

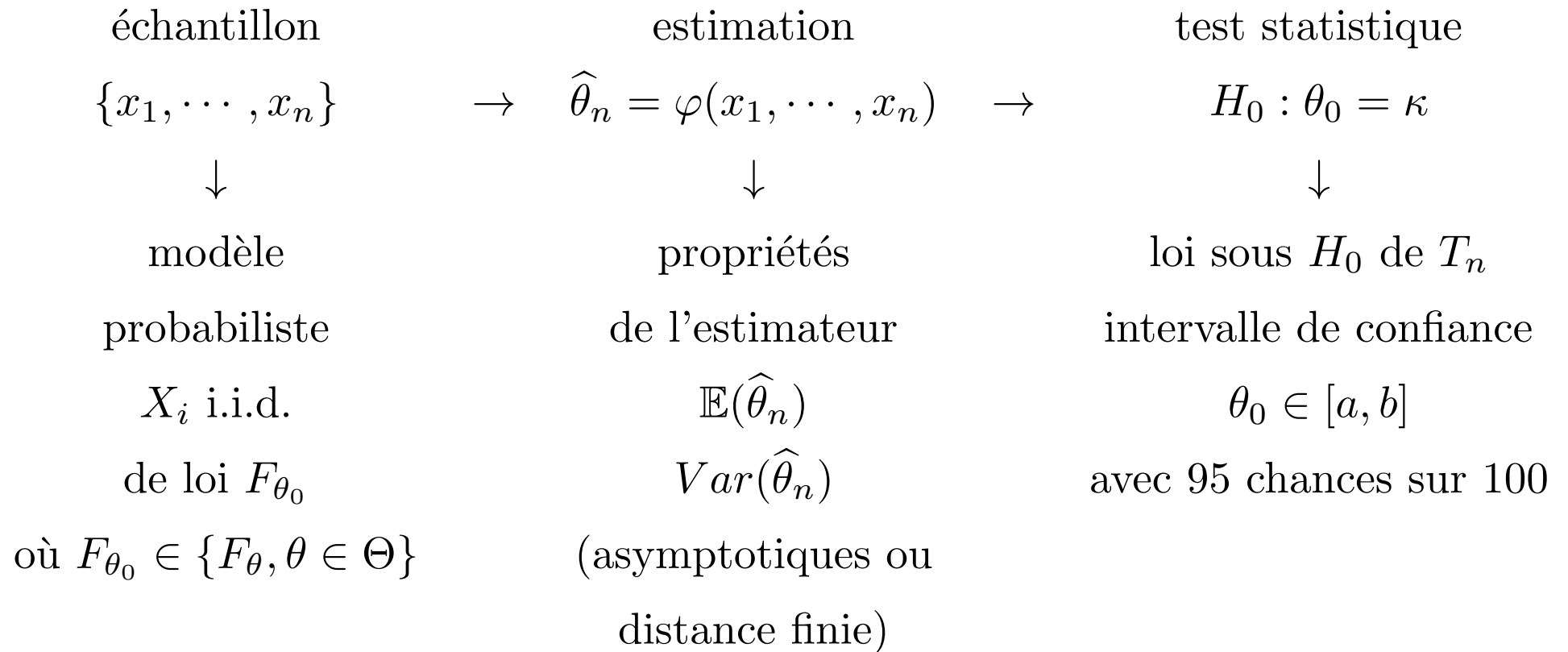
Probabilités & statistiques quelques brefs rappels # 1

Arthur Charpentier, 2014

<http://freakonometrics.hypotheses.org/category/courses/m1-statistique>

Plan du cours

- Introduction, la modélisation statistique
- **Rappels de probabilité**
- Fonctions usuelles, \mathbb{P} , F , f , \mathbb{E} , Var
- Lois usuelles, discrètes et continues
- Conditionnement, espérance conditionnelle et mélanges
- Convergence, approximations et théorèmes limites
 - Loi(s) des grands nombres
 - Théorème central limite
- **Rappels de statistique (mathématique)**
- De la statistique descriptive à la statistique mathématique
- Échantillonnage, moyenne et variance
- Intervalle de confiance
- Introduction aux tests



L'espace de probabilité

On suppose être dans un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

- Ω est l'espace fondamental, $\Omega = \{\omega_i, i \in I\}$ est l'ensemble de tous les résultats possible d'une expérience aléatoire.
- \mathcal{A} est la tribu des évènements, ou l'ensemble des parties de Ω , i.e. l'ensemble des évènements .
- \mathbb{P} est une mesure de probabilité sur Ω , i.e.
 - $\mathbb{P}(\Omega) = 1$
 - pour tout évènement A de Ω , $0 \leq \mathbb{P}(A) \leq 1$,
 - pour tous A_1, \dots, A_n disjoints ($A_i \cap A_j = \emptyset$), $\mathbb{P}(\cup A_i) = \sum \mathbb{P}(A_i)$.

Rappelons qu'une variable aléatoire X est une fonction $\Omega \rightarrow \mathbb{R}$.

Quelques rappels sur les fonctions usuelles

Definition 1. Soit X une variable aléatoire. La fonction de répartition, cumulative distribution function (cdf) de X est

$$F(x) = \mathbb{P}(X \leq x), \text{ pour tout } x \in \mathbb{R}.$$

Formellement, $F(x) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \leq x\})$.

Notons que

- F est une fonction croissante sur \mathbb{R} ,
- $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

On dira que X et Y sont égales en loi, $X \stackrel{\mathcal{L}}{=} Y$ si pour tout x

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(Y \leq x) = F_Y(x).$$

On appelle fonction de survie $\bar{F}(x) = 1 - F(x) = \mathbb{P}(X > x)$.

Sous R, `pexp()` ou `ppois()` renvoient les fonctions de répartition des lois exponentielle ($\mathcal{E}(1)$) et de Poisson.

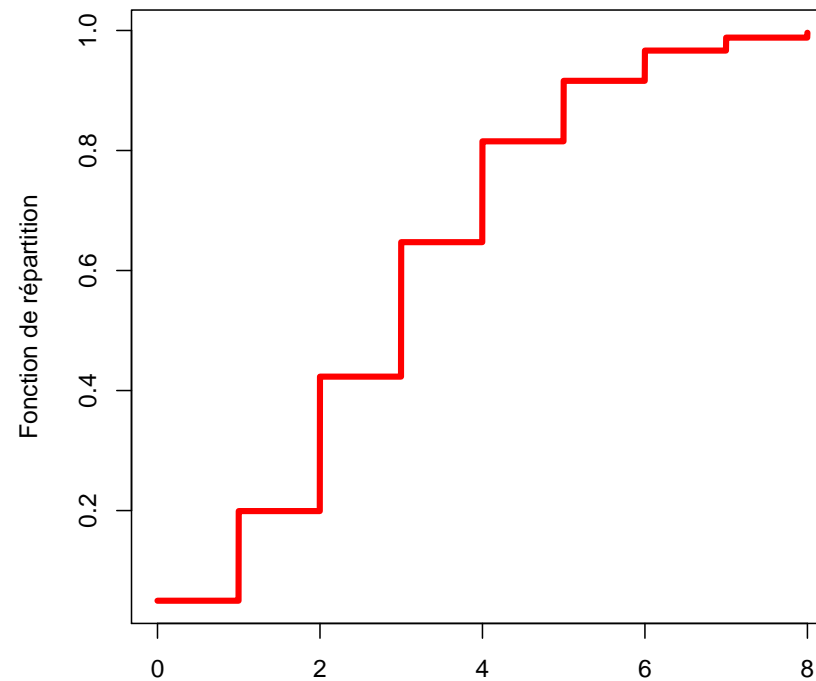
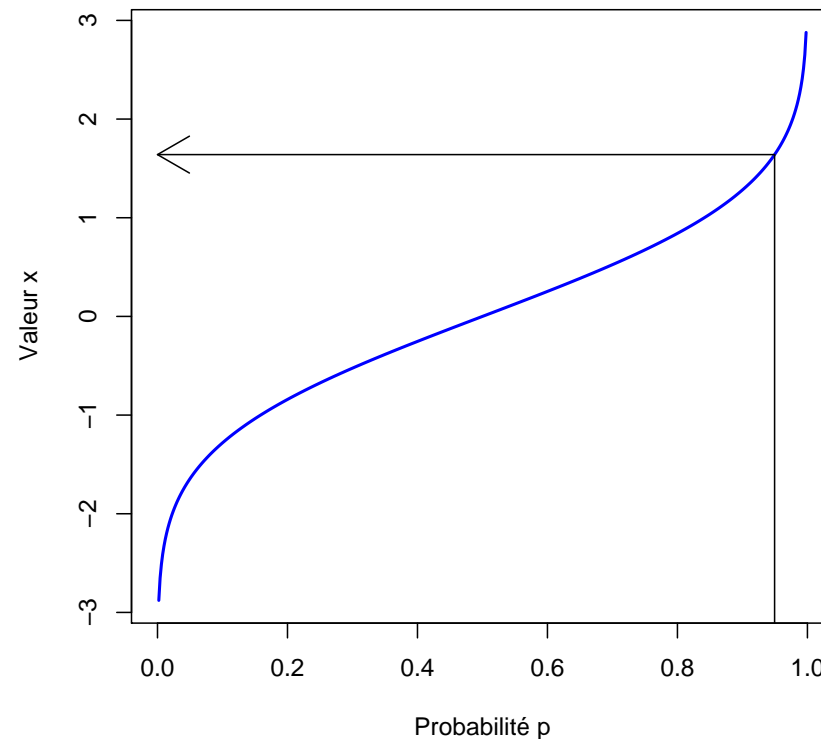


FIGURE 1 – Fonction de répartition $F(x) = \mathbb{P}(X \leq x)$.

