta^k \Gamma(k)} & \text{si } t > 0 \\ 0 & \text{sinon} \end{cases}$$

où Γ désigne la fonction Gamma d'Euler ($\Gamma : z \mapsto \int_0^{+\infty} t^{z-1} e^{-t} dt$).

On note cette loi $\Gamma(k, \theta)$.

Lemme 4.3. Soient X_1, \dots, X_n des variables aléatoires indépendantes, de loi $\Gamma(k_1, \theta), \dots, \Gamma(k_n, \theta)$. Alors $X_1 + \dots + X_n$ suit une loi $\Gamma(\sum_{i=1}^n k_i, \theta)$.

Lemme 4.4. Soit $X \sim \Gamma(k, \theta)$ et $t > 0$. Alors $tX \sim \Gamma(k, t\theta)$.

Exemple 4.4. La **loi normale** est définie suivant deux paramètres ($\mu \in \mathbb{R}, \sigma > 0$) par sa densité:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

On note la loi $\mathcal{N}(\mu, \sigma^2)$.

Exemple 4.5. Soit X une variable aléatoire qui suit une loi normale $\mathcal{N}(\mu, \sigma^2)$. Alors $\exp(X)$ suit une loi **log-normale**.

4.2 Indépendance

Définition 4.2. Soient X, Y deux variables aléatoires réelles. On dit que les variables aléatoires X et Y sont **indépendantes** si les événements $\{X \leq x\}$ et $\{Y \leq y\}$ sont indépendants pour tous x, y réels.

Théorème 4.1. *Soient X, Y deux variables aléatoires, f, g deux fonctions telles que $f(X), g(Y)$ soient des variables aléatoires.*

Si X et Y sont indépendantes, il en est de même pour $f(X)$ et $g(Y)$.“

4.3 Espérance

Définition 4.3. L'espérance d'une variable aléatoire X de densité f est définie par

$$\mathbb{E}X = \int_{-\infty}^{+\infty} xf(x)dx$$

lorsque cette intégrale existe.

Théorème 4.2. *Si X et $g(X)$ sont des variables aléatoires, alors*

$$\mathbb{E}g(X) = \int_{-\infty}^{+\infty} g(x)f(x)dx$$

Lemme 4.5. *Si X admet une densité nulle sur \mathbb{R}^- et une fonction de répartition F , alors*

$$\mathbb{E}X = \int_0^{+\infty} (1 - F(x))dx = \int_0^{+\infty} \mathbb{P}(X > x)dx$$

Les autres propriétés de l'espérance, vues dans le chapitre sur les variables aléatoires discrètes, s'étendent sans difficultés aux variables continues (moments, moments centrés, variance, covariance, corrélation).

4.4 Exemples de variables aléatoires continues

Théorème 4.3. *L'espérance et la variance des lois continues de référence sont recensés dans le tableau ci-après :*

Loi	Espérance	Variance
$\mathcal{U}(a, b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\mathcal{E}(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\Gamma(k, \theta)$	$k\theta$	$k\theta^2$
$\mathcal{N}(\mu, \sigma^2)$	μ	σ^2

4.5 Dépendance

Définition 4.4. Soient X, Y deux variables aléatoires réelles. Ces variables sont **conjointement continues** s'il existe une fonction positive f telle que:

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv$$

f est une densité du couple (X, Y) . Si F est dérivable, on peut prendre

$$f = \frac{\partial^2 F}{\partial x \partial y}$$

Lemme 4.6. *On a*

$$\mathbb{P}(a \leq X \leq b; c \leq Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c) = \int_a^b \int_c^d f(u, v) du dv$$

Plus généralement, si \mathcal{E} est un évènement, on a :

$$\mathbb{P}((X, Y) \in \mathcal{E}) = \iint_{\mathcal{E}} f(u, v) du dv$$

Définition 4.5. Les **fonctions de répartition marginales** de X et Y sont définies par

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x \int_{-\infty}^{+\infty} f(u, v) du dv; F_Y(y) = \mathbb{P}(Y \leq y) = \int_{-\infty}^{+\infty} \int_{-\infty}^y f(u, v) du dv$$

On en déduit les **densités marginales** selon X et Y :

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy; f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

Théorème 4.4. Soit $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ telle que $g(X, Y)$ soit une variable aléatoire. Alors,

$$\mathbb{E}g(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(u, v) f(u, v) du dv$$

dès que cette intégrale est définie.

