PROBABILITÉS & STATISTIQUE

a Crash Course

par
Anthony Phan

Département de Mathématiques
Boulevard Marie et Pierre Curie,
Téléport 2,
BP 30179, F-86962 Chasseneuil–Futuroscope cedex.

Ce texte reprend les notes d'un cours effectué en première année de l'ÉSIP (École Supérieure d'Ingénieurs de Poitiers), 2008–09, avec un volume de 8 heures de cours et 10 heures de travaux dirigés. Évidemment, une partie des notions enseignées n'auront été introduites qu'en travaux dirigés, et ce, au détriment du temps consacré aux exercices.

TABLE DES MATIÈRES

Chapitre premier

ÉLÉMENTS DE PROBABILITÉS

Le propos de ce chapitre est seulement d'exposer les bases du langage du Calcul des Probabilités qui pourraient se montrer utiles à un utilisateur occasionnel de la Statistique. La théorie dite classique ou moderne est plus complexe elle se fonde largement sur l'intégration abstraite de Lebesgue.

De nombreuses remarques sont faites. Elles donnent souvent des précisions quant aux différents énoncés et peuvent être laissées en seconde lecture.

1. **Formalisme simplifié des Probabilités**

1.1. Généralités

Définition 1. — Soit E un ensemble non vide. Une *mesure de probabilité sur* E est une application

$$\begin{aligned} \mu : \mathcal{P}(E) &\longrightarrow [0,1] \\ B \subset E &\longmapsto \mu(B) \end{aligned}$$

vérifiant :

(i) $\mu(E) = 1$;

(ii) si $A, B \subset E$, A et B disjoints ($A \cap B = \varnothing$), alors $\mu(A \cup B) = \mu(A) + \mu(B)$ (additivité).

Le couple (E, μ) est alors appelé *espace probabilisé.*

Remarques. — *a*) La notation $\mathcal{P}(E)$ désigne l'ensemble des parties de E. On rappelle que lorsque E est fini de cardinal n, alors $\mathcal{P}(E)$ est fini de cardinal 2^n. Si E est infini, $\mathcal{P}(E)$ l'est aussi (et de cardinal strictement supérieur à celui de E).

b) La définition moins naïve d'une mesure de probabilité nécessite l'introduction de la notion de *tribu* ou *σ-algèbre*. Une tribu $\mathcal{E}$ sur E est un ensemble de parties de E contenant $\varnothing$, E, stable par passage au complémentaire et stable par réunions et intersections de familles finies ou dénombrables de ses éléments. Les tribus triviale (ou grossière) $\mathcal{E}_{\text{triviale}} = \{\varnothing, E\}$ et discrète $\mathcal{E}_{\text{discrète}} = \mathcal{P}(E)$ sont des tribus sur E.

c) Pour des questions techniques ou des exigences de modélisation, il n'est pas toujours possible ou souhaitable de considérer des mesures de probabilité définies sur $\mathcal{P}(E)$ tout entier ; on se restreint à les définir sur une tribu $\mathcal{E}$, sous-tribu de $\mathcal{P}(E)$; dans ce cadre $\mu(B)$ n'a un sens que si $B \in \mathcal{E}$ et B est alors dit être un sous-ensemble mesurable, ou probabilisable, de E. Un couple $(E, \mathcal{E})$, où E est un ensemble et $\mathcal{E}$ est une tribu sur E est appelé espace mesurable ou espace probabilisable. Si μ est une mesure de probabilité sur $(E, \mathcal{E})$, le triplet $(E, \mathcal{E}, \mu)$ est un espace probabilisé (ou de probabilité).

d) Le support d'une mesure de probabilité μ sur un ensemble E est la plus petite partie $\operatorname{supp}(\mu)$ (de type convenable) de E telle que $\mu(\operatorname{supp}(\mu)) = 1$. Cette notion pourra être évoquée de manière évasive dans la suite.

Théorème 1 (propriétés élémentaires). — *Soit* (E, μ) *un espace probabilisé.*

(i) *On a* $\mu(\varnothing) = 0$.

(ii) *Pour tout* $B \subset E$*, en notant* B^c *le complémentaire de* B *dans* E*, on a*

$$\mu(B^c) = 1 - \mu(B) \qquad \textit{(passage au complémentaire).}$$

(iii) *Si* $A, B \subset E$*,* $A \subset B$*, alors* $\mu(A) \leqslant \mu(B)$ *(monotonie).*

(iv) *Si* $A_1, \ldots, A_n \subset E$ *sont des sous-ensembles disjoints de* E *(pour tous* $k, \ell \in \{1, \ldots, n\}$*,* $k \neq \ell$*,* $A_k \cap A_\ell = \varnothing$*), alors*

$$\mu(A \cup \ldots \cup A_n) = \mu(A_1) + \cdots + \mu(A_n) \qquad \textit{(additivité finie).}$$

(v) *Si* $A_1, \ldots, A_n \subset E$ *sont des sous-ensembles quelconques de* E*, alors*

$$\mu(A \cup \ldots \cup A_n) \leqslant \mu(A_1) + \cdots + \mu(A_n) \qquad \textit{(inégalité de Boole).}$$

Démonstration. — (i) On a $E = \varnothing \cup E$, et $\varnothing \cap E = \varnothing$, donc $\varnothing$ et E sont disjoints de réunion totale. Ainsi $1 = \mu(E) = \mu(\varnothing \cup E) = \mu(\varnothing) + \mu(E) = \mu(\varnothing) + 1$. En simplifiant par 1 (qui est un nombre fini), on obtient $\mu(\varnothing) = 0$.

(ii) Pour $B \subset E$, on a $B \cup B^c = E$ et $B \cap B^c = \varnothing$, donc $1 = \mu(E) = \mu(B \cup B^c) = \mu(B) + \mu(B^c)$, donc $\mu(B^c) = 1 - \mu(B)$.

(iii) Soient $A, B \subset E$. On a $A \cup B = (A \backslash B) \cup (B \backslash A) \cup (A \cap B)$ union d'ensembles disjoints, donc $\mu(A \cup B) = \mu(A \backslash B) + \mu(B \backslash A) + \mu(A \cap B)$. or $A = (A \backslash B) \cup (A \cap B)$ est aussi une union d'ensembles disjoints, et ainsi $\mu(A) = \mu(A \backslash B) + \mu(A \cap B)$. De même, $\mu(B) = \mu(B \backslash A) + \mu(A \cap B)$. Finalement,

$$\begin{aligned} \mu(A \cup B) &= \big(\mu(A) - \mu(A \cap B)\big) + \big(\mu(B) - \mu(A \cap B)\big) + \mu(A \cap B) \\ &= \mu(A) + \mu(B) - \mu(A \cap B). \end{aligned}$$

(iv) Si $A \subset B \subset E$, on a $B = A \cup (B \backslash A)$ réunion disjointe, donc $\mu(B) = \mu(A) + \mu(B \backslash A) \geqslant \mu(A)$.

(v) L'additivité finie s'obtient par une récurrence immédiate à partir de la propriété d'additivité de la mesure de probabililité μ.

(vi) D'après (iv), $\mu(A \cap B) = \mu(A) + \mu(B) - \mu(A \cap B) \leqslant \mu(A) + \mu(B)$. L'extension de cette inégalité s'obtient par une récurrence immédiate. □

Remarque. — Le véritable axiome d'additivité relatif aux mesures (de probabilité) est l'axiome d'additivité dénombrable : si $A_1, \ldots, A_n, \ldots$ est une suite finie ou dénombrable de sous-ensembles (mesurables) de E disjoints, alors

$$\mu\Big(\bigcup_n A_n\Big) = \mu(A_1 \cup \ldots \cup A_n \cup \ldots) = \mu(A_1) + \cdots + \mu(A_n) + \cdots = \sum_n \mu(A_n).$$

Cette axiome est plus fort et donc plus restrictif que celui donné dans la définition. Il est nécessaire de le postuler si on désire obtenir des résultats asymptotiques ou limite, mais il ne sert que rarement pour des calculs élémentaires.