Quelques rappels sur les fonctions usuelles

Definition 2. Soit X une variable aléatoire. La fonction quantile de X est

$$Q(p) = F^{-1}(p) = \inf\{x \in \mathbb{R} \text{ tel que } F(x) > p\}, \text{ pour tout } p \in [0, 1].$$



Sous R, `qexp()` ou `qpois()` renvoient quantiles des lois exponentielle ($\mathcal{E}(1)$) et de Poisson.

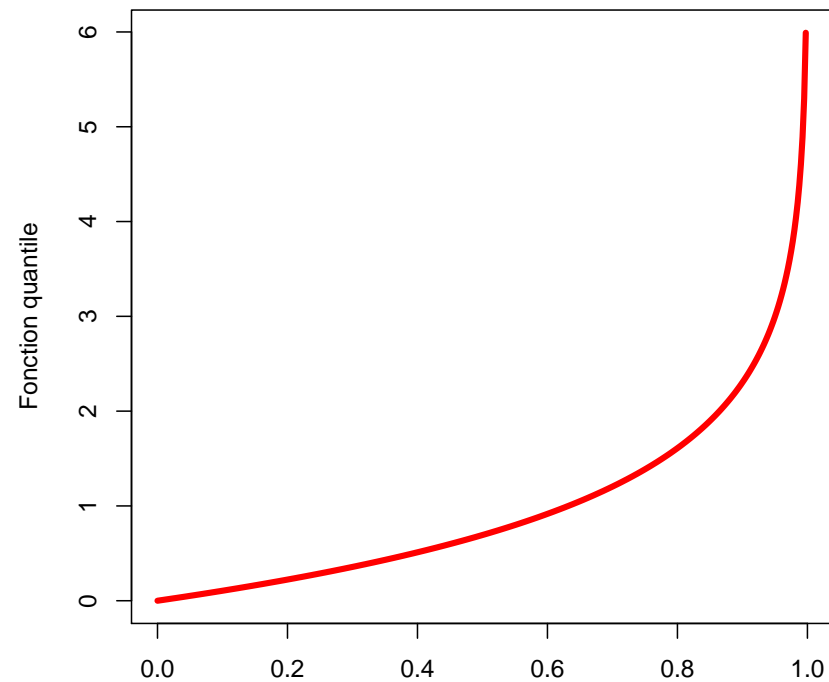


FIGURE 2 – Fonction quantile $Q(p) = F^{-1}(p)$.

Quelques rappels sur les fonctions usuelles

Definition 3. Soit X une variable aléatoire. La fonction de densité ou la fonction de probabilité de X est

$$f(x) = \begin{cases} \frac{dF(x)}{dx} = F'(x) & \text{dans le cas (abs.) continu, } x \in \mathbb{R} \\ \mathbb{P}(X = x) & \text{dans le cas discret, } x \in \mathbb{N} \\ dF(x), & \text{dans un cas plus général} \end{cases}$$

F étant croissante, une densité est *toujours* positive. Dans le cas continu, il est possible d'avoir $f(x) > 1$.

Aussi, $F(x) = \int_{-\infty}^x f(s)ds$ dans le cas continu, $F(x) = \sum_{s=0}^x f(s)$ dans le cas discret.

Sous R, `dexp()` ou `dpois()` renvoient les densités des lois exponentielle ($\mathcal{E}(1)$) et de Poisson.

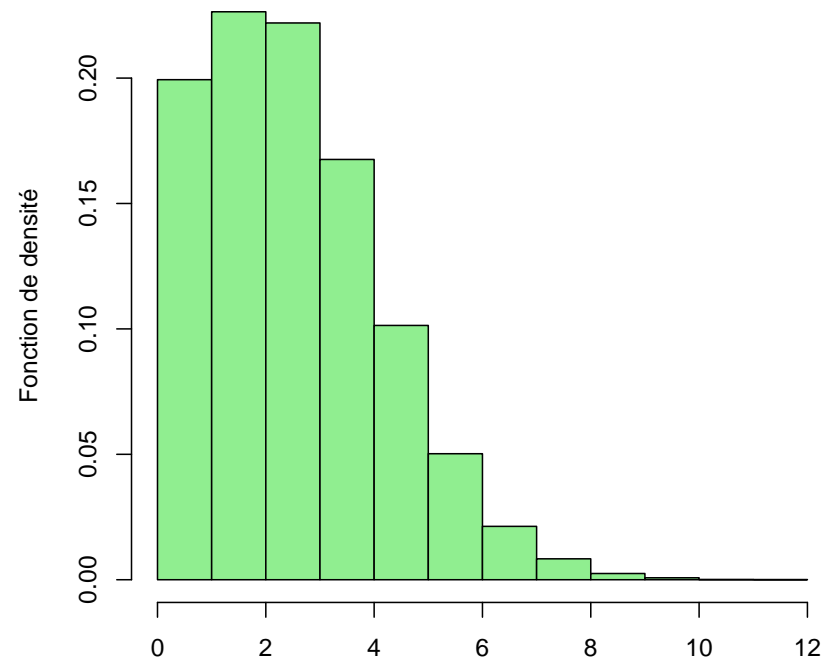
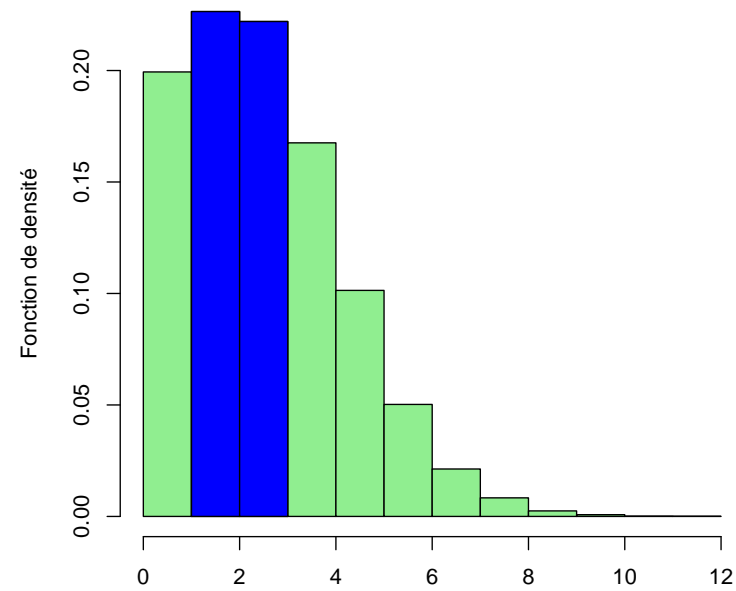


FIGURE 3 – Fonction de densité $f(x) = F'(x)$ ou $f(x) = \mathbb{P}(X = x)$.

$$\mathbb{P}(X \in [a, b]) = \int_a^b f(s)ds \text{ ou } \sum_{s=a}^b f(s).$$

FIGURE 4 – Probabilité $\mathbb{P}(X \in [1, 3[)$.

Quelques rappels sur les vecteurs aléatoires

Definition 4. Soit $\mathbf{Z} = (X, Y)$ un vecteur aléatoire. La fonction de répartition de \mathbf{Z} est

$$F(\mathbf{z}) = F(x, y) = \mathbb{P}(X \leq x, Y \leq y), \text{ pour tout } \mathbf{z} = (x, y) \in \mathbb{R} \times \mathbb{R}.$$

Definition 5. Soit $\mathbf{Z} = (X, Y)$ un vecteur aléatoire. La fonction de densité de \mathbf{Z} est

$$f(\mathbf{z}) = f(x, y) = \begin{cases} \frac{\partial F(x, y)}{\partial x \partial y} & \text{dans le cas continu, } \mathbf{z} = (x, y) \in \mathbb{R} \times \mathbb{R} \\ \mathbb{P}(X = x, Y = y) & \text{dans le cas discret, } \mathbf{z} = (x, y) \in \mathbb{N} \times \mathbb{N} \end{cases}$$

Quelques concepts sur les vecteurs aléatoires

Etant donné un vecteur aléatoire $\mathbf{Z} = (X, Y)$ de fonction de répartition F et de densité f , on peut en déduire les lois marginales de X (ou de Y) en notant simplement que

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x, Y \leq +\infty) = \lim_{y \rightarrow \infty} F(x, y),$$

$$f_X(x) = \mathbb{P}(X = x) = \sum_{y=0}^{\infty} \mathbb{P}(X = x, Y = y) = \sum_{y=0}^{\infty} f(x, y), \text{ dans le cas discret}$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \text{ dans le cas continu}$$

La loi conditionnelle $Y|X$

De plus, on peut définir **loi conditionnelle de Y sachant $X = x$** , dont la fonction de densité est donnée par la **formule de Bayes**

$$\mathbb{P}(Y = y|X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)} \text{ dans le cas discret,}$$

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)}, \text{ dans le cas continu.}$$

On peut également considérer la fonction de répartition

$$\mathbb{P}(Y \leq y|X = x) = \sum_{t=0}^y \mathbb{P}(Y = t|X = x) = \sum_{t=0}^y \frac{\mathbb{P}(X = x, Y = t)}{\mathbb{P}(X = x)} \text{ dans le cas discret,}$$

$$F_{Y|X=x}(y) = \int_{-\infty}^y f_{Y|X=x}(t) dt = \frac{1}{f_X(x)} \int_{-\infty}^y f(x, t) dt, \text{ dans le cas continu.}$$

Quelques concepts sur les vecteurs aléatoires

Definition 6. Soient X et Y deux variables aléatoires. On dira que X et Y sont *indépendantes* si une des conditions suivantes est satisfaite

- $F(x, y) = F_X(x)F_Y(y)$ pour tout x, y , ou
 $\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \times \mathbb{P}(Y \leq y)$,
- $f(x, y) = f_X(x)f_Y(y)$ pour tout x, y , ou
 $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y)$,
- $F_{Y|X=x}(y) = F_Y(y)$ pour tout x, y , ou $f_{Y|X=x}(y) = f_Y(y)$,
- $F_{X|Y=y}(x) = F_X(x)$ pour tout x, y , ou $f_{X|Y=y}(x) = f_X(x)$.