Théorème 4.5. Les variables X, Y sont indépendantes si et seulement si

$$F_{X,Y} = F_X F_Y \iff f_{X,Y} = f_X f_Y.$$

Théorème 4.6. Soit X, Y deux variables aléatoires continues. Alors

$$(\mathbb{E}XY)^2 \leq (\mathbb{E}X^2) (\mathbb{E}Y^2)$$

avec égalité si et seulement si il existe des réels λ, μ non tous nuls tels que

$$\mathbb{P}(\lambda X + \mu Y = 0) = 1.$$

4.6 Loi conditionnelle et espérance conditionnelle

Soient X, Y deux variables aléatoires continues, de densité jointe f .

Définition 4.6. La **fonction de répartition conditionnelle** de Y sachant $X = x$ est définie par

$$F_{Y|X}(y|x) = \int_{-\infty}^y \frac{f(x, v)}{f_X(x)} dv$$

pour tout x tel que $f_X(x) > 0$. Elle est également notée

$$\mathbb{P}(Y \leq y | X = x)$$

Définition 4.7. La **densité conditionnelle** de Y sachant $X = x$ est définie par

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}$$

pour tout x tel que $f_X(x) > 0$. Encore,

$$f_{Y|X} = \frac{f_{X,Y}}{f_X}$$

Théorème 4.7. *L'espérance conditionnelle de Y sachant X est définie comme*

$$\mathbb{E}(Y|X = x) = \int_{-\infty}^{+\infty} y f_{Y|X}(y|x) dy;$$

Elle vérifie la relation

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

Plus généralement, pour toute fonction g pour laquelle l'expression a un sens, on a :

$$\mathbb{E}(g(X)\mathbb{E}(Y|X)) = \mathbb{E}Yg(X).$$

4.7 Fonctions de variables aléatoires

Théorème 4.8. *Soit X une variable aléatoire avec une densité f , et soit g une fonction réelle telle que $Y = g(X)$ soit encore une variable aléatoire. La fonction de répartition de Y est définie par :*

$$\mathbb{P}(Y \leq y) = \mathbb{P}(X \in g^{-1}(]-\infty, y])).$$

Plus généralement, on a le théorème suivant:

Théorème 4.9. Soient \mathcal{U}, \mathcal{V} deux ouverts de \mathbb{R}^d , Soit X une variable aléatoire à valeurs dans \mathcal{U} de densité f et soit T un difféomorphisme de classe $\mathcal{C}^1 : \mathcal{U} \rightarrow \mathcal{V}$. La densité de la variable aléatoire $Y = \varphi(X)$ est définie par :

$$f_Y(y) = \begin{cases} f_X(\varphi^{-1}(y)) |J_{\varphi^{-1}}(y)| & \text{si } y \in \mathcal{V} \\ 0 & \text{sinon} \end{cases}$$

en notant $|J_{\varphi^{-1}}(y)|$ le déterminant du jacobien de φ^{-1} , c'est-à-dire de la matrice de terme général $\left(\frac{\partial \varphi_j^{-1}}{\partial y_i} \right)_{i,j}$.

Lemme 4.7. Pour $d = 2$, le changement de variables “passage en coordonnées polaires” transforme une densité $f(x, y)$ en $f(r \cos \theta, r \sin \theta)r$.

4.8 Somme de variables aléatoires

Théorème 4.10. Si X, Y ont pour densité jointe f , alors la densité de $Z = X + Y$ est définie par:

$$f_Z(z) = \int_{-\infty}^{+\infty} f(x, z - x) dx$$

Si, X et Y sont indépendantes, le résultat devient :

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z - y) f_Y(y) dy = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x) dx$$

On dit que f_Z est le **produit de convolution** de f_X et f_Y , noté $f_X \star f_Y$.