Lorsqu'un espace probabilisé modélise une *expérience aléatoire*, telle que par exemple le tirage d'une boule dans une urne, on le note généralement $(\Omega, \mathbb{P})$. L'ensemble Ω est l'univers des possibles, chacun de ses éléments $\omega \in \Omega$ représente ce qu'il peut se passer, c'est-à-dire une éventualité, les parties $A \subset \Omega$ sont appelées événements, c'est-à-dire des collections d'éventualités ayant un trait commun (appartenir à A !), et auxquelles on peut attribuer une probabilité de réalisation $\mathbb{P}(A)$.

Dans ce cadre (simplifié), si E est un ensemble non vide, les applications $X : \Omega \to E$ sont appelées variables aléatoires. Lorsque $E \subset \mathbb{R}$, elles sont dites réelles ; lorsque $E \subset \mathbb{R}^n$, elles sont dites vectorielles. Une variable aléatoire correspond à une mesure effectuée lors d'une expérience aléatoire, la valeur mesurée $X(\omega)$ dépend de l'éventualité $\omega \in \Omega$ rencontrée.

Définition 2 et théorème. — Soient $(\Omega, \mathbb{P})$ un espace probabilisé, E un ensemble non vide et $X : \Omega \to E$ une variable aléatoire. L'application

$$\begin{aligned} P_X : \mathcal{P}(E) &\longrightarrow [0,1] \\ B \subset E &\longmapsto P_X(B) = \mathbb{P}\big(X^{-1}(B)\big) = \mathbb{P}\{X \in B\} \end{aligned}$$

est une mesure de probabilité sur E appelée loi de la variable aléatoire X (ou encore, image directe de la mesure de probabilité $\mathbb{P}$ par l'application X).

Démonstration. — Tout repose sur les propriétés de l'image réciproque d'ensembles par une application :

(i) $P_X(E) = \mathbb{P}(X^{-1}(E)) = \mathbb{P}(\Omega) = 1$;

(ii) si $B', B'' \subset E$ sont tels que $B' \cap B'' = \varnothing$, alors $X^{-1}(B' \cup B'') = X^{-1}(B') \cup X^{-1}(B'')$ et $X^{-1}(B') \cap X^{-1}(B'') = X^{-1}(B' \cap B'') = X^{-1}(\varnothing) = \varnothing$; donc $P_X(B' \cup B'') = \mathbb{P}(X^{-1}(B' \cup B'')) = \mathbb{P}(X^{-1}(B') \cup X^{-1}(B'')) = \mathbb{P}(X^{-1}(B')) + \mathbb{P}(X^{-1}(B'')) = P_X(B') + P_X(B'')$. □

Définition 3. — Soit $(\Omega, \mathbb{P})$ un espace probabilisé.

(i) Si $E_1, \ldots, E_n$ sont des ensembles non vides et $X_i : \Omega \to E_i$, $i = 1, \ldots, n$, sont des variables aléatoires, la loi de $X = (X_1, \ldots, X_n) : \Omega \to E_1 \times \cdots \times E_n$ est appelée *loi conjointe* de $X_1, \ldots, X_n$. Pour $i = 1, \ldots, n$, la loi de X_i est appelée i-ième *loi marginale* de $X = (X_1, \ldots, X_n)$.

(ii) Si $E_1, \ldots, E_n$ sont des ensembles non vides et $X_i : \Omega \to E_i$, $i = 1, \ldots, n$, sont des variables aléatoires, les variables aléatoires $X_1, \ldots, X_n$ sont *indépendantes* si et seulement si la loi de $X = (X_1, \ldots, X_n)$ est la *loi produit* des lois de $X_1, \ldots, X_n$: pour tout $B = B_1 \times \cdots \times B_n$, $B_i \subset E_i$, $i = 1, \ldots, n$,

$$P_X(B) = P_{X_1}(B_1) \times \cdots \times P_{X_n}(B_n).$$

(iii) Une suite infinie de variables aléatoires $X_i : \Omega \to E_i$, $1 \leqslant i < \infty$, est une *suite de variables aléatoires indépendantes* si et seulement si, pour tout $n \geqslant 1$, les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes.

Remarque. — Pour la définition d'indépendance de variables aléatoires, nous détaillons l'écriture de la propriété requise : pour tout $B = B_1 \times \cdots \times B_n$, $B_i \subset E_i$, $i = 1, \ldots, n$,

$$\begin{aligned} P_X(B) &= \mathbb{P}\{X \in B\} \\ &= \mathbb{P}\{(X_1, \ldots, X_n) \in B_1 \times \cdots \times B_n\} \\ &= \mathbb{P}\{(X_1 \in B_1, \ldots, X_n \in B_n\} \\ \text{(indépendance)} &= \mathbb{P}\{X_1 \in B_1\} \times \cdots \times \mathbb{P}\{X_n \in B_n\} \\ &= P_{X_1}(B_1) \times \cdots \times P_{X_n}(B_n), \end{aligned}$$

où on notera bien quelle égalité rend compte de l'hypothèse d'indépendance, les autres n'étant que des égalités entre notations.

Exercice 1. — Soit $\Omega, \mathbb{P})$ un espace probabilisé. Pour $A \subset \Omega$ un événement, on définit la fonction indicatrice $\mathbb{1}_A$ de A par

$$\begin{aligned} \mathbb{1}_A : \Omega &\longrightarrow \{0,1\} \\ \omega &\longmapsto \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon.} \end{cases} \end{aligned}$$

(i) Montrer que $\mathbb{1}_A$ et $\mathbb{1}_{A'}$ sont des variables aléatoires indépendantes si et seulement si $\mathbb{P}(A \cap A') = \mathbb{P}(A) \times \mathbb{P}(A')$ (indépendance de deux événements).

(ii) Énoncer une caractérisation similaire dans le cas de n événements $A_1, \ldots, A_n$. (*Avertissement.* — Ce n'est pas une question naïve.)

Exercice 2. — Soient $n \geqslant 1$, $\Omega = \{0,1\}^n$, $\mathbb{P} : A \subset \Omega \mapsto \operatorname{Card} A / \operatorname{Card} \Omega$ et pour $i \in \{1, \ldots, n\}$, $X_i : \Omega \to \{0,1\}$, $\omega = (\omega_1, \ldots, \omega_n) \in \Omega \mapsto X_i(\omega) = \omega_i$.

(i) Montrer que $(\Omega, \mathbb{P})$ est un espace probabilisé.

(ii) Montrer que les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes, de même loi, loi à expliciter.

Correction. — (i) Non disponible.

(ii) Non disponible.

1.2. Variables aléatoires réelles

Lorsque $X : \Omega \to \mathbb{R}, \mathbb{C}, \ldots$ est une variable aléatoire à valeurs dans $\mathbb{R}, \mathbb{C}, \ldots$, on peut envisager la moyenne ou *espérance* de X, ou encore si $X : \Omega \to E$ et $f : E \to \mathbb{R}, \mathbb{C}, \ldots$, la moyenne ou *espérance* de la variable aléatoire $f(X) : \Omega \to \mathbb{R}, \mathbb{C}, \ldots$: c'est simplement la somme des valeurs que peut prendre cette variable pondérées respectivement par les probabilités d'être prises. En fait, on considère respectivement la somme pondérée des valeurs positives et celle des valeurs négatives de la variable, puis on prend la différence des deux sommes. Elle peut être non définie (situations du type « $\infty - \infty$ »), définie sans être finie (situations du type « $\infty -$fini » ou « fini$- \infty$ »), ou définie et finie (dans ce dernier cas, la variable X est dite intégrable). *Les méthodes pour calculer de telles moyennes seront précisées dans des cas particuliers.*

Définition 4 et notations. — Soient $(\Omega, \mathbb{P})$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ est une variable aléatoire réelle de loi P_X.

(i) La moyenne, lorsqu'elle est définie, des valeurs observées de la variable aléatoire X est appelé *espérance* de X et notée $\mathbb{E}[X]$, c'est un nombre réel. Lorsque l'espérance de X est finie, ont dit que X est intégrable.