On notera $X \perp Y$ pour signifier l'indépendance.

Etude de l'indépendance

Les tableaux suivants donnent les valeurs de $\mathbb{P}(X = \cdot, Y = \cdot)$

	$X = 0$	$X = 1$
$Y = 0$	0.15	0.2
$Y = 1$	0.45	0.2

	$X = 0$	$X = 1$
$Y = 0$	0.15	0.1
$Y = 1$	0.45	0.3

Dans les deux cas,

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 0, Y = 0) + \mathbb{P}(X = 0, Y = 1) = 0.15 + 0.45 = 0.6, \text{ i.e.}$$

$$X \sim \mathcal{B}(0.4)$$

Dans le premier cas, X et Y ne sont pas indépendantes, mais le sont dans le second.

Notion d'indépendance conditionnelle

On dira que X et Y sont conditionnellement indépendantes sachant Z si pour tout z tel que $\mathbb{P}(Z \leq z) > 0$,

$$\mathbb{P}(X \leq x, Y \leq y \mid Z = z) = \mathbb{P}(X \leq x \mid Z = z) \cdot \mathbb{P}(Y \leq y \mid Z = z)$$

Quelques rappels sur les moments

Definition 7. Soit X une variable aléatoire. L'espérance (expected value) de X est

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx \text{ ou } \sum_{x=0}^{\infty} x \cdot \mathbb{P}(X = x)$$

Definition 8. Soit $\mathbf{Z} = (X, Y)$ un vecteur aléatoire. L'espérance de \mathbf{Z} est

$$\mathbb{E}(\mathbf{Z}) = \begin{pmatrix} \mathbb{E}(X) \\ \mathbb{E}(Y) \end{pmatrix}$$

Proposition 9. L'espérance de $Y = g(X)$, où X a pour densité f , est

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

Notons que si g n'est pas linéaire, $\mathbb{E}(g(X)) \neq g(\mathbb{E}(X))$.

Quelques rappels sur l'espérance

Proposition 10. *Soient X et Y deux variables aléatoires d'espérance finie,*

- $\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y)$, pour tout α, β , i.e. l'espérance est linéaire,
- $\mathbb{E}(XY) \neq \mathbb{E}(X) \times \mathbb{E}(Y)$ de manière générale, mais l'égalité est vérifiée si $X \perp Y$.

L'espérance d'une variable aléatoire est un nombre réel.

Exemple Considérons une loi uniforme sur $[a, b]$, de densité

$$f(x) = \frac{1}{b-a} \mathbf{1}(x \in [a, b]),$$

$$\begin{aligned} \mathbb{E}(X) &= \int_{\mathbb{R}} x f(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b \\ &= \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{1}{b-a} \frac{(b-a)(a+b)}{2} = \frac{a+b}{2}. \end{aligned}$$

Il existe des variables aléatoires dont l'espérance n'existe pas (ou est infinie).

Exemple 11 (Paradoxe de Saint Peterbroug). *Considérons un jeu de pile ou face où les gains sont doublés à chaque lancer, et le jeu est arrêté au 1er pile.*

$$\begin{aligned}\mathbb{E}(X) &= 1 \times \mathbb{P}(\text{pile au 1er lancer}) \\ &\quad + 1 \times 2 \times \mathbb{P}(\text{1er pile au 2ème lancer}) \\ &\quad + 2 \times 2 \times \mathbb{P}(\text{1er pile au 3ème lancer}) \\ &\quad + 4 \times 2 \times \mathbb{P}(\text{1er pile au 4ème lancer}) \\ &\quad + 8 \times 2 \times \mathbb{P}(\text{1er pile au 5ème lancer}) \\ &\quad + 16 \times 2 \times \mathbb{P}(\text{1er pile au 6ème lancer}) \\ &\quad + 32 \times 2 \times \mathbb{P}(\text{1er pile au 7ème lancer}) + \dots \\ &= \frac{1}{2} + \frac{2}{4} + \frac{4}{8} + \frac{8}{16} + \frac{16}{32} + \frac{32}{64} + \dots = \infty.\end{aligned}$$

L'espérance conditionnelle

Definition 12. Soient X et Y deux variables aléatoires. L'espérance conditionnelle (conditional expectation) de Y sachant $X = x$ est l'espérance associée à la loi conditionnelle $Y|X = x$,

$$\mathbb{E}(Y|X = x) = \int_{-\infty}^{\infty} y \cdot f_{Y|X=x}(y) dy \text{ ou } \sum_{x=0}^{\infty} y \cdot \mathbb{P}(Y = y|X = x).$$

$\mathbb{E}(Y|X = x)$ est donc fonction de x , $\mathbb{E}(Y|X = x) = \varphi(x)$. Par abus de notation, on appellera $\mathbb{E}(Y|X)$ la variable aléatoire $\varphi(X)$.

Proposition 13. $\mathbb{E}(Y|X)$ étant une variable aléatoire, notons que

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

Démonstration.

$$\begin{aligned}\mathbb{E}(\mathbb{E}(X|Y)) &= \sum_y \mathbb{E}(X|Y = y) \cdot \mathbb{P}(Y = y) \\ &= \sum_y \left(\sum_x x \cdot \mathbb{P}(X = x|Y = y) \right) \cdot \mathbb{P}(Y = y) \\ &= \sum_y \sum_x x \cdot \mathbb{P}(X = x|Y = y) \cdot \mathbb{P}(Y = y) \\ &= \sum_x \sum_y x \cdot \mathbb{P}(Y = y|X = x) \cdot \mathbb{P}(X = x) \\ &= \sum_x x \cdot \mathbb{P}(X = x) \cdot \left(\sum_y \mathbb{P}(Y = y|X = x) \right) \\ &= \sum_x x \cdot \mathbb{P}(X = x) = \mathbb{E}(X).\end{aligned}$$

□

Quelques rappels sur les moments

Pour travailler sur le **moment d'ordre 2**, rappelons que

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

$$\mathbb{E}(g(X, Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) \cdot f(x, y) dx dy.$$

Definition 14. *Soit X une variable aléatoire. La **variance** (variance) de X est*

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 \cdot f(x) dx \text{ ou } \sum_{x=0}^{\infty} (x - \mathbb{E}(X))^2 \cdot \mathbb{P}(X = x).$$

Une écriture équivalente est $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

La variance mesure la dispersion des valeurs prises par X autour de $\mathbb{E}(X)$, c'est un nombre réel. $\sqrt{\text{Var}(X)}$ est appelé **écart-type** (standard deviation), ou volatilité.

Quelques rappels sur les moments

Definition 15. Soit $\mathbf{Z} = (X, Y)$ un vecteur aléatoire. La *matrice de variance-covariance* (variance-covariance matrix) de \mathbf{Z} est

$$\text{Var}(\mathbf{Z}) = \begin{pmatrix} \text{Var}(X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \text{Var}(Y) \end{pmatrix}$$

où $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$ et

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

Definition 16. Soit $\mathbf{Z} = (X, Y)$ un vecteur aléatoire. La *corrélation* entre X et Y est

$$\text{corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} = \frac{\mathbb{E}[(X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))]}{\sqrt{\mathbb{E}[(X - \mathbb{E}(X))^2] \cdot \mathbb{E}[(Y - \mathbb{E}(Y))^2]}}.$$

Quelques rappels sur la variance

Proposition 17. *La variance est toujours positive, et $\text{Var}(X) = 0$ si et seulement si X est une constante.*

Proposition 18. *La variance n'est pas linéaire, mais*

$$\text{Var}(\alpha X + \beta Y) = \alpha^2 \text{Var}(X) + 2\alpha\beta \text{Cov}(X, Y) + \beta^2 \text{Var}(Y).$$

Une conséquence est que

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{j \neq i} \text{Cov}(X_i, X_j) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{j > i} \text{Cov}(X_i, X_j).$$

Proposition 19. *La variance n'est pas linéaire, mais $\text{Var}(\alpha + \beta X) = \beta^2 \text{Var}(X)$.*

Quelques rappels sur la covariance

Proposition 20. *Pour toutes variables aléatoires X , X_1 , X_2 et Y ,*

- $Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$,
- $Cov(\alpha X_1 + \beta X_2, Y) = \alpha Cov(X_1, Y) + \beta Cov(X_2, Y)$.

$$Cov(X, Y) = \sum_{\omega \in \Omega} [X(\omega) - \mathbb{E}(X)] \cdot [Y(\omega) - \mathbb{E}(Y)] \cdot \mathbb{P}(\omega)$$

Un interprétation heuristique d'une covariance positive est qu'il doit y avoir une majorité d'évènements ω pour lesquels

$$[X(\omega) - \mathbb{E}(X)] \cdot [Y(\omega) - \mathbb{E}(Y)] \geq 0$$

- $X(\omega) \geq \mathbb{E}(X)$ et $Y(\omega) \geq \mathbb{E}(Y)$, i.e. X et Y prennent ensemble des grandes valeurs,
- $X(\omega) \leq \mathbb{E}(X)$ et $Y(\omega) \leq \mathbb{E}(Y)$, i.e. X et Y prennent ensemble des petites valeurs.

Proposition 21. *Si X et Y sont indépendantes ($X \perp Y$), alors $Cov(X, Y) = 0$, mais la réciproque n'est généralement pas vraie.*

La variance conditionnelle

Definition 22. Soient X et Y deux variables aléatoires. La *variance conditionnelle* (conditional variance) de Y sachant $X = x$ est la variance associée à la loi conditionnelle $Y|X = x$,

$$\text{Var}(Y|X = x) = \int_{-\infty}^{\infty} [y - \mathbb{E}(Y|X = x)]^2 \cdot f_{Y|X=x}(y) dy.$$

$\text{Var}(Y|X = x)$ est donc fonction de x , $\mathbb{E}(Y|X = x) = \psi(x)$. Par abus de notation, on appellera $\text{Var}(Y|X)$ la variable aléatoire $\psi(X)$.