Lemme 4.8. Soit $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Supposons que X et Y sont indépendantes. Alors $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

4.9 La loi normale multivariée

Définition 4.8. Le vecteur $X = (X_1, \dots, X_n)$ suit une **loi normale multivariée** (notée $\mathcal{N}(\mu, \Sigma)$) s'il existe un vecteur $\mu = (\mu_1, \dots, \mu_n)$ de réels et une matrice symétrique définie positive Σ telle que la densité jointe de X soit:

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp \left(-\frac{1}{2} (x - \mu) \Sigma^{-1} (x - \mu)' \right)$$

De façon équivalente, on dit que le vecteur X suit une loi normale multivariée si et seulement si pour tout vecteur $a \in \mathbb{R}^n$, Xa' suit une loi normale.

Théorème 4.11. Si X suit une loi $\mathcal{N}(\mu, \Sigma)$, alors

- $\mathbb{E}X = \mu$

- la matrice Σ est la matrice de covariance de X :

$$\Sigma_{i,j} = \text{cov}(X_i, X_j)$$

- Si A est une matrice de rang $m \leq n$, alors XA suit la loi $\mathcal{N}(\mu A, A' \Sigma A)$

4.10 Lois issues de la loi normale

On dispose de variables aléatoires correspondants aux résultats d'un même expérience aléatoire X_1, \dots, X_n dont on suppose qu'elle sont de loi normale $\mathcal{N}(\mu, \sigma^2)$ pour des quantités μ, σ non observée. On cherche à estimer ces quantités. Pour cela, on va considérer la moyenne empirique

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

comme estimateur de μ et la variance empirique

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

comme estimateur de σ^2 .

Théorème 4.12. Soit X_1, \dots, X_n des variables aléatoires indépendantes, de même loi $\mathcal{N}(\mu, \sigma^2)$.

- \bar{X} et S^2 sont indépendants;
- $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$
- $(n-1)S^2/\sigma^2 \sim \Gamma(\frac{1}{2}, \frac{n-1}{2})$. Cette dernière loi s'appelle la **loi du χ^2 à $n-1$ degrés de liberté**, notée $\chi^2(n-1)$; c'est la loi de la somme des carrés de $n-1$ lois normales centrées réduites indépendantes.
- Les variables aléatoires

$$U = \frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$$

et

$$V = \frac{\sqrt{n}}{\sigma} (\bar{X} - \mu) \sim \mathcal{N}(0, 1).$$

Elles ne dépendent pas de σ . Le ratio

$$T = \frac{V}{\sqrt{U/(n-1)}}$$

dont le numérateur suit $\mathcal{N}(0, 1)$ et le dénominateur est la racine d'un $\chi^2(n-1)$, divisé par $n-1$ suit une **loi de Student**, de paramètre $n-1$, notée $t(n-1)$.

- Soient U, V deux variables aléatoires qui suivent respectivement les lois $\chi^2(r)$ et $\chi^2(s)$, alors

$$F = \frac{U/r}{V/s}$$

suit une **loi de Fisher** de paramètres (r, s) , notée $F(r, s)$. Dans ce cas, $1/F$ suit la loi $F(s, r)$ et si $T \sim t(r)$, $T^2 \sim F(1, r)$

Chapter 5

Convergence de variables aléatoires

5.1 Modes de convergence

Définition 5.1. Soient X, X_1, \dots, X_n des variables aléatoires. On dit que :

- $X_n \rightarrow X$ **presque sûrement** (noté $X_n \xrightarrow{p.s} X$) si

$$\mathbb{P} \left(\left\{ \omega \in \Omega : X_n(\omega) \xrightarrow[n \rightarrow +\infty]{} X(\omega) \right\} \right) = 1$$

- $X_n \rightarrow X$ **dans L^p** ($p \geq 1$) (noté $X_n \xrightarrow{L^p} X$) si $\mathbb{E}|X_n|^p < +\infty$ et

$$\mathbb{E}|X_n - X|^p \xrightarrow[n \rightarrow +\infty]{} 0$$

- $X_n \rightarrow X$ **en probabilité** (noté $X_n \xrightarrow{\mathbb{P}} X$) si

$$\mathbb{P}(|X_n - X| > \epsilon) \xrightarrow[n \rightarrow +\infty]{} 0 \forall \epsilon > 0$$

- $X_n \rightarrow X$ **en loi**, ou **converge faiblement** (noté $X_n \xrightarrow{\mathcal{L}} X$) si

$$\mathbb{P}(X_n \leq x) \xrightarrow[n \rightarrow +\infty]{} \mathbb{P}(X \leq x) \text{ pour tous les } x \text{ où la fonction } F_X \text{ est continue.}$$

Notons que si $X_n \xrightarrow{\mathcal{L}} X$, $X_n \xrightarrow{\mathcal{L}} Y$ pour tout Y avec la même loi que X .