(ii) Si l'espérance $\mathbb{E}[X]$ est définie, $(X - \mathbb{E}[X])^2 : \Omega \to \mathbb{R}$ est une variable aléatoire réelle positive dont l'espérance est appelée *variance* de la variable aléatoire X et notée $\operatorname{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$. (On peut montrer par linéarité de la moyenne qu'on a aussi $\operatorname{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.) La racine carrée de la variance est appelée écart-type et notée $\sigma_X = \sqrt{\operatorname{Var}(X)}$.

(iii) Pour tout $\theta \in \mathbb{R}$, $\mathrm{e}^{\mathrm{i}\theta X} : \Omega \to \mathbb{S}^1 \subset \mathbb{C}$ est une variable aléatoire complexe bornée. Son espérance est notée

$$\varphi_X(\theta) = \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta X}\big].$$

L'application $\varphi_X : \mathbb{R} \to \mathbb{C}$, qui est continue, bornée en module par 1, $\varphi_X(0) = 1$, est appelée *fonction caractéristique* de la loi de la variable aléatoire X. On montre, en effet, que deux variables aléatoires réelles X et Y ayant même fonction caractéristique ($\varphi_X = \varphi_Y$) ont même loi ($P_X = P_Y$).

(iv) La fonction, appelée fonction de répartition de la loi de X,

$$\begin{aligned} F_X : \mathbb{R} &\longrightarrow [0,1] \\ x &\longmapsto F_X(x) = P_X(]-\infty, x]) = \mathbb{P}\{X \leqslant x\} \end{aligned}$$

est une fonction croissante, continue à droite, limitée à gauche vérifiant

$$\lim_{x \to -\infty} F_X(x) = 0 \qquad \text{et} \qquad \lim_{x \to +\infty} F_X(x) = 1.$$

Deux variables aléatoires réelles X et Y ayant même fonction de répartition ($F_X = F_Y$) ont même loi ($P_X = P_Y$).

Remarques. — *a*) Toutes les quantités et fonctions définies ci-dessus ne dépendent que de la loi P_X de la variable aléatoire X. Nous aurions pu les définir pour une mesure de probabilité μ sur $\mathbb{R}$ sans faire appel à une quelconque variable aléatoire. Ceci apparaîtra clairement lorsque les méthodes usuelles de calcul seront données.

b) L'espérance est bien sûr associée à la mesure de probabilité $\mathbb{P}$ dont on a muni Ω. Elle vérifie en particulier $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$ pour tout événement A : la variable aléatoire $\mathbb{1}_A : \Omega \to \{0,1\}$ prend la valeur 1 sur A et la valeur 0 sur A^c ; sa moyenne est donc $1 \times \mathbb{P}(A) + 0 \times \mathbb{P}(A^c) = \mathbb{P}(A)$. Une mesure de probabilité $\mathbb{P}'$ sur Ω différente de $\mathbb{P}$ donne lieu à une espérance $\mathbb{E}'$ différente de $\mathbb{E}$.

c) La notion de variance n'est pas toujours comprise, il peut être utile de détailler son rôle : si une variable aléatoire X s'exprime en mètres, sa variance s'exprime en mètres-carrés, et son écart-type en mètres. Une variance, ou un écart-type, « petite » assure qu'il est peu probable d'observer des valeurs éloignées de la moyenne $\mathbb{E}[X]$, lorsqu'elle est nulle, on a même $\mathbb{P}\{X = \mathbb{E}[X]\} = 1$, c'est-à-dire que X est constante (presque sûrement) égale à sa moyenne ; à l'opposé, une variance, ou un écart-type, « grande » assure qu'il est probable observer des valeurs de X assez éloignées de la valeur moyenne, ces valeurs sont *dispersées*. C'est pourquoi l'écart-type (et non la variance) sert d'indicateur de dispersion de la loi de la variable considérée.

d) La notion d'espérance a bien sûr un sens pour des variables aléatoires complexes (voir la définition de fonction caractéristique) et aussi pour des variables aléatoires vectorielles. Pour des variables aléatoires à valeurs dans $\mathbb{R}^n$, les notions de fonctions caractéristiques et de répartition s'étendent avec des arguments qui sont alors vectoriels : pour $X = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$, d'une part, pour $\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n$,

$$\varphi_X(\theta) = \mathbb{E}[\mathrm{e}^{\mathrm{i}\langle \theta, X\rangle}] = \mathbb{E}[\exp \mathrm{i}(\theta_1 X_1 + \cdots + \theta_n X_n)],$$

d'autre part, pour $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$F_X(x) = P_X(]-\infty, x_1] \times \cdots \times]-\infty, x_n]) = \mathbb{P}\{X_1 \leqslant x_1, \ldots, X_n \leqslant x_n\}.$$

Ces fonctions sont encore caractéristiques des lois.

e) Si X est une variable aléatoire réelle et $\alpha \in \mathbb{R}$, X^α, lorsque cela a un sens, est une variable aléatoire réelle et $\mathbb{E}[X^\alpha]$, lorsque c'est défini, est le moment d'ordre α de X. On dit que X est de puissance α-intégrable si $\mathbb{E}[|X|^\alpha] < \infty$.

f) Deux variables aléatoires X et Y peuvent avoir même loi sans être égales ou même définies sur un même espace probabilisé, l'égalité de leur loi ne signifiant que l'identité de la manière dont elles se répartissent. Si elles sont même loi, elles ont les mêmes moments ; dans certains cas, on peut avoir une réciproque, mais généralement il vaut mieux se fier aux fonctions caractéristiques ou aux fonctions de répartition.

Théorème 2 (propriétés usuelles de l'espérance). — *Soient $X, Y : \Omega \to \mathbb{R}$ deux variables aléatoires réelles intégrables.*

(i) *Si $\mathbb{P}\{X \geqslant 0\} = 1$, en particulier si $X \geqslant 0$, alors*

$$\mathbb{E}[X] \geqslant 0 \qquad \textit{(positivité)};$$

si de plus $\mathbb{E}[X] = 0$, *alors* $\mathbb{P}\{X = 0\} = 1$.

(ii) *Si* $\mathbb{P}\{X \leqslant Y\} = 1$, *en particulier si* $X \leqslant Y$, *alors*

$$\mathbb{E}[X] \leqslant \mathbb{E}[Y] \qquad \textit{(monotonie).}$$

(iii) *Pour tous* $\alpha, \beta, \gamma \in \mathbb{R}$, $\alpha X + \beta Y + \gamma : \Omega \to \mathbb{R}$ *est une variable aléatoire intégrable et on a*

$$\mathbb{E}[\alpha X + \beta Y + \gamma] = \alpha\, \mathbb{E}[X] + \beta\, \mathbb{E}[Y] + \gamma \qquad \textit{(linéarité).}$$

(iv) *Si pour* $p \geqslant 1$, $X^p : \Omega \to \mathbb{R}$ *est intégrable, alors pour tout* $1 \leqslant q \leqslant p$, $X^q : \Omega \to \mathbb{R}$ *est intégrable (remplacer* X *par* $|X|$ *pour des exposants non nécessairement entiers).*

(v) *Pour tout* $\alpha, \beta \in \mathbb{R}$, *on a* $\operatorname{Var}(\alpha X) = \alpha^2 \operatorname{Var}(X)$ *et* $\operatorname{Var}(X + \beta) = \operatorname{Var}(X)$. *De plus, si* X *et* Y *sont de carré intégrable, ou, plus généralement, si* $X \times Y$ *est intégrable,*

$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{cov}(X, Y),$$

où

$$\operatorname{cov}(X, Y) = \mathbb{E}\big[\big(X - \mathbb{E}[X]\big) \times \big(Y - \mathbb{E}[Y]\big)\big] = \mathbb{E}[X \times Y] - \mathbb{E}[X] \times \mathbb{E}[Y]$$

est la covariance de X *et de* Y.