Proposition 23. *Var(Y|X) étant une variable aléatoire, notons que*

$$\text{Var}(Y) = \text{Var}[\mathbb{E}(Y|X)] + \mathbb{E}[\text{Var}(Y|X)],$$

formule dite de décomposition de la variance (ou théorème de Pythagore).

Démonstration. L'idée est de faire un développement de la forme suivante

$$\begin{aligned} \text{Var}(Y) &= \mathbb{E}[(Y - \mathbb{E}(Y))^2] = \mathbb{E}[(Y - \mathbb{E}(Y|X) + \mathbb{E}(Y|X) - \mathbb{E}(Y))^2] \\ &= \mathbb{E}[(Y - \mathbb{E}(Y|X)) + (\mathbb{E}(Y|X) - \mathbb{E}(Y))]^2 \\ &= \mathbb{E}[(Y - \mathbb{E}(Y|X))]^2 + \mathbb{E}[(\mathbb{E}(Y|X) - \mathbb{E}(Y))]^2 \\ &\quad + 2\mathbb{E}[(Y - \mathbb{E}(Y|X)) \cdot (\mathbb{E}(Y|X) - \mathbb{E}(Y))] \end{aligned}$$

On note alors que

$$\mathbb{E}[(Y - \mathbb{E}(Y|X))]^2 = \mathbb{E}(\mathbb{E}((Y - \mathbb{E}(Y|X))^2 | X)) = \mathbb{E}[\text{Var}(Y|X)],$$

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

Et l'espérance du produit croisé est nulle (en conditionnant par X). □

Un tout petit peu de géométrie

Si L^2 désigne l'ensemble des variables aléatoires de variance finie,

- $\langle X, Y \rangle = \mathbb{E}(XY)$ est un produit scalaire,
- $\|X\| = \sqrt{\mathbb{E}(X^2)}$ est une norme (parfois notée $\|\cdot\|_2$).

$\mathbb{E}(X)$ est la projection orthogonale de X sur l'ensemble des constantes,

$$\mathbb{E}(X) = \operatorname{argmin}_{a \in \mathbb{R}} \{ \|X - a\|^2 = \mathbb{E}([X - a]^2) \}.$$

La corrélation est le cosinus de l'angle entre $X - \mathbb{E}(X)$ et $Y - \mathbb{E}(Y)$: si $\operatorname{corr}(X, Y) = 0$ on dira que les variables sont orthogonales, $X \perp Y$ (et non pas indépendantes, $X \perp\!\!\!\perp Y$).

Si L_X^2 désigne l'ensemble des variables aléatoires engendrées par X (de la forme $\varphi(X)$) de variance finie,

$\mathbb{E}(Y|X)$ est la projection orthogonale de Y sur l'ensemble L_X^2

$$\mathbb{E}(Y|X) = \operatorname{argmin}_{\varphi} \{ \|Y - \varphi(X)\|^2 = \mathbb{E}([Y - \varphi(X)]^2) \}.$$

$\mathbb{E}(Y|X)$ est la meilleure approximation de Y par une fonction de X .

L'espérance conditionnelle

De manière générale, $\mathbb{E}(Y|X) = \varphi(X)$. En **économétrie**, on cherche à expliquer Y par X .

- **économétrie linéaire**, $EL(Y|X) = \beta_0 + \beta_1 X$.
- **économétrie nonlinéaire**, $\mathbb{E}(Y|X) = \varphi(X)$.

ou plus généralement on cherche à expliquer Y par \mathbf{X} .

- **économétrie linéaire**, $EL(Y|\mathbf{X}) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$.
- **économétrie nonlinéaire**, $\mathbb{E}(Y|\mathbf{X}) = \varphi(\mathbf{X}) = \varphi(X_1, \cdots, X_k)$.

En **séries temporelle**, on cherche à expliquer X_t par X_{t-1}, X_{t-2}, \cdots .

- **séries temporelles linéaires**,
 $EL(X_t|X_{t-1}, X_{t-2}, \cdots) = \beta_0 + \beta_1 X_{t-1} + \cdots + \beta_k X_{t-k}$ (forme autorégressive).
- **séries temporelles nonlinéaires**, $\mathbb{E}(X_t|X_{t-1}, X_{t-2}, \cdots) = \varphi(X_{t-1}, X_{t-2}, \cdots)$.

Un petit complément sur l'espérance conditionnelle

L'espérance conditionnelle n'est pas un réel, mais une variable aléatoire.

Formellement, on ne conditionne pas par rapport à une variable aléatoire X , mais la tribu engendrée par X ,

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

Plus généralement, on peut conditionner par une tribu quelconque \mathcal{F} , qui sera interprétée comme de l'information disponible pour expliquer Y .

En séries temporelles, on dispose d'une suite de variables aléatoires $X_1, X_2, \dots, X_n, \dots$. On note \mathcal{F}_n l'information disponible à la date n , i.e. $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, et on cherchera $\mathbb{E}(X_{n+1}|\mathcal{F}_n)$.

Si $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$, on parlera de martingale.

Additionner des variables aléatoires

Proposition 24. *Soient X et Y sont deux variables discrètes indépendantes. Alors la loi de $S = X + Y$ est*

$$\mathbb{P}(S = s) = \sum_{k=0}^{\infty} \mathbb{P}(X = k) \times \mathbb{P}(Y = s - k).$$

Soient X et Y sont deux variables continues indépendantes. Alors la loi de $S = X + Y$ est

$$f_S(s) = \int_{-\infty}^{\infty} f_X(x) \times f_Y(s - x) dx.$$

*On notera $f_S = f_X \star f_Y$ où \star est l'opérateur de *convolution*.*

Un petit exercice

Montrer que pour X positive et dans L^1

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{\infty} \mathbb{P}(X > x) dx \\ &= \int_0^{\infty} (1 - F(x)) dx \text{ pour une variable continue,} \\ &= \sum_{k=0}^{\infty} \mathbb{P}(X > k) \text{ pour une variable discrète.} \end{aligned}$$

En effet, dans le cas continue,

$$\begin{aligned} \int_0^{\infty} (1 - F(x)) dx &= \int_0^{\infty} \left\{ \int_x^{\infty} f(t) dt \right\} dx = \int_0^{\infty} \left\{ \int_0^t f(t) dx \right\} dt \\ &= \int_0^{\infty} \left\{ f(t) \int_0^t dx \right\} dt = \int_0^{\infty} f(t) t dt = \mathbb{E}(X), \end{aligned}$$

en intervertissant les signes sommes (théorème de Tonelli). Et dans le cas discret,

$$\begin{aligned} \sum_{k=0}^{\infty} \mathbb{P}(X > k) &= \sum_{k=0}^{\infty} \left\{ \sum_{i=k+1}^{\infty} \mathbb{P}(X = i) \right\} = \sum_{i=1}^{\infty} \left\{ \sum_{k=1}^i \mathbb{P}(X = i) \right\} \\ &= \sum_{i=1}^{\infty} \left\{ \mathbb{P}(X = i) \sum_{k=1}^i 1 \right\} = \sum_{i=1}^{\infty} i \mathbb{P}(X = i) = \mathbb{E}(X), \end{aligned}$$

en utilisant le même argument que précédemment.

En fait, l'intégration par partie ne peut pas être utilisé pour la raison suivante : on aurait souhaité écrire, dans le cas continue,

$$\mathbb{E}(X) = \int_0^{\infty} x f(x) dx = [x(F(x) - 1)]_0^{\infty} - \int_0^{\infty} (F(x) - 1) dx,$$

en dérivant x , et en prenant comme primitive de $f(x)$ $F(x) - 1$ (afin d'avoir au moins la nullité en $+\infty$, sinon l'intégrale de droite n'aurait aucune chance de converger). Le problème est que rien ne garantie, a priori, que $x(F(x) - 1) \rightarrow 0$ quand $x \rightarrow \infty$.

Quelques rappels sur les moments

On définit le moment d'ordre n d'une variable aléatoire X comme $\mu_n = \mathbb{E}[X^n]$, si cette valeur est finie. On notera également μ'_n les moments centrés (et éventuellement réduits).

Certains moments sont utiles pour caractériser une variable aléatoire :

- Le moment d'ordre un de la variable : $\mu = \mathbb{E}[X]$ correspond à l'**espérance**
- Le moment d'ordre deux de la variable **centrée** : $\mu'_2 = \mathbb{E}[(X - \mu)^2]$ correspond à la **variance**.
- Le moment d'ordre trois de la variable **centrée-réduite** :

$$\mu'_3 = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right]$$
 correspond au coefficient d'asymétrie, la **skewness**.
- Le moment d'ordre quatre de la variable **centrée-réduite** :

$$\mu'_4 = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right]$$
 correspond à la **kurtosis**.

Quelques rappels sur les lois de probabilité

La loi de **Bernoulli** $\mathcal{B}(p)$, $p \in (0, 1)$

$$\mathbb{P}(X = 0) = 1 - p \text{ et } \mathbb{P}(X = 1) = p.$$

Alors $\mathbb{E}(X) = p$ et $\text{Var}(X) = p(1 - p)$.

Quelques rappels sur les lois de probabilité

La loi **Binomiale** $\mathcal{B}(n, p)$, $p \in (0, 1)$ et $n \in \mathbb{N}^*$

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \text{ où } k = 0, 1, \dots, n, \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Alors $\mathbb{E}(X) = np$ et $\text{Var}(X) = np(1 - p)$.

Notons que si $X_1, \dots, X_n \sim \mathcal{B}(p)$ sont indépendantes, alors

$$X = X_1 + \dots + X_n \sim \mathcal{B}(n, p).$$

Sous R, `dbinom(x, size, prob)`, `qbinom()` et `pbinom()` désignent respectivement la fonction de répartition, la fonction quantile et la loi de probabilité.