En particulier, on a les résultats suivants pour la convergence faible.

Théorème 5.1. Si $X_n \xrightarrow{\mathcal{L}} X$ et g est une fonction continue, alors $g(X_n) \xrightarrow{\mathcal{L}} g(X)$. Par ailleurs,

on a les équivalences :

- $X_n \xrightarrow{\mathcal{L}} X$
- $\mathbb{E}g(X_n) \rightarrow \mathbb{E}g(X)$ pour toute fonction g continue bornée
- $\mathbb{E}g(X_n) \rightarrow \mathbb{E}g(X)$ pour toute fonction g de la forme $g(x) = f(x)\mathbb{1}_{[a,b]}(x)$ où f est continue sur $[a,b]$ et a, b sont des points de continuité de la fonction de répartition de X .

Théorème 5.2. Si $X_n \xrightarrow{\mathbb{P}} X$ et g est une fonction continue, alors $g(X_n) \xrightarrow{\mathbb{P}} g(X)$

Théorème 5.3. Ces modes de convergence ne sont pas équivalents:

- $(X_n \xrightarrow{p.s} X) \Rightarrow (X_n \xrightarrow{\mathbb{P}} X)$
- $(X_n \xrightarrow{L^p} X) \Rightarrow (X_n \xrightarrow{\mathbb{P}} X)$
- $(X_n \xrightarrow{\mathbb{P}} X) \Rightarrow (X_n \xrightarrow{\mathcal{L}} X)$

Les implications inverses sont fausses en général.

De plus, si $q > p \geq 1$, alors

$$(X_n \xrightarrow{L^q} X) \Rightarrow (X_n \xrightarrow{L^p} X)$$

Théorème 5.4. On a :

- Si $X_n \xrightarrow{p.s} X$ et $Y_n \xrightarrow{p.s} Y$, alors $X_n + Y_n \xrightarrow{p.s} X + Y$
- Si $X_n \xrightarrow{L^p} X$ et $Y_n \xrightarrow{L^p} Y$, alors $X_n + Y_n \xrightarrow{L^p} X + Y$
- Si $X_n \xrightarrow{\mathbb{P}} X$ et $Y_n \xrightarrow{\mathbb{P}} Y$, alors $X_n + Y_n \xrightarrow{\mathbb{P}} X + Y$: Ce résultat est faux en général pour la convergence faible (prendre X telle que X et $-X$ ont la même loi, $X_n = Y_n$ et $Y = -X$).
- Si $X_n \xrightarrow{\mathbb{P}} X$ et $Y_n \xrightarrow{\mathbb{P}} Y$, alors $X_n Y_n \xrightarrow{\mathbb{P}} XY$

5.2 Inégalités en probabilités

Théorème 5.5. Soit X une variable aléatoire, h une fonction positive telle que $h(X)$ admette une espérance. Alors

$$\mathbb{P}(h(X) \geq a) \leq \frac{\mathbb{E}h(X)}{a}, \forall a > 0$$

Proof. Soit $A = \{h(X) \geq a\}$.

Alors $X \geq a\mathbf{1}_A$ et le résultat suit en passant à l'espérance. □

En particulier, on a:

Théorème 5.6 (Inégalité de Markov). *Soit X une variable aléatoire qui admet une moyenne. Alors*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}|X|}{a}, \forall a > 0$$

Théorème 5.7 (Inégalité de Bienaymé-Tchébychev). *Soit X une variable aléatoire qui admet une moyenne. Alors*

$$\mathbb{P}(|X|^2 \geq a) \leq \frac{\mathbb{E}|X|^2}{a}, \forall a > 0$$

Théorème 5.8. *Soit X une variable aléatoire qui admet un moment d'ordre 2. Alors*

$$\mathbb{P}(|X - \mathbb{E}X| \geq a) \leq \frac{\mathbb{V}X}{a^2}, \forall a > 0$$

Le théorème suivant permet de proposer une borne supérieure à $\mathbb{P}(|X| \geq a)$:

Théorème 5.9. *Si g est une fonction positive strictement croissante, alors*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}g(X)}{g(a)}, \forall a > 0$$

Les bornes inférieures sont plus difficiles à trouver en général, mais on a le

Théorème 5.10. *Soit $h : \mathbb{R} \rightarrow [0, M]$ une fonction positive à valeurs bornées. Alors*

$$\mathbb{P}(h(X) \geq a) \leq \frac{\mathbb{E}h(X) - a}{M - a}, \forall M > a \geq 0$$

Proof. Soit

$$A = \{h(X) \geq a\}$$

.