(vi) *Si de plus,* X *et* Y *sont deux variables aléatoires indépendantes, intégrables, alors la variable aléatoire* $X \times Y : \Omega \to \mathbb{R}$ *est intégrable et on a*

$$\mathbb{E}[X \times Y] = \mathbb{E}[X] \times \mathbb{E}[Y].$$

On a alors $\operatorname{cov}(X, Y) = 0$ *— X et Y sont dites non corrélées — et*

$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y).$$

Remarques. — *a*) Les trois premières propriétés sont celles qu'on attend d'une moyenne. Elles se résument en disant que l'espérance est une forme linéaire (définie) positive. Si on ajoute $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$ pour tout événement A et une propriété de continuité (convergence monotone ou dominée), on caractérise complètement l'espérance par rapport à la mesure de probabilité $\mathbb{P}$.

b) La propriété (iv) est propre à l'intégration par rapport à une mesure de probabilité et peut se démontrer sans trop de difficulté en séparant les événements $\{|X| \leqslant 1\}$ et $\{|X| > 1\}$. La propriété (v) indique qu'il y a une structure de type euclidien sur l'ensemble des variables aléatoires de carré intégrables. Les inégalités classiques (Cauchy–Schwarz, ...) s'appliquent.

c) La propriété de produit liée à l'hypothèse d'indépendance est assez particulière. Ceux qui sont familiers avec l'Analyse savent que le produit de deux applications intégrables n'est généralement pas intégrable et qu'il n'y a pas de raison particulière pour que de plus l'intégrale du produit soit égal au produit des intégrales. Même avec une « cuisine probabiliste » ceci peut être en défaut : pour $\Omega =]0, 1]$ muni de la mesure de Lebesgue $\mathbb{P}(\mathrm{d}x) = \mathrm{d}x$, la variable aléatoire $X : x \in]0, 1] \mapsto x^{-1/2}$ est intégrable et $X \times X : x \in]0, 1] \mapsto 1/x$ ne l'est pas ; pour (Ω, P) donné avec $X : \Omega \to \{-1, 1\}$, $\mathbb{P}\{X = 1\} = \mathbb{P}\{X = -1\} = 1/2$, alors X est intégrable d'espérance nulle, $X^2 = 1$ est intégrable d'espérance 1 et ainsi $\mathbb{E}[X^2] \neq \mathbb{E}[X] \times \mathbb{E}[X]$.

d) La propriété (vi) est simplificatrice lorsqu'on calcule la variance d'une somme de variables aléatoires indépendantes, de carré intégrables, donc intégrables : soient $X_1, \ldots, X_n$ de telles variables ; on a en développant

$$\operatorname{Var}(X_1 + \cdots + X_n) = \sum_{k=1}^{n} \operatorname{Var}(X_k) + 2 \sum_{1 \leqslant k < \ell \leqslant n} \operatorname{cov}(X_k, X_\ell)\,;$$

par indépendance et donc non corrélation, on a $\mathrm{cov}(X_k, X_\ell) = 0$ pour $k \neq \ell$; et ainsi

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

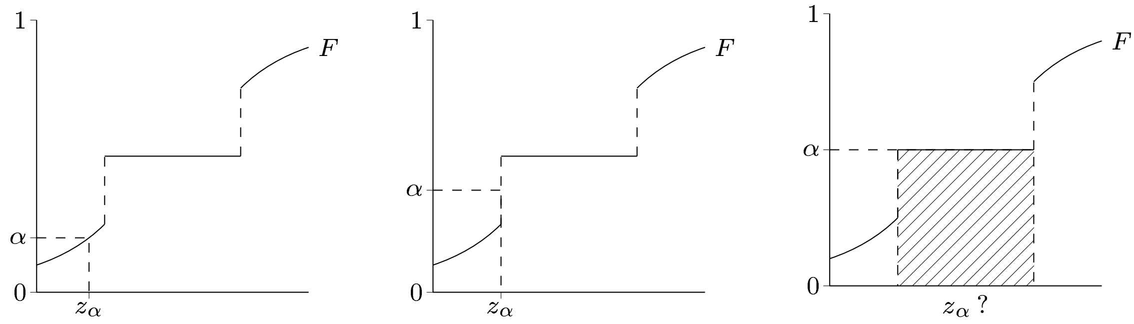
e) La propriété (vi) est utile lorsqu'on essaie de déterminer la loi d'une somme de variables aléatoires réelles indépendantes : la fonction caractéristique de cette somme est alors égale au produit des fonctions caractéristiques de chacun de ses termes, et si on sait reconnaître en l'expression obtenue la fonction caractéristique d'une loi, celle-ci est alors la loi de la somme.

Pour finir cette section, nous introduisons la notion de quantile d'une loi de probabilité sur $\mathbb{R}$.

DÉFINITION 5. — Soit μ une mesure de probabilité sur $\mathbb{R}$, par exemple loi d'une variable aléatoire réelle X, de fonction de répartition F. Soit $\alpha \in]0,1[$. On appelle *quantile d'ordre* α de μ, ou de la loi de X, tout réel $x_\alpha \in \mathbb{R}$ tel que

$$F(x_\alpha -) = \mu(]-\infty, x_\alpha[) = \mathbb{P}\{X < x_\alpha\} \leqslant \alpha \quad \text{et} \quad F(x_\alpha) = \mu(]-\infty, x_\alpha]) = \mathbb{P}\{X \leqslant x_\alpha\} \geqslant \alpha.$$

On ne peut écrire $x_\alpha = F^{-1}(\alpha)$ que s'il existe un et un seul $x_\alpha \in \mathbb{R}$ tel que $F(x_\alpha) = \alpha$, ce qui en pratique n'est pas toujours le cas. Rappelons que la fonction de répartition F est croissante, continue à droite et de limites 0 et 1 en, respectivement, $-\infty$ et $+\infty$. Les graphiques suivants illustrent les trois situations possibles en fonction de α :



FIGURES 1–3. — *Détermination des quantiles.*

Dans les deux premières situations, z_α est déterminé sans ambigüité. Pour la troisième, l'ensemble des valeur admissibles pour z_α est un intervalle de la forme $[a, b[$, on convient alors de prendre $z_\alpha = (a + b)/2$.

Les quantiles correspondants à certains ordres reçoivent une dénomination particulière :

- la *médiane* est le quantile d'ordre $1/2$ et est notée me ;
- les *quartiles* sont les quantiles d'ordre $k/4$ et sont notés q_k ;
- les *déciles* sont les quantiles d'ordre $k/10$;
- les *centiles* sont les quantiles d'ordre $k/100$;
- etc.

2. **Cadre discret**

2.1. GÉNÉRALITÉS

DÉFINITION 6. — (i) Un espace probabilisé discret est un espace probabilisé (E, μ) où E est un ensemble fini ou dénombrable, $E = \{x_1, \ldots, x_n, \ldots\}$. Dans ce cas, la mesure de probabilité μ est définie par la suite de ses valeurs sur les singletons : $\mu_n = \mu\{x_n\}$, $n = 1, \ldots$ (*Explication.* — Si $B \subset E$, B est la réunion disjointe de la famille au plus dénombrable des singletons qui le composent ; ainsi, par additivité dénombrable, $\mu(B) = \sum_{x_n \in B} \mu\{x_n\} = \sum_{x_n \in B} \mu_n$.)

(ii) Une variable aléatoire discrète est une application $X : \Omega \to E$ où $(\Omega, \mathbb{P})$ est un espace probabilisé et E un ensemble fini ou dénombrable. Alors, l'espace probabilisé (E, P_X) — P_X étant la loi de X — est un espace probabilisé discret. Plus généralement, une variable aléatoire $X : \Omega \to E$ est discrète s'il existe $B \subset E$ fini ou dénombrable tel que $\mathbb{P}\{X \in B\} = 1$, autrement dit si le support $\mathrm{supp}(P_X)$ de la loi de X est fini ou dénombrable.

(iii) Soient $(\Omega, \mathbb{P})$ un espace probabilisé, $X : \Omega \to E$ une variable aléatoire discrète, et $f : E \to \mathbb{R}$ une application. Alors $f(X) = f \circ X : \Omega \to \mathbb{R}$ est une variable aléatoire réelle discrète. Son espérance, ou moyenne, $\mathbb{E}[f(X)]$ est définie par

$$\mathbb{E}[f(X)] = \sum_{x \in E} f(x) \times P_X\{x\} = \sum_{x \in E} f(x) \times \mathbb{P}\{X = x\}$$

lorsque cette somme a un sens.

Pratique. — Soit $X : \Omega \to E \subset \mathbb{R}$ une variable aléatoire réelle discrète.