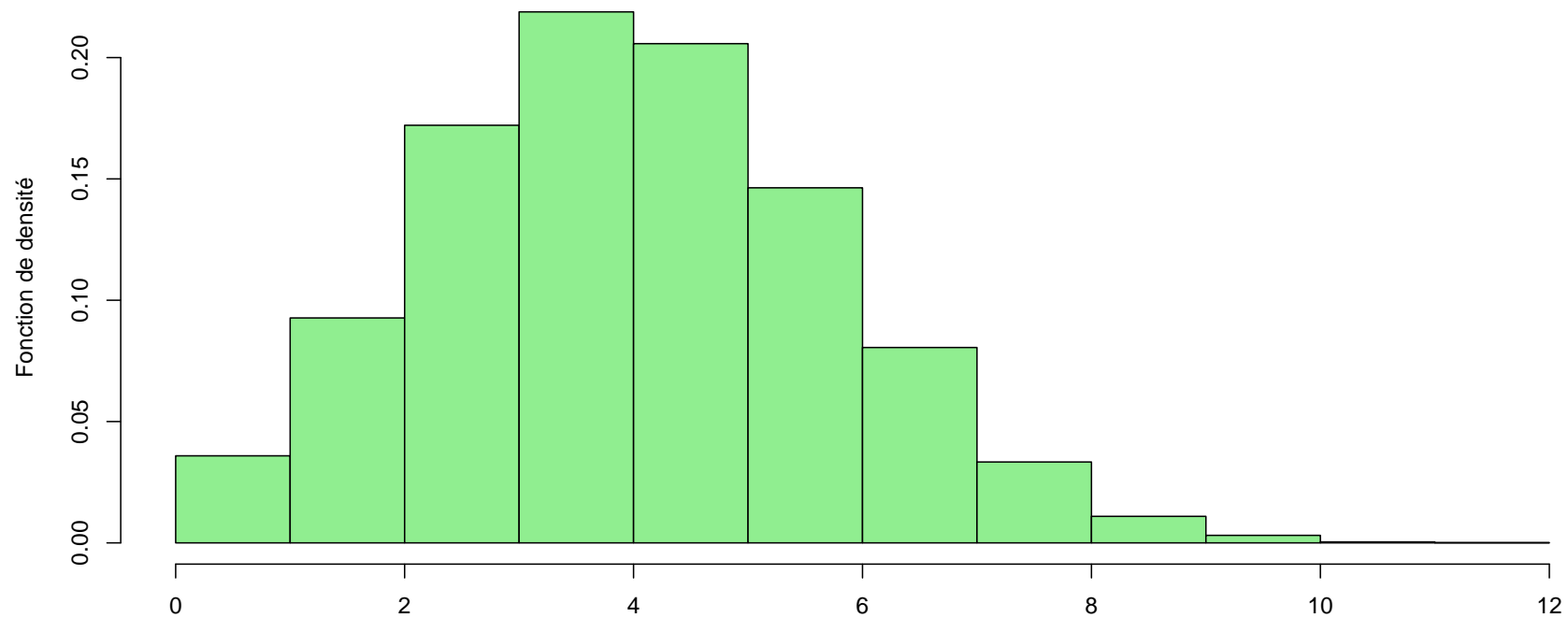


FIGURE 5 – Loi binomiale, $\mathcal{B}(n, p)$.

Quelques rappels sur les lois de probabilité

La loi de **Poisson** $\mathcal{P}(\lambda)$, $\lambda > 0$

$$\mathbb{P}(X = k) = \exp(-\lambda) \frac{\lambda^k}{k!} \text{ où } k = 0, 1, \dots$$

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

De plus, si $X_1 \sim \mathcal{P}(\lambda_1)$ et $X_2 \sim \mathcal{P}(\lambda_2)$ sont deux variables indépendantes, alors

$$X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$$

On peut noter que cette loi vérifie une relation de récurrence de la forme

$$\frac{\mathbb{P}(X = k + 1)}{\mathbb{P}(X = k)} = \frac{\lambda}{k + 1} \text{ pour } k \geq 1$$

Sous R, `dpois(x, lambda)`, `qpois()` et `ppois()` désignent respectivement la fonction de répartition, la fonction quantile et la loi de probabilité.

Quelques rappels sur les lois de probabilité

La loi de géométrie^a $\mathcal{G}(p)$, $p \in]0, 1[$

$$\mathbb{P}(X = k) = p(1 - p)^{k-1} \text{ pour } k = 1, 2, \dots$$

dont la fonction de répartition est $\mathbb{P}(N \leq k) = 1 - p^k$.

On peut noter que cette loi vérifie une relation de récurrence de la forme

$$\frac{\mathbb{P}(X = k + 1)}{\mathbb{P}(X = k)} = 1 - p (= \text{constante}) \text{ pour } k \geq 1$$

Les premiers moments sont alors

$$\mathbb{E}(X) = \frac{1}{p} \text{ et } \text{Var}(X) = \frac{1 - p}{p^2}.$$

a. Il est également possible de définir cette loi sur \mathbb{N} , et non pas comme ici sur $\mathbb{N} \setminus \{0\}$.

Quelques rappels sur les lois de probabilité

La loi **exponentielle** $\mathcal{E}(\lambda)$, $\lambda > 0$

$$F(x) = \mathbb{P}(X \leq x) = e^{-\lambda x} \text{ où } x \geq 0, f(x) = \lambda e^{-\lambda x}.$$

Alors $\mathbb{E}(X) = 1/\lambda$ et $\text{Var}(X) = 1/\lambda^2$.

Cette loi est parfois dite *sans mémoire*, i.e.

$$\mathbb{P}(X > x + t | X > x) = \mathbb{P}(X > t).$$

Sous R, `dexp(x, rate)`, `qexp()` et `pexp()` désigne respectivement la fonction de répartition, la fonction quantile et la densité.

Quelques rappels sur les lois de probabilité

La loi **Gaussienne** (ou **normale**) $\mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ et $\sigma > 0$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \text{ pour } x \in \mathbb{R}.$$

Alors $\mathbb{E}(X) = \mu$ et $\text{Var}(X) = \sigma^2$.

Notons que si $Z \sim \mathcal{N}(0, 1)$, $X = \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$.

Sous R, `dnorm(x, mean, sd)`, `qnorm()` et `pnorm()` désigne respectivement la fonction de répartition, la fonction quantile et la densité.

Sous R, `dnorm(x, mean=a, sd=b)` ($\mathcal{N}(a, b)$).

Quelques rappels sur les lois de probabilité

Le vecteur Gaussien $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

$X = (X_1, \dots, X_n)$ est un vecteur gaussien de moyenne $\mathbb{E}(X) = \boldsymbol{\mu}$ et de matrice de variance covariance $\boldsymbol{\Sigma} = \mathbb{E}\left((X - \boldsymbol{\mu})(X - \boldsymbol{\mu})^\top\right)$ non dégénéré ($\boldsymbol{\Sigma}$ est inversible) si et seulement si sa densité s'écrit

$$f(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2}(x - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (x - \boldsymbol{\mu})\right), \quad x \in \mathbb{R}^d,$$

ou de façon équivalente, si sa fonction caractéristique $\phi(t) = \exp(it^\top X)$, s'écrit

$$\phi(t) = \exp\left(it^\top \boldsymbol{\mu} - \frac{1}{2}t^\top \boldsymbol{\Sigma} t\right), \quad t \in \mathbb{R}^d.$$

Proposition 25. *Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur aléatoire à valeurs dans \mathbb{R}^d , alors \mathbf{X} sera dit être un vecteur gaussien si et seulement si, pour tout $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^d$, $\mathbf{a}^\top \mathbf{X} = a_1 X_1 + \dots + a_n X_n$ est une variable gaussienne.*

Quelques rappels sur les lois de probabilité

Soit \mathbf{X} un vecteur gaussien, alors pour tout i , X_i suit une loi Gaussienne. Mais la réciproque est fautive : si les X_i sont des variables gaussiennes, alors

$\mathbf{X} = (X_1, \dots, X_n)$ n'est pas nécessairement un vecteur gaussien.

Proposition 26. *Si $\mathbf{X} = (X_1, \dots, X_n)$ est un vecteur gaussien de moyenne $\mathbb{E}(\mathbf{X}) = \mu$ et de matrice de variance covariance Σ , et A est une matrice $k \times n$, et $b \in \mathbb{R}^k$, alors $\mathbf{Y} = A\mathbf{X} + b$ est un vecteur gaussien de \mathbb{R}^k , de loi $\mathcal{N}(A\mu, A\Sigma A^T)$.*

Notons que si (X_1, X_2) est un vecteur gaussien, alors X_1 et X_2 sont des variables indépendantes si et seulement si

$$\text{Cov}(X_1, X_2) = \mathbb{E}((X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))) = 0,$$

cette équivalence étant en général fautive dans un cadre non-gaussien.

Quelques rappels sur les lois de probabilité

Soit \mathbf{X} un vecteur gaussien, alors pour tout i , X_i suit une loi Gaussienne. Mais la réciproque est fautive : si les X_i sont des variables gaussiennes, alors $\mathbf{X} = (X_1, \dots, X_n)$ n'est pas nécessairement un vecteur gaussien.