Alors

$$h(X) \leq M\mathbf{1}_A + a\mathbf{1}_{A^c}$$

et le résultat suit en passant à l'espérance. \square

Une inégalité comparable à l'inégalité de Markov, mais plus précise, est l'inégalité de Hoeffding :

Théorème 5.11 (Inégalité de Hoeffding). *Soient Y_1, Y_2, \dots des variables aléatoires indépendantes, d'espérance nulle et bornée (pour tout entier positif i , il existe des réels a_i, b_i tels que $a_i \leq Y_i \leq b_i$). Soit $\epsilon > 0$. Alors pour tout $t > 0$,*

$$\mathbb{P}\left(\sum_{i=1}^n Y_i \geq \epsilon\right) \leq e^{-t\epsilon} \prod_{i=1}^n \exp\left(\frac{t^2(b_i^2 - a_i^2)}{8}\right)$$

Théorème 5.12 (Inégalité de Cauchy-Schwarz). *Soient X, Y deux variables aléatoires admettant des moments d'ordre 2, alors*

$$\mathbb{E}XY \leq \sqrt{\mathbb{E}(X^2) \mathbb{E}(Y^2)}$$

Théorème 5.13 (Inégalité de Jensen). *Soit g une fonction convexe, X une variable aléatoire dont l'espérance existe. Alors $g(X)$ est encore une variable aléatoire, et*

$$\mathbb{E}g(X) \geq g(\mathbb{E}X)$$

5.3 Applications

Soient X_1, X_2, \dots des variables aléatoires indépendantes identiquement distribuées. Soit $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ la *moyenne empirique* des X_i .

Théorème 5.14 (Loi faible des grands nombres). *La moyenne empirique converge en probabilité vers μ , l'espérance de X_1 .*

Théorème 5.15 (Loi forte des grands nombres). *La moyenne empirique converge presque sûrement vers μ , l'espérance de X_1 .*

Théorème 5.16 (Théorème central limite). *Supposons que X_1 admette une variance, notée σ^2 . Alors la variable aléatoire $\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$ converge en loi vers une variable aléatoire de loi normale centrée réduite $\mathcal{N}(0, 1)$.*

Part II

Statistiques

Chapter 6

Statistique inférentielle

6.1 Introduction

En statistiques, on est confronté au problème suivant : observant les résultats d'une expérience aléatoire X_1, \dots, X_n que peut-on dire de la loi du processus associé ?

Une approche courante consiste à proposer un modèle décrivant le processus observé, et à inférer les caractéristiques de ce modèle à partir des données observées. Quand ces caractéristiques sont paramétrées par un nombre fini de paramètres, on dit qu'il s'agit d'un *modèle paramétrique*, sinon il s'agit d'un *modèle non-paramétrique*.

6.2 Estimation ponctuelle

Définition 6.1. Soit θ une quantité à estimer. Un estimateur de θ , noté $\hat{\theta}_n$ est une variable aléatoire

$$\hat{\theta}_n = g(X_1, \dots, X_n)$$

Ce paramètre n'est pas nécessairement un des paramètres qui définit la loi, mais peut en être la somme, le produit...

L'idée est de construire $\hat{\theta}_n \simeq \theta$ (θ étant une quantité non observable, mais fixée).

Définition 6.2. Le biais de l'estimateur $\hat{\theta}_n$ est défini par

$$\mathbb{E}\hat{\theta}_n - \theta$$

. L'espérance étant prise suivant la distribution qui a généré les données (donc avec θ observé). Lorsqu'il est nul, on dit que l'estimateur est *sans biais*. Lorsqu'il tend vers 0 lorsque n tend vers $+\infty$, on dit que l'estimateur est *asymptotiquement sans biais*.