- Pour tout $B \subset E$,

$$P_X(B) = \mathbb{P}\{X \in B\} = \sum_{x \in B} \mathbb{P}\{X = x\} = \sum_{x \in B} P_X\{x\}.$$

- Si la somme, ou série, $\sum_{x \in E} x\, \mathbb{P}\{X = x\}$ est définie, en particulier si elle est absolument convergente, alors l'espérance de X est définie, et égale à cette somme ; si elle est finie, la variable aléatoire X est dite intégrable.
- Si la variable aléatoire X est intégrable d'espérance $\mathbb{E}[X]$, la variance de X est définie et égale à

$$\sum_{x \in E} \big(x - \mathbb{E}[X]\big)^2\, \mathbb{P}\{X = x\} = \sum_{x \in E} x^2\, \mathbb{P}\{X = x\} - \mathbb{E}[X]^2.$$

  La variance est finie si et seulement si X est de carré intégrable.
- La fonction caractéristique $\varphi_X : \mathbb{R} \to \mathbb{C}$ de la loi de X est définie et

$$\varphi_X(x) = \sum_{x \in E} \mathrm{e}^{\mathrm{i}\theta x}\, \mathbb{P}\{X = x\}, \qquad \text{pour tout } \theta \in \mathbb{R}.$$

- La fonction de répartition de la loi de X est une fonction en escalier ne présentant des sauts qu'aux points de E, sauts de hauteur $\mathbb{P}\{X = x\}$, $x \in E$.
- Soient deux variables aléatoires discrètes $X : \Omega \to E$, $Y : \Omega \to F$, où E et F sont deux ensembles finis ou dénombrables quelconques. Les variables X et Y sont indépendantes si et seulement si

$$\mathbb{P}\{X = x, Y = y\} = \mathbb{P}\{X = x\} \times \mathbb{P}\{Y = y\}, \qquad \text{pour tous } x \in E,\ y \in F.$$

  L'énoncé similaire pour n variables aléatoires discrètes est vrai.

Dans la suite, nous donnons quelques exemples usuels de variables ou plus précisément de lois discrètes. On aura $(\Omega, \mathbb{P})$ un espace probabilisé et $X : \Omega \to E \subset \mathbb{R}$ une variable aléatoire discrète réelle (E fini ou dénombrable).

2.2. Lois unité

Soit $E = \{x_0\}$ un ensemble réduit à un singleton. La loi d'une variable aléatoire X telle que $P_X\{x_0\} = \mathbb{P}\{X = x_0\} = 1$ est dite loi unité ou masse de Dirac en x_0. Pour toute partie B, on a $P_X(B) = 1$ si $x_0 \in B$, 0 sinon. Ces lois sont les lois des variables aléatoires constantes. Si $x_0 \in \mathbb{R}$, on a

$$\mathbb{E}[X] = x_0, \qquad \mathbb{E}\big[X^2\big] = x_0^2, \qquad \mathrm{Var}(X) = \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2 = 0,$$
$$\varphi_X(\theta) = \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta X}\big] = \mathrm{e}^{\mathrm{i}\theta x_0}, \qquad F_X(x) = \begin{cases} 0 & \text{si } x < x_0 \\ 1 & \text{si } x \geqslant x_0. \end{cases}$$

2.3. Lois uniformes sur des ensembles finis

Soit $E = \{x_1, \ldots, x_n\}$. Si la loi de X vérifie

$$P_X\{x_k\} = 1/n, \qquad \text{pour tout } k \in \{1, \ldots, n\},$$

alors la loi de X est la *loi uniforme sur* E, elle est notée $\mathcal{U}(E)$.

Remarque. — Les calculs qui suivent sont peu utilisés. Ils peuvent être vus comme exercice.

Lorsque E est un sous-ensemble de points régulièrement espacés de $\mathbb{R}$, par exemple $E = \{1, \ldots, n\}$, on a

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{n} k\,\mathbb{P}\{X = k\} = \sum_{k=1}^{n} k/n = (n+1)/2 \\
\mathbb{E}\big[X^2\big] &= \sum_{k=1}^{n} k^2\,\mathbb{P}\{X = k\} = \sum_{k=1}^{n} k^2/n = (n+1)(2n+1)/6 \\
\mathrm{Var}(X) &= \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2 = (n+1)(n-1)/12 = (n^2-1)/12 \\
\varphi_X(\theta) &= \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta X}\big] = \sum_{k=1}^{n} \mathrm{e}^{\mathrm{i}\theta k}/n = \frac{1}{n} \times \frac{\mathrm{e}^{\mathrm{i}\theta(n+1)} - \mathrm{e}^{\mathrm{i}\theta}}{\mathrm{e}^{\mathrm{i}\theta} - 1} = \frac{\mathrm{e}^{\mathrm{i}(n+1)\theta/2}}{n} \times \frac{\sin(n\theta/2)}{\sin(\theta/2)} \\
F_X(x) &= \sum_{x' \in \{1,\ldots,n\}, x' \leqslant x} \mathbb{P}\{X = x'\} = \begin{cases} 0 & \text{si } x < 1 \\ k/n & \text{si } k \leqslant x < k+1 \text{ avec } k \in \{0, \ldots, n-1\} \\ 1 & \text{si } x \geqslant n. \end{cases}
\end{aligned}$$

Si l'espacement des points est de pas $h > 0$, par exemple $E = \{h, \ldots, nh\}$, en considérant $X = hX'$ avec X' de loi uniforme sur $\{1, \ldots, n\}$, on obtient

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[hX'] = (n+1)h/2 \\
\mathbb{E}\big[X^2\big] &= \mathbb{E}\big[(hX')^2\big] = (n+1)(2n+1)h^2/6 \\
\mathrm{Var}(X) &= \mathrm{Var}\big(hX'\big) = (n^2-1)h^2/12 \\
\varphi_X(\theta) &= \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta X}\big] = \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta h X'}\big] = \varphi_{X'}(h\theta) = \frac{\mathrm{e}^{\mathrm{i}(n+1)h\theta/2}}{n} \times \frac{\sin(nh\theta/2)}{\sin(h\theta/2)} \\
F_X(x) &= \sum_{x' \in \{h,\ldots,nh\}, x' \leqslant x} \mathbb{P}\{X = x'\} = \begin{cases} 0 & \text{si } x < h \\ k/n & \text{si } kh \leqslant x < (k+1)h \text{ avec } k \in \{1, \ldots, n\} \\ 1 & \text{si } x \geqslant nh. \end{cases}
\end{aligned}$$

Si E est un ensemble de n points uniformément espacés d'extrémités $a < b$, en prenant $X = a - h + hX'$ avec X' uniformément répartie sur $\{1, \ldots, n\}$ et $h = (b-a)/(n-1)$, alors

X est uniformément répartie sur $E = \{a, a+h, \ldots, b-h, b\}$ et on a

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[a-h+hX'] = a-h+h\times\frac{n+1}{2} = a+h\times\frac{n-1}{2} = a+\frac{b-a}{2} = \frac{a+b}{2}\\
\mathbb{E}\big[X^2\big] &= \text{peu intéressant}\ldots\\
\operatorname{Var}(X) &= \operatorname{Var}\big(a-h+hX'\big) = \operatorname{Var}\big(hX'\big) = (n^2-1)h^2/12 = \frac{n+1}{n-1}\times\frac{(b-a)^2}{12}\\
\varphi_X(\theta) &= \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta X}\big] = \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta(a-h)}\times\mathrm{e}^{\mathrm{i}\theta hX'}\big] = \mathrm{e}^{\mathrm{i}(a-h)\theta}\times\varphi_{X'}(h\theta)\\
&= \mathrm{e}^{\mathrm{i}(a-h)\theta}\times\frac{\mathrm{e}^{\mathrm{i}(n+1)h\theta/2}}{n}\times\frac{\sin(nh\theta/2)}{\sin(h\theta/2)} = \mathrm{e}^{\mathrm{i}(a+b)\theta/2}\times\frac{\sin\big(\frac{n}{n-1}(b-a)\theta/2\big)}{n\sin\big(\frac{1}{n-1}(b-a)\theta/2\big)}\\
F_X(x) &= \sum_{x'\in\{a,a+h,\ldots,b\},x'\leqslant x}\mathbb{P}\{X=x'\} = \begin{cases}0 & \text{si } x<a\\ k/n & \text{si } a+(k-1)h\leqslant x<a+kh\\ & \quad\text{avec } k\in\{1,\ldots,n-1\}\\ 1 & \text{si } x\geqslant b.\end{cases}
\end{aligned}$$

Il est intéressant de noter la convergence des expressions précédentes lorsque n tend vers l'infini vers celles qui correspondent à la loi uniforme sur un intervalle $[a, b]$ dans le cadre « continu » (*voir* section suivante).