Proposition 27. *Si $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ est un vecteur gaussien de moyenne*

$$\mathbb{E}(\mathbf{X}) = \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \text{ et de matrice de variance covariance } \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \text{ alors}$$

$$\mathbf{X}_2 | \mathbf{X}_1 = \mathbf{x}_1 \sim \mathcal{N}(\mu_2 + \Sigma_{12} \Sigma_{11}^{-1} (\mathbf{x}_1 - \mu_1), \Sigma_{22} - \Sigma_{12} \Sigma_{11}^{-1} \Sigma_{21}).$$

Considérons une série temporelle définie par $X_t = \rho X_{t-1} + \varepsilon_t$, avec $X_0 = 0$, $\varepsilon_1, \dots, \varepsilon_n$ i.i.d. $\mathcal{N}(0, \sigma^2)$, i.e. $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbb{I})$. Alors

$$\mathbf{X} = (X_1, \dots, X_n) \sim \mathcal{N}(\mathbf{0}, \Sigma), \Sigma = [\Sigma_{i,j}] = [\text{Cov}(X_i, X_j)] = [\rho^{|i-j|}].$$

Quelques rappels sur les lois de probabilité

En dimension 2, pour un vecteur (X, Y) centré (i.e. $\mu = \mathbf{0}$)

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2\rho xy}{(\sigma_x\sigma_y)}\right)\right)$$

où la matrice de variance-covariance Σ est

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}.$$

Quelques rappels sur les lois de probabilité

La loi du **chi-deux** $\chi^2(\nu)$, $\nu \in \mathbb{N}^*$ admet pour densité

$$\frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x/2}, x \in [0; +\infty[.$$

Notons que $\mathbb{E}(X) = \nu$ et $\text{Var}(X) = 2\nu$.

Si $X_1, \dots, X_\nu \sim \mathcal{N}(0, 1)$ sont indépendantes, alors $Y = \sum_{i=1}^{\nu} X_i^2 \sim \chi^2(\nu)$.

Sous R, `dchisq(x, df)`, `qchisq()` et `pchisq()` désignent respectivement la fonction de répartition, la fonction quantile et la densité.

Quelques rappels sur les lois de probabilité

La loi de Student $St(\nu)$, de densité

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)},$$

où Γ désigne la fonction Gamma ($\Gamma(n+1) = n!$). Notons que

$$\mathbb{E}(X) = 0 \text{ et } \text{Var}(X) = \frac{\nu}{\nu-2} \text{ pour } \nu > 2.$$

Si $X \sim \mathcal{N}(0, 1)$ et $Y \sim \chi^2(\nu)$ sont indépendantes, alors

$$T = \frac{X}{\sqrt{Y/\nu}} \sim St(\nu).$$

Quelques rappels sur les lois de probabilité

Soient X_1, \dots, X_n des variables $\mathcal{N}(\mu, \sigma^2)$ indépendantes. Posons

$$\bar{X}_n = (X_1 + \dots + X_n)/n \text{ et } S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Alors $(n-1)S_n^2/\sigma^2$ suit une loi $\chi^2(n-1)$, et on peut montrer que

$$T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim \text{St}(n-1).$$

Sous R, `dt(x, df)`, `qt()` et `pt()` désignent respectivement la fonction de répartition, la fonction quantile et la densité.

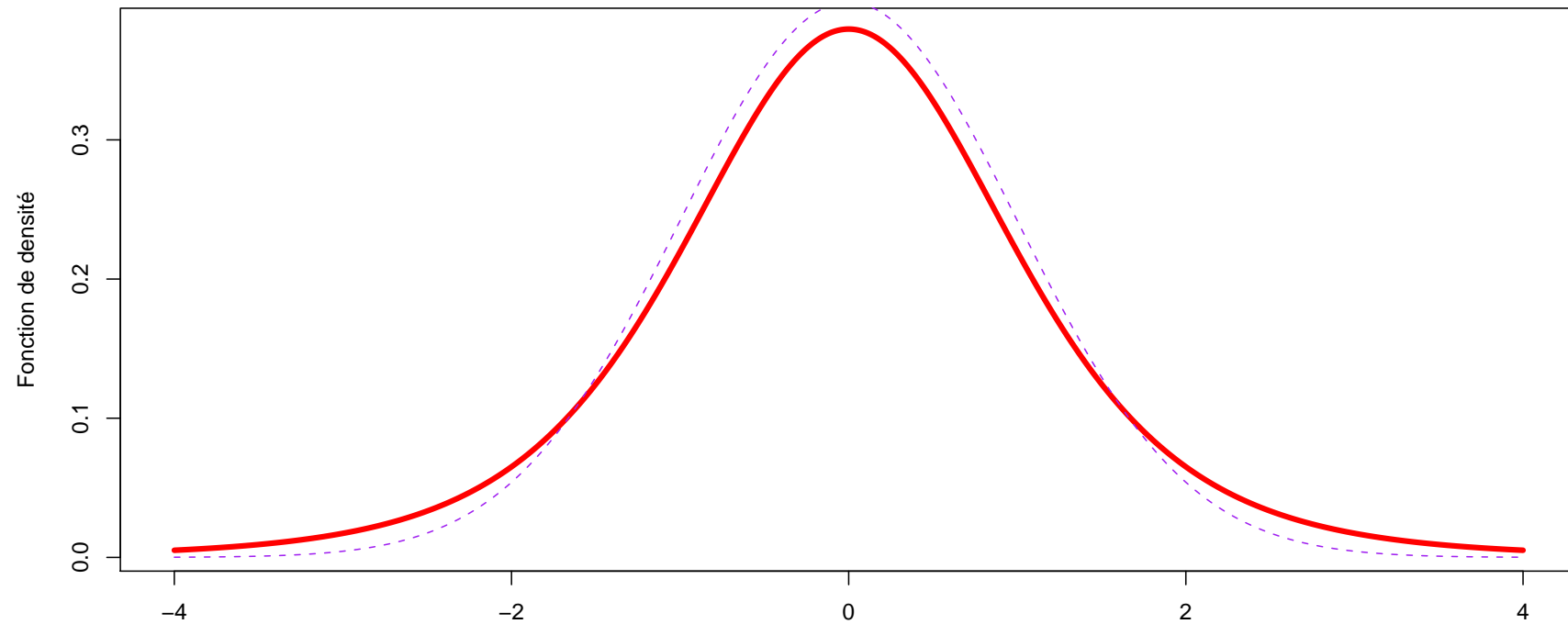


FIGURE 6 – Loi de Student, $St(\nu)$.

Quelques rappels sur les lois de probabilité

La loi de **Fischer** $\mathcal{F}(d_1, d_2)$, admet pour densité

$$f(x) = \frac{\left(\frac{d_1 x}{d_1 x + d_2}\right)^{d_1/2} \left(1 - \frac{d_1 x}{d_1 x + d_2}\right)^{d_2/2}}{x B(d_1/2, d_2/2)}$$

for $x \geq 0$ pour $d_1, d_2 \in \mathbb{N}$ et où B désigne la fonction beta.

Notons que $\mathbb{E}(X) = \frac{d_2}{d_2 - 2}$ pour $d_2 > 2$ et $Var(X) = \frac{2 d_2^2 (d_1 + d_2 - 2)}{d_1 (d_2 - 2)^2 (d_2 - 4)}$ pour $d_2 > 4$.

Si $X \sim \mathbb{F}(\nu_1, \nu_2)$, alors $\frac{1}{X} \sim F(\nu_2, \nu_1)$.

Si $X_1 \sim \chi^2(\nu_1)$ et $X_2 \sim \chi^2(\nu_2)$ sont indépendantes, alors $Y = \frac{X_1/\nu_1}{X_2/\nu_2} \sim \mathcal{F}(\nu_1, \nu_2)$.

Sous R, `df(x, df1, df2)`, `qf()` et `pf()` désigne respectivement la fonction de répartition, la fonction quantile et la densité.

Les lois conditionnelles

- **Mélange de deux lois de Bernoulli $\mathcal{B}(\Theta)$**

Soit Θ une variable aléatoire pouvant prendre deux valeurs, $\theta_1, \theta_2 \in [0, 1]$ avec probabilité p_1 et p_2 ($p_1 + p_2 = 1$). Supposons que

$$X|\Theta = \theta_1 \sim \mathcal{B}(\theta_1) \text{ et } X|\Theta = \theta_2 \sim \mathcal{B}(\theta_2).$$

La loi **non-conditionnelle** de X est alors

$$\mathbb{P}(X = x) = \sum_{\theta} \mathbb{P}(X = x, \Theta = \theta) = \sum_{\theta} \mathbb{P}(X = x|\Theta = \theta) \cdot \mathbb{P}(\Theta = \theta) = \mathbb{P}(X = x|\Theta = \theta_1) \cdot p_1 + \mathbb{P}(X = x|\Theta = \theta_2) \cdot p_2$$

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 0|\Theta = \theta_1) \cdot p_1 + \mathbb{P}(X = 0|\Theta = \theta_2) \cdot p_2 = (1 - \theta_1)p_1 + (1 - \theta_2)p_2 = 1 - \theta_1 p_1 - \theta_2 p_2$$

$$\mathbb{P}(X = 1) = \mathbb{P}(X = 1|\Theta = \theta_1) \cdot p_1 + \mathbb{P}(X = 1|\Theta = \theta_2) \cdot p_2 = \theta_1 p_1 + \theta_2 p_2$$

i.e. $X \sim \mathcal{B}(\theta_1 p_1 + \theta_2 p_2)$.