2.4. Lois de Bernoulli

Soient $E = \{0, 1\}$ et $p = \mathbb{P}\{X = 1\} = 1 - \mathbb{P}\{X = 0\}$. La loi de X est appelée *loi de Bernoulli de paramètre* p, elle notée $\mathcal{B}(1, p)$. On a

$$\begin{aligned}
\mathbb{E}[X] &= 0\times\mathbb{P}\{X=0\}+1\times\mathbb{P}\{X=1\} = 0\times(1-p)+1\times p = p\\
\mathbb{E}\big[X^2\big] &= 0^2\times\mathbb{P}\{X=0\}+1^2\times\mathbb{P}\{X=1\} = 0\times(1-p)+1\times p = p\\
\operatorname{Var}(X) &= \mathbb{E}\big[X^2\big]-\mathbb{E}[X]^2 = p-p^2 = p(1-p)\\
\varphi_X(\theta) &= \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta X}\big] = \mathrm{e}^{\mathrm{i}\theta 0}\,\mathbb{P}\{X=0\}+\mathrm{e}^{\mathrm{i}\theta 1}\,\mathbb{P}\{X=1\} = p\,\mathrm{e}^{\mathrm{i}\theta}+1-p\\
F_X(x) &= \sum_{x'\in\{0,1\},x'\leqslant x}\mathbb{P}\{X=x'\} = \begin{cases}0 & \text{si } x<0\\ 1-p & \text{si } 0\leqslant x<1\\ 1 & \text{si } x\geqslant 1.\end{cases}
\end{aligned}$$

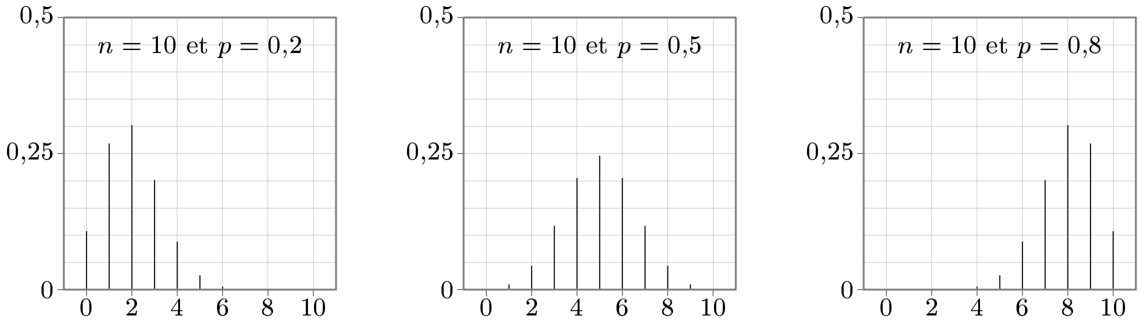
Une variable aléatoire (de loi) de Bernoulli est l'indicatrice de l'événement $\{X = 1\}$ et l'indicatrice d'un événement A est une variable aléatoire de loi de Bernoulli de paramètre $\mathbb{P}(A)$.

2.5. Lois binomiales

Soient $E = \{0, 1, \ldots, n\}$, $n \geqslant 1$, $p \in [0, 1]$. Si la loi de X vérifie

$$P_X\{k\} = \mathbb{P}\{X=k\} = \begin{cases}\dbinom{n}{k}p^k(1-p)^{n-k} & \text{si } k\in\{0,1,\ldots,n\}\\ 0 & \text{sinon,}\end{cases}$$

alors la loi de X est la *loi binomiale de paramètres n et p*, elle est notée $\mathcal{B}(n,p)$.



FIGURES 4–6. — *Distributions de lois binomiales pour $n = 10$ et différentes valeurs de p.*

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{n} k\,\mathbb{P}\{X=k\} = \sum_{k=0}^{n} k \times \binom{n}{k} p^k(1-p)^{n-k} = \cdots = np \\
\mathbb{E}\big[X^2\big] &= \sum_{k=0}^{n} k^2\,\mathbb{P}\{X=k\} = \sum_{k=0}^{n} k^2 \times \binom{n}{k} p^k(1-p)^{n-k} = \cdots = np - np^2 + n^2p^2 \\
\mathrm{Var}(X) &= \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2 = np - np^2 = np(1-p) \\
\varphi_X(\theta) &= \sum_{k=0}^{n} \mathrm{e}^{\mathrm{i}\theta k}\,\mathbb{P}\{X=k\} = \sum_{k=0}^{n} \mathrm{e}^{\mathrm{i}\theta k}\binom{n}{k} p^k(1-p)^{n-k} = \cdots = \big(p\,\mathrm{e}^{\mathrm{i}\theta} + 1 - p\big)^n \\
F_X(x) &= \sum_{x'\in\{0,\dots,n\},x'\leqslant x} \mathbb{P}\{X=x'\} = \text{pas de formule sommatoire particulière.}
\end{aligned}$$

Si X est de loi $\mathcal{B}(m,p)$, Y de loi $\mathcal{B}(n,p)$, et si de plus X et Y sont indépendantes, alors $X+Y$ est de loi $\mathcal{B}(m+n,p)$. En particulier, si $X_1,\dots,X_n : \Omega \to \{0,1\}$ sont des variables aléatoires indépendantes de même loi la loi de Bernoulli de paramètre p, alors $X = X_1 + \cdots + X_n : \Omega \to \{0,1,\dots,n\}$ est de loi $\mathcal{B}(n,p)$.

Remarque. — Les calculs de l'espérance et de la variance d'une variable aléatoire de loi binomiale est immédiat en utilisant la propriété précédente : $\mathbb{E}[X] = \mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = p + \cdots + p = np$, et $\mathrm{Var}(X) = \mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = p + \cdots + p = np$. Cette propriété se prouve facilement en comparant les fonctions caractéristiques, une démonstration « à la main » est en exercice.

2.6. Lois géométriques

Soient $E = \mathbb{N}^* = \{1,2,\dots\}$, $p \in \,]0,1]$. Si la loi de X vérifie

$$P_X\{n\} = \mathbb{P}\{X=n\} = \begin{cases} p(1-p)^{n-1} & \text{si } n \in \mathbb{N}^* \\ 0 & \text{sinon,} \end{cases}$$

alors la loi de X est la loi géométrique de paramètre p, elle est notée $\mathcal{G}(p)$.

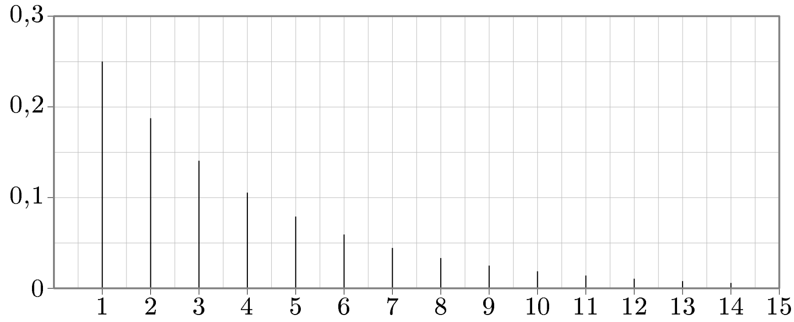


FIGURE 7. — *Distribution de la loi de géométrique de paramètre $p = 0{,}25$.*

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{n\geqslant 1} n\,\mathbb{P}\{X=n\} = \sum_{n\geqslant 1} np(1-p)^{n-1} = \cdots = 1/p\\
\mathbb{E}\big[X^2\big] &= \sum_{n\geqslant 1} n^2\,\mathbb{P}\{X=n\} = \sum_{n\geqslant 1} n^2p(1-p)^{n-1} = \cdots = (2-p)/p^2\\
\mathrm{Var}(X) &= \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2 = (1-p)/p^2\\
\varphi_X(\theta) &= \sum_{n\geqslant 1} \mathrm{e}^{\mathrm{i}\theta n}\,\mathbb{P}\{X=n\} = \sum_{n\geqslant 1} \mathrm{e}^{\mathrm{i}\theta n} p(1-p)^{n-1} = \cdots = \frac{p\,\mathrm{e}^{\mathrm{i}\theta}}{1-(1-p)\,\mathrm{e}^{\mathrm{i}\theta}}\\
F_X(x) &= \sum_{x'\in\mathbb{N}^*, x'\leqslant x} \mathbb{P}\{X=x'\} = \begin{cases} 0 & \text{si } x<1\\ 1-(1-p)^{\lfloor n\rfloor} & \text{si } x\geqslant 1\\ & \text{avec } \lfloor x\rfloor \text{ la partie entière de } x.\end{cases}
\end{aligned}$$

Remarques. — *a*) Les lois géométriques apparaissent souvent comme loi du rang de premier succès dans une suite de « pile ou face » infini (avec p la probabilité d'obtenir « pile » ou « succès »), ce qu'on appelle aussi *schéma de Bernoulli infini*.

b) Pour $r \in \mathbb{N}^*$, la loi du rang de r-ième succès X_r vérifie

$$P_{X_r}\{n\} = \mathbb{P}\{X_r = n\} = \begin{cases} \dbinom{n-1}{r-1} p^r(1-p)^{n-r} & \text{si } n \in \{r, r+1, \ldots\}\\ 0 & \text{sinon,}\end{cases}$$

loi qui est appelée loi binomiale négative de paramètres r et p, et parfois notée $\mathcal{B}inneg(r,p)$. Une somme de r variables aléatoires indépendantes toutes de loi $\mathcal{G}(p)$ — ce qui est aussi le cas de X_r — est de loi binomiale négative de paramètres r et p.

c) Les lois géométriques possèdent une propriété caractéristique d'absence de mémoire : avoir perdu jusqu'au rang n n'informe de rien sur le rang ultérieur de succès.

d) Le cas $p = 0$ est dégénéré : quoiqu'on fasse, on perd. Ainsi, une variable aléatoire de loi $\mathcal{G}(0)$ vérifie $\mathbb{P}\{X = \infty\} = 1$.

2.7. Lois de Poisson

Soient $E = \mathbb{N} = \{0, 1, 2, \ldots\}$, $\lambda \geqslant 0$. Si la loi de X vérifie

$$P_X\{n\} = \mathbb{P}\{X = n\} = \begin{cases} \mathrm{e}^{-\lambda}\dfrac{\lambda^n}{n!} & \text{si } n \in \mathbb{N}\\ 0 & \text{sinon,}\end{cases}$$

alors la loi de X est la loi de Poisson de paramètre λ, elle est notée $\mathcal{P}(\lambda)$.

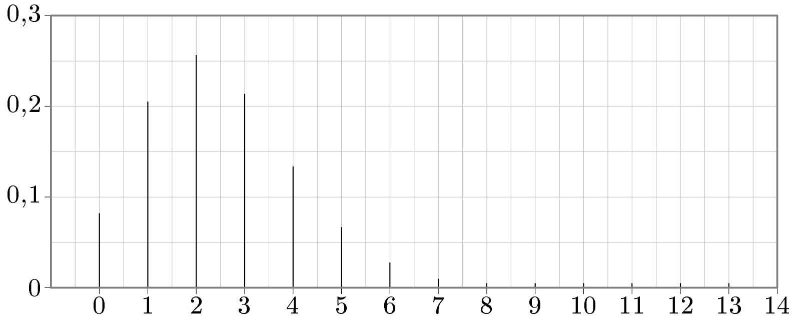


FIGURE 8. — *Distribution de la loi de Poisson de paramètre $\lambda = 2{,}5$.*

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{n\geqslant 0} n\,\mathbb{P}\{X=n\} = \sum_{n\geqslant 0} n\,\mathrm{e}^{-\lambda}\frac{\lambda^n}{n\,!} = \cdots = \lambda \\
\mathbb{E}\big[X^2\big] &= \sum_{n\geqslant 0} n^2\,\mathbb{P}\{X=n\} = \sum_{n\geqslant 0} n^2\,\mathrm{e}^{-\lambda}\frac{\lambda^n}{n\,!} = \cdots = \lambda+\lambda^2 \\
\mathrm{Var}(X) &= \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2 = \lambda \\
\varphi_X(\theta) &= \sum_{n\geqslant 0} \mathrm{e}^{\mathrm{i}\theta n}\,\mathbb{P}\{X=n\} = \sum_{n\geqslant 0} \mathrm{e}^{\mathrm{i}\theta n}\,\mathrm{e}^{-\lambda}\frac{\lambda^n}{n\,!} = \cdots = \exp\big(\lambda\,\mathrm{e}^{\mathrm{i}\theta} - \lambda\big) \\
F_X(x) &= \sum_{x'\in\mathbb{N},\,x'\leqslant x} \mathbb{P}\{X=x'\} = \text{pas de formule sommatoire particulière.}
\end{aligned}$$

Si X est de loi $\mathcal{P}(\lambda)$, Y de loi $\mathcal{P}(\mu)$, et si de plus X et Y sont indépendantes, alors $X+Y$ est de loi $\mathcal{P}(\lambda+\mu)$. Ceci peut se montrer en utilisant les fonctions caractéristiques ou manuellement. Notons que le résultat est cohérent avec les calculs d'espérance et de variance.

Remarque. — Les lois de Poisson apparaissent dans des modèles où on compte durant une unité de temps le nombre d'occurences d'un certain phénomène… Elles apparaissent aussi dans des calculs approchés portant sur des lois binomiales $\mathcal{B}(n,p)$ avec n grand et p proche de 0 ou, par symétrie, de 1.

Le cas $\lambda = 0$ est particulier : une variable aléatoire de loi $\mathcal{P}(0)$ vérifie $\mathbb{P}\{X=0\} = 1$, c'est-à-dire que rien ne se passe.

3. Cadre « continu »

Ici $E = \mathbb{R}$ ou $\mathbb{R}^n$. Le trait particulier de ses espaces est qu'ils disposent de mesures naturelles qui sont celles de longueur, d'aire ou de volume suivant la dimension. Ces mesures, considérées dans le cadre de l'intégration de Lebesgue, sont justement les mesures de Lebesgue. Elles donnent lieu à une notion d'intégrale qui généralise les calculs usuels d'intégrales sur $\mathbb{R}^n$. Cette généralisation est rarement apparente dans les applications simples qu'on rencontrera ici. Les intégrales

$$\int_{\mathbb{R}} f(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x \quad \text{et} \quad \int_{\mathbb{R}^n} f(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(x_1,\ldots,x_n)\,\mathrm{d}x_1\ldots\mathrm{d}x_n$$

auront donc le sens le plus ordinaire qu'il soit, de même que la notion d'intégrabilité.

DÉFINITION 7. — (i) Une mesure de probabilité μ sur E est dite absolument continue (par rapport à la mesure de Lebesgue) s'il existe une application $p : E \to \mathbb{R}_+$ positive, intégrable, telle que

$$\mu(B) = \int_B p(x)\,\mathrm{d}x = \int_E \mathbb{1}_B(x)p(x)\,\mathrm{d}x, \qquad \text{pour tout } B \subset E \text{ (mesurable)}.$$

Une telle application est alors une fonction de densité (de probabilité) de la mesure de probabilité μ et elle vérifie $\int_E p(x)\,\mathrm{d}x = 1$.