Notons que

$$\begin{aligned}\mathbb{E}(X) &= \theta_1 p_1 + \theta_2 p_2 \\ &= \mathbb{E}(X|\Theta = \theta_1)\mathbb{P}(\Theta = \theta_1) + \mathbb{E}(X|\Theta = \theta_2)\mathbb{P}(\Theta = \theta_2) = \mathbb{E}(\mathbb{E}(X|\Theta))\end{aligned}$$

$$\begin{aligned}\text{Var}(X) &= [\theta_1 p_1 + \theta_2 p_2][1 - \theta_1 p_1 - \theta_2 p_2] \\ &= \theta_1^2 p_1 + \theta_2^2 p_2 - [\theta_1 p_1 + \theta_2 p_2]^2 \\ &+ [\theta_1(1 - \theta_1)]p_1 + [\theta_2(1 - \theta_2)]p_2 \\ &= \mathbb{E}(X|\Theta = \theta_1)^2 \mathbb{P}(\Theta = \theta_1) + \mathbb{E}(X|\Theta = \theta_2)^2 \mathbb{P}(\Theta = \theta_2) \\ &- [\mathbb{E}(X|\Theta = \theta_1)\mathbb{P}(\Theta = \theta_1) + \mathbb{E}(X|\Theta = \theta_2)\mathbb{P}(\Theta = \theta_2)]^2 \\ &+ \text{Var}(X|\Theta = \theta_1)\mathbb{P}(\Theta = \theta_1) + \text{Var}(X|\Theta = \theta_2)\mathbb{P}(\Theta = \theta_2) \\ &= \underbrace{\mathbb{E}([\mathbb{E}(X|\Theta)]^2) - [\mathbb{E}(\mathbb{E}(X|\Theta))]^2}_{\text{Var}(\mathbb{E}(X|\Theta))} + \mathbb{E}(\text{Var}(X|\Theta))\end{aligned}$$

Les lois conditionnelles

- Mélange de deux lois de Poisson $\mathcal{P}(\Theta)$

Soit Θ une variable aléatoire pouvant prendre deux valeurs, $\theta_1, \theta_2 \in [0, 1]$ avec probabilité p_1 et p_2 ($p_1 + p_2 = 1$). Supposons que

$$X|\Theta = \theta_1 \sim \mathcal{P}(\theta_1) \text{ et } X|\Theta = \theta_2 \sim \mathcal{P}(\theta_2).$$

Alors

$$\mathbb{P}(X = x) = \frac{e^{-\theta_1} \theta_1^x}{x!} \cdot p_1 + \frac{e^{-\theta_2} \theta_2^x}{x!} \cdot p_2,$$

Il ne s'agit pas d'une loi *connue*.

Les lois conditionnelles

- Mélange continu de lois de Poisson $\mathcal{P}(\Theta)$

Soit Θ une variable aléatoire à valeurs dans $]0, \infty[$, de densité $\pi(\cdot)$. Supposons que

$$X|\Theta = \theta \sim \mathcal{P}(\theta) \text{ pour tout } \theta > 0$$

Alors

$$\mathbb{P}(X = x) = \int_0^{\infty} \mathbb{P}(X = x|\Theta = \theta)\pi(\theta)d\theta.$$

De plus,

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

$$\text{Var}(X) = \text{Var}(\mathbb{E}(X|\Theta)) + \mathbb{E}(\text{Var}(X|\Theta)) = \text{Var}(\Theta) + \mathbb{E}(\Theta) > \mathbb{E}(\Theta).$$

Remarque Les lois mélanges sont la base de la [statistique bayésienne](#).

Les lois conditionnelles, mélange et hétérogénéité

$$f(x) = f(x|\Theta = \theta_1) \times \mathbb{P}(\Theta = \theta_1) + f(x|\Theta = \theta_2) \times \mathbb{P}(\Theta = \theta_2).$$

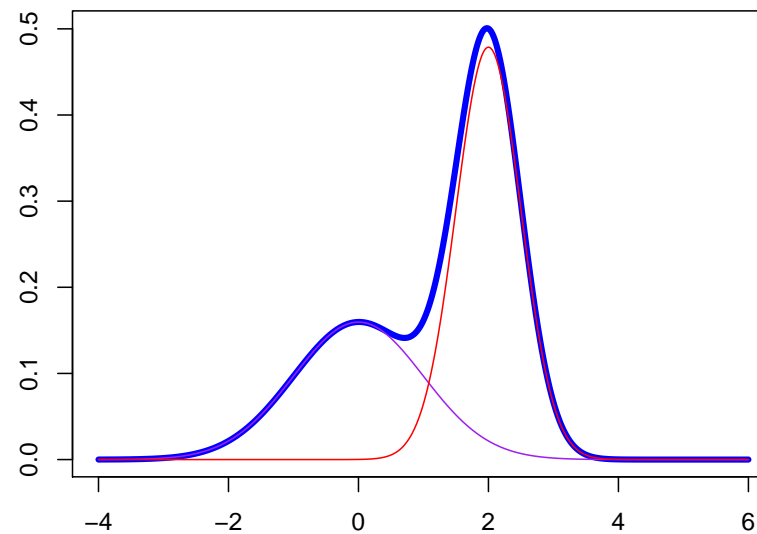


FIGURE 7 – Mélange de lois normales.

Les lois conditionnelles et l'hétérogénéité

La prise en compte de l'hétérogénéité est la base des méthodes économétriques.

On suppose en **économétrie** que Y suit une loi f_θ où θ est fonction de \mathbf{X} .

- **économétrie linéaire**, $Y|\mathbf{X} = \mathbf{x} \sim \mathcal{N}(\boldsymbol{\beta}'\mathbf{x}, \sigma^2)$.
- **économétrie des variables qualitatives (logit/probit)**, $Y|\mathbf{X} = \mathbf{x} \sim \mathcal{B}(\varphi[\boldsymbol{\beta}'\mathbf{x}])$.
- **régression Poissonienne**, $Y|\mathbf{X} = \mathbf{x} \sim \mathcal{P}(\exp[\boldsymbol{\beta}'\mathbf{x}])$.

E.g. $Y|X_1 = \text{homme} \sim \mathcal{B}(p_h)$ et $Y|X_1 = \text{femme} \sim \mathcal{B}(p_f)$.

Un peu de convergence

La suite (X_n) converge presque sûrement vers X , noté $X_n \xrightarrow{p.s.} X$, si

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \text{ pour tout } \omega \in A,$$

où $\mathbb{P}(A) = 1$. On dit aussi que (X_n) converge vers X avec probabilité 1 - on parle également de convergence trajectorielle. Notons que $X_n \xrightarrow{p.s.} X$ si et seulement si

$$\forall \varepsilon > 0, \mathbb{P}(\limsup \{|X_n - X| > \varepsilon\}) = 0.$$

Il est aussi possible de contrôler les accroissements de la suite (X_n) : soit (ε_n) telle que $\sum_{n \geq 0} \mathbb{P}(|X_n - X| > \varepsilon_n) < \infty$ où $\sum_{n \geq 0} \varepsilon_n < \infty$, alors (X_n) converge presque sûrement vers X .

Un peu de convergence

La suite (X_n) converge dans L^p vers X , ou en moyenne d'ordre p , noté $X_n \xrightarrow{L^p} X$, si

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

Pour $p = 1$ on parle de convergence en moyenne, et pour $p = 2$, on parle de convergence en moyenne quadratique.

Supposons que $X_n \xrightarrow{p.s.} X$ et qu'il existe une variable aléatoire Y telle que pour $n \geq 0$, $|X_n| \leq Y$ \mathbb{P} -presque sûrement, avec $Y \in L^p$, alors $X_n \in L^p$ et $X_n \xrightarrow{L^p} X$ (*théorème dit de la convergence dominée*).

Un peu de convergence

La suite (X_n) converge en probabilité vers X , noté $X_n \xrightarrow{\mathbb{P}} X$, si

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

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ (ce résultat reste toutefois vrai pour des fonctions $f : \mathbb{R}^p \rightarrow \mathbb{R}^q$) une application continue, si $X_n \xrightarrow{\mathbb{P}} X$ alors $f(X_n) \xrightarrow{\mathbb{P}} f(X)$.

De plus, si $X_n \xrightarrow{p.s.} X$ ou si $X_n \xrightarrow{L^1} X$ alors $X_n \xrightarrow{\mathbb{P}} X$. Enfin, si $X_n \xrightarrow{\mathbb{P}} X$, il existe une sous-suite (X_{n_k}) qui converge vers X presque sûrement.

Attention $X_n \xrightarrow{\mathbb{P}} c$ n'implique pas forcément $\mathbb{E}(X_n) \rightarrow c$

Une condition suffisante pour que $X_n \xrightarrow{\mathbb{P}} a$ est que

$$\lim_{n \rightarrow \infty} \mathbb{E}X_n = a \text{ et } \lim_{n \rightarrow \infty} \text{Var}(X_n) = 0$$

Un peu de convergence

- Loi forte des grands nombres

On suppose que les variables X_i sont i.i.d. et qu'elles admettent une moyenne $\mu = \mathbb{E}(X_i)$. Alors $S_n \xrightarrow{p.s.} \mu$ lorsque $n \rightarrow \infty$.

- Loi faible des grands nombres

On suppose que les variables X_i sont i.i.d. et qu'elles admettent une moyenne $\mu = \mathbb{E}(X_i)$ finie. Alors $S_n = (1/n) \sum_{i=1}^n X_i$, alors $S_n \xrightarrow{\mathbb{P}} \mu$ quand $n \rightarrow +\infty$.

Un peu de convergence

(iv) La suite (X_n) converge en loi vers X , noté $X_n \xrightarrow{\mathcal{L}} X$, si pour toute fonction f continue bornée,

$$\lim_{n \rightarrow \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)).$$

La convergence en loi est équivalente à la convergence (simple) des fonctions de répartition : $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si pour tout $t \in \mathbb{R}$ où F_X est continue

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

Un peu de convergence

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ (ce résultat reste toutefois vrai pour des fonctions $f : \mathbb{R}^p \rightarrow \mathbb{R}^q$) une application continue, si $X_n \xrightarrow{\mathcal{L}} X$ alors $f(X_n) \xrightarrow{\mathcal{L}} f(X)$. De plus, si $X_n \xrightarrow{\mathbb{P}} X$ alors $X_n \xrightarrow{\mathcal{L}} X$ (la réciproque n'étant vraie que si (X_n) converge en loi vers une constante).

○ Théorème central limite

Soient $X_1, X_2 \dots$ i.i.d. de moyenne μ et de variance σ^2 , alors :

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} = \sqrt{n} \left(\frac{S_n - \mu}{\sigma} \right) \xrightarrow{\mathcal{L}} X \text{ où } X \sim \mathcal{N}(0, 1)$$

De la convergence aux approximations

Proposition 28. *Soit (X_n) une suite i.i.d. de variables $\mathcal{B}(n, p)$. Alors si $np \rightarrow \infty$, $X_n \xrightarrow{\mathcal{L}} X$ où $X \sim \mathcal{P}(\lambda)$, et $\lambda = np$*

Démonstration. Ce résultat découle de l'approximation suivante,

$$\binom{n}{k} p^k [1 - p]^{n-k} \approx \exp[-np] \frac{[np]^k}{k!}$$

□

La loi de Poisson $\mathcal{P}(np)$ est une approximation de la loi Binomiale $\mathcal{B}(n, p)$ pour n suffisamment grand, avec $np \rightarrow \infty$ (et donc p petit).

En pratique l'approximation est valide pour $n > 30$ et $np < 5$.

De la convergence aux approximations

Proposition 29. *Soit (X_n) une suite i.i.d. de variables $\mathcal{B}(n, p)$. Alors si $np \rightarrow \infty$, $[X_n - np]/\sqrt{np(1-p)} \xrightarrow{\mathcal{L}} X$ où $X \sim \mathcal{N}(0, 1)$.*

En pratique l'approximation est valide pour $n > 30$ et $np > 5$ et $n(1-p) > 5$.

La loi de Normale $\mathcal{N}(np, np(1-p))$ est une approximation de la loi Binomiale $\mathcal{B}(n, p)$ pour n suffisamment grand, avec $np, n(1-p) \rightarrow \infty$.

De la convergence aux approximations

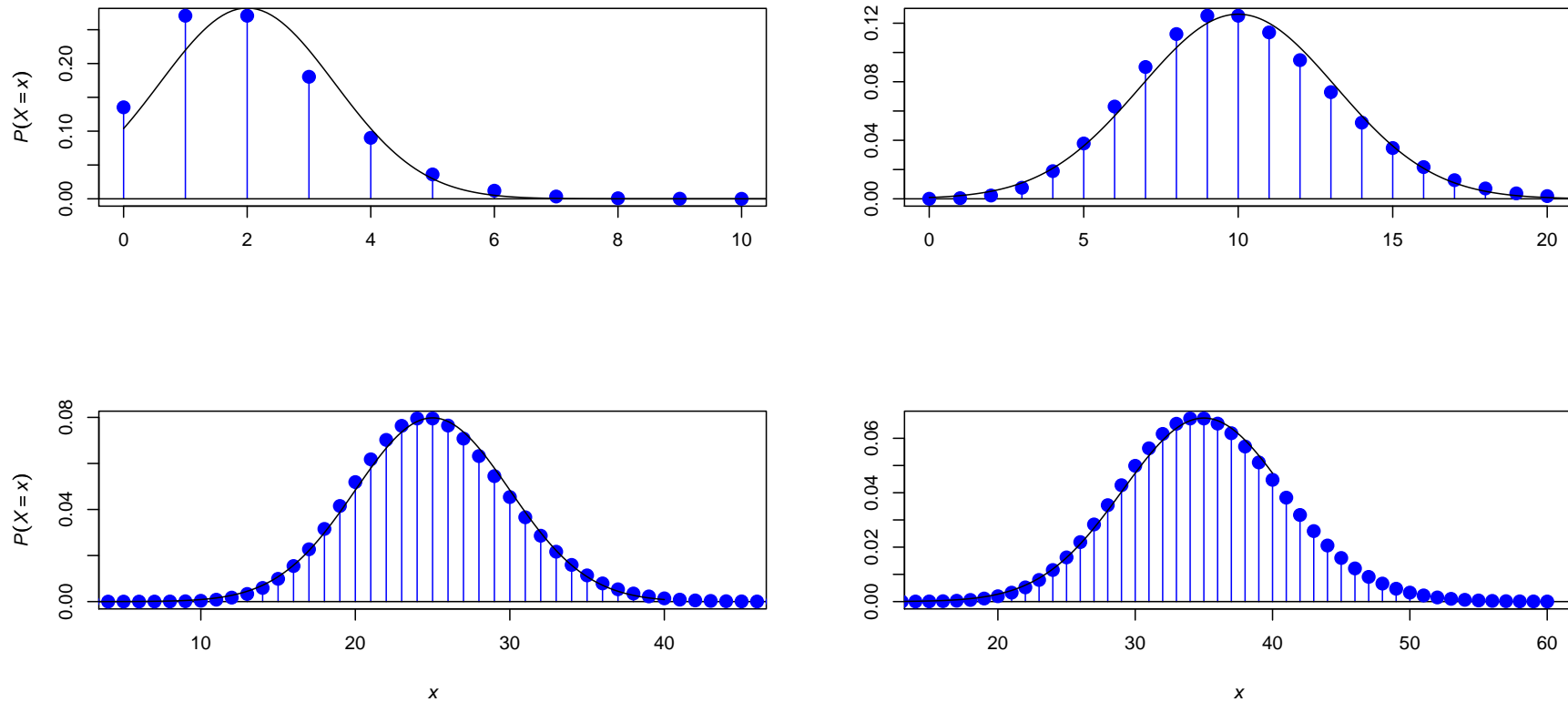


FIGURE 8 – Approximation de la de Poisson par une loi normale.

Lois de transformations de variables aléatoires

Soit X une v.a. absolument continue de densité $f(x)$, on cherche à connaître la densité de probabilité de $Y = \phi(X)$.

Proposition 30. *si la fonction ϕ est dérivable et bijective : La variable Y admet une densité g telle que*

$$g(y) = \frac{f(\phi^{-1}(y))}{\phi'(\phi^{-1}(y))}$$

Exemple de transformation de variables aléatoires

Proposition 31. *Soit X une variable aléatoire continue de fonction de répartition F , i.e. $F(x) = \mathbb{P}(X \leq x)$. Alors $Y = F(X)$ suit une loi uniforme sur $[0, 1]$.*

Proposition 32. *Soit X une variable aléatoire uniforme sur $[0, 1]$ et F une fonction de répartition. Alors $Y = F^{-1}(X)$ admet pour fonction de répartition F .*

Remarque Cette propriété est la base des méthodes de [Monte Carlo](#).

Lois de transformations de variables aléatoires

Soit (X, Y) un couple de variables aléatoires continues, la loi du couple admet une densité de probabilité $f(x, y)$. Soit $(U, V) = \phi(X, Y)$. Si on note J_ϕ le Jacobien associé, i.e.

$$J_\phi = \left| \det \begin{pmatrix} \partial U / \partial X & \partial V / \partial X \\ \partial U / \partial Y & \partial V / \partial Y \end{pmatrix} \right|$$

alors la loi de (U, V) est donnée par la densité :

$$g(u, v) = \frac{1}{J_\phi} f(\phi^{-1}(u, v))$$

Quelques exercices ?

Soit (U, V) un couple de variables aléatoires indépendantes, uniformément distribuées sur $[0, 1]$ respectivement. Montrer que les variables

$$\begin{cases} X = \sqrt{-2 \log U} \cos(2\pi V) \\ Y = \sqrt{-2 \log U} \sin(2\pi V) \end{cases}$$

sont indépendantes et de même loi $\mathcal{N}(0, 1)$.

si h est une fonction borélienne, si l'on peut écrire

$$\mathbb{E}(h(X)) = \int h(x) f(x) dx = \int yg(y) dy$$

alors g sera la densité de $h(X)$.

Soit h une application borélienne de \mathbb{R}^2 dans \mathbb{R} .

$$\mathbb{E}(h(X, Y)) = \mathbb{E}\left(h\left(\sqrt{-2 \log U} \cos(2\pi V), \sqrt{-2 \log U} \sin(2\pi V)\right)\right)$$

Considérons le changement de variable

$T : (u, v) \rightarrow (x, y) = (\sqrt{-2 \log u} \cos(2\pi v), \sqrt{-2 \log u} \sin(2\pi v))$, s'inversant en

$$T^{-1}(x, y) = (u, v) = \left(\exp \left[-\frac{x^2 + y^2}{2} \right], \frac{1}{2\pi} \arctan \frac{y}{x} \right)$$

dont le Jacobien est

$$J = \begin{vmatrix} -x \exp \left(-\frac{x^2 + y^2}{2} \right) & -y \exp \left(-\frac{x^2 + y^2}{2} \right) \\ -\frac{1}{2\pi} \frac{y}{x^2 + y^2} & \frac{1}{2\pi} \frac{x}{x^2 + y^2} \end{vmatrix} = \frac{-1}{2\pi} \exp \left(-\frac{x^2 + y^2}{2} \right) \quad (\neq 0).$$

(ce calcul permet de vérifier que le changement de variable correspond effectivement à un C^1 -difféomorphisme de $[0, 1]$ dans \mathbb{R}^2). D'où finalement,

$$\begin{aligned} \mathbb{E}(h(X, Y)) &= \mathbb{E} \left(h \left(\sqrt{-2 \log U} \cos(2\pi V), \sqrt{-2 \log U} \sin(2\pi V) \right) \right) \\ &= \int_{\mathbb{R}^2} h(x, y) \frac{1}{2\pi} \exp \left(-\frac{x^2 + y^2}{2} \right) dx dy. \end{aligned}$$

Aussi, (X, Y) admet pour densité

$$f(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \times \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right).$$

Aussi, les variables X et Y sont indépendantes, et de même loi $\mathcal{N}(0, 1)$.

Lois de transformations de variables aléatoires

Nous avons noté que $\mathbb{E}(g(X)) \neq g(\mathbb{E}(X))$, ormis dans le cas où g est une application linéaire. L'inégalité de Jensen permet d'avoir des résultats dans certains cas,

Proposition 33. *Soit g une fonction convexe, alors $\mathbb{E}(g(X)) \geq g(\mathbb{E}(X))$.*