Réciproquement, si $p : E \to \mathbb{R}_+$ est une application positive, intégrable, vérifiant de plus $\int_E p(x)\,\mathrm{d}x = 1$, alors elle définit une unique mesure de probabilité μ_p sur E par $\mu_p(B) = \int_B p(x)\,\mathrm{d}x$ pour $B \subset E$ (appartenant à une classe raisonnable de parties de E). On note alors $\mu_p(\mathrm{d}x) = p(x)\,\mathrm{d}x$.

(ii) Une variable aléatoire $X : \Omega \to E$ est dite « continue » (ou, plus justement, de loi absolument continue [par rapport à la mesure de Lebesgue]) si P_X est une mesure de probabilité absolument continue sur E. On note p_X une densité de la loi de X, et $P_X(\mathrm{d}x) = p_X(x)\,\mathrm{d}x$.

(iii) Soient $(\Omega, \mathbb{P})$ un espace probabilisé, et $X : \Omega \to E$ une variable aléatoire de loi absolument continue, $P_X(\mathrm{d}x) = p_X(x)\,\mathrm{d}x$. Soit $f : E \to \mathbb{R}$ (ou $\mathbb{C}$, $\mathbb{R}^m$) une application (mesurable). Alors $f(X) = f \circ X : \Omega \to \mathbb{R}$ une variable aléatoire réelle dont l'espérance s'écrit

$$\mathbb{E}[f(X)] = \int_E f(x) \times p_X(x)\,\mathrm{d}x$$

lorsque cette intégrale est définie.

Remarque. — Une partie des définitions devrait être justifiée, nous admettrons l'ensemble. Par ailleurs, à une mesure μ est absolument continue correspond une infinité de fonctions de densité : si on modifie une fonction en un nombre fini ou dénombrable de points, on ne modifie pas sont intégrale (par rapport à la mesure de Lebesgue), ainsi deux fonctions de densité ne différant que sur un petit ensemble de points (fini ou dénombrable, par exemple) définissent la même mesure de probabilité.

Pratique. — Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle de loi absolument continue, $P_X(\mathrm{d}x) = p_X(x)\,\mathrm{d}x$.

- Pour tout B intervalle de $\mathbb{R}$,

$$P_X(B) = \mathbb{P}\{X \in B\} = \int_B p_X(x)\,\mathrm{d}x.$$

  Si $B \subset \mathbb{R}$ est une réunion finie ou dénombrable d'intervalles B_n *disjoints*, $B = \bigcup_n B_n$,

$$P_X(B) = \mathbb{P}\{X \in B\} = \sum_n \mathbb{P}\{X \in B_n\} = \sum_n \int_{B_n} p_X(x)\,\mathrm{d}x.$$

- Si l'intégrale $\int_{\mathbb{R}} x\, p_X(x)\,\mathrm{d}x$ est définie, en particulier si elle est absolument convergente, alors l'espérance de X est définie, et égale à cette intégrale ; si elle est finie, la variable aléatoire X est dite intégrable.
- Si la variable aléatoire X est intégrable d'espérance $\mathbb{E}[X]$, la variance de X est définie et égale à

$$\int_{\mathbb{R}} \big(x - \mathbb{E}[X]\big)^2 p_X(x)\,\mathrm{d}x = \int_{\mathbb{R}} x^2 p_X(x)\,\mathrm{d}x - \mathbb{E}[X]^2.$$

  La variance est finie si et seulement si X est de carré intégrable.

- La fonction caractéristique $\varphi_X : \mathbb{R} \to \mathbb{C}$ de la loi de X est définie et

$$\varphi_X(x) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\theta x} p_X(x)\,\mathrm{d}x, \qquad \text{pour tout } \theta \in \mathbb{R}.$$

- La fonction de répartition de la loi de X est une fonction continue, dérivable de dérivée p_X presque partout (sauf en un nombre fini ou dénombrable de points, par exemple).
- Soient deux variables aléatoires réelles de lois absolument continues $X : \Omega \to \mathbb{R}$ de loi $P_X(\mathrm{d}x) = p(x)\,\mathrm{d}x$, $Y : \Omega \to \mathbb{R}$ de loi $P_Y(\mathrm{d}x) = q(x)\,\mathrm{d}x$. Les variables X et Y sont indépendantes si et seulement si

$$P_{(X,Y)}(\mathrm{d}x\,\mathrm{d}y) = \big(p(x)\,\mathrm{d}x\big) \times \big(q(y)\,\mathrm{d}y\big),$$

c'est-à-dire que la loi du couple (X, Y) est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R} \times \mathbb{R}$ et admet pour densité $(x, y) \in \mathbb{R} \times \mathbb{R} \mapsto p(x)q(y)$. L'énoncé similaire pour n variables aléatoires réelles ou vectorielles de loi absolument continues est vrai.

Dans la suite, nous donnons quelques exemples usuels de lois de variables aléatoires réelles, absolument continues. On aura $(\Omega, \mathbb{P})$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle de loi absolument continue, $P_X(\mathrm{d}x) = p_X(x)\,\mathrm{d}x$.

3.1. Lois uniformes sur des intervalles bornés

Soit E un intervalle borné non vide de $\mathbb{R}$ d'extrémités $a < b$ ($E = [a, b]$, ou $[a, b[$, $]a, b]$, $]a, b[$). Si la loi de X vérifie

$$P_X(\mathrm{d}x) = p_X(x)\,\mathrm{d}x = \mathbb{1}_E(x)\,\frac{\mathrm{d}x}{b-a}\,,$$

alors la loi de X est la loi uniforme sur E, elle est notée $\mathcal{U}(E)$. On remarquera que l'ouverture ou la fermeture des bornes ne change rien à la loi considérée.

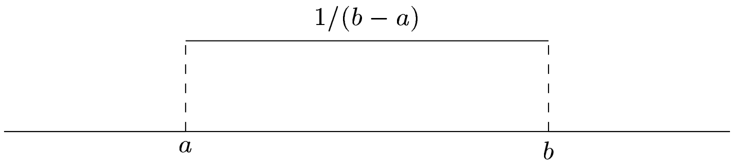


Figure 9. — *Densité de loi uniforme.*

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{+\infty} x\,p_X(x)\,\mathrm{d}x = \int_a^b x\,\frac{\mathrm{d}x}{b-a} = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}\\
\mathbb{E}\big[X^2\big] &= \int_{-\infty}^{+\infty} x^2 p_X(x)\,\mathrm{d}x = \int_a^b x^2\,\frac{\mathrm{d}x}{b-a} = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}\\
\mathrm{Var}(X) &= \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2 = \frac{4a^2+4ab+4b^2-3a^2-6ab-3b^2}{12} = \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}\\
\varphi_X(\theta) &= \int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}\theta x} p_X(x)\,\mathrm{d}x = \int_a^b \mathrm{e}^{\mathrm{i}\theta x}\,\frac{\mathrm{d}x}{b-a} = \frac{\mathrm{e}^{\mathrm{i}\theta b}-\mathrm{e}^{\mathrm{i}\theta a}}{i\theta(b-a)} = \mathrm{e}^{\mathrm{i}\theta(a+b)\theta/2} \times \frac{\sin((b-a)\theta/2)}{(b-a)\theta/2}\\
F_X(x) &= \int_{-\infty}^{x} p_X(y)\,\mathrm{d}y = \begin{cases} 0 & \text{si } x \leqslant a\\ (x-a)/(b-a) & \text{si } a \leqslant x \leqslant b\\ 1 & \text{si } x \geqslant b.\end{cases}
\end{aligned}$$

Si $X : \Omega \to [0, 1]$ est une variable aléatoire de loi uniforme sur l'intervalle $[0, 1]$, alors pour tous $-\infty < a < b < +\infty$, $Y = (b-a)X + a$ est une variable aléatoire de loi uniforme sur l'intervalle $[a, b]$. Ceci est immédiat. On peut le voir facilement en considérant les fonctions de répartition ou en calculant les fonctions caractéristiques. De même, si $Y : \Omega \to [a, b]$ est de loi uniforme sur $[a, b]$, alors $X = (Y - a)/(b - a)$ est de loi uniforme sur $[0, 1]$.

Remarque. — La plupart des générateurs aléatoires (calculatrices, logiciels de calcul scientifique) simulent une variable aléatoire de loi uniforme sur $[0, 1]$ ou permettent aisément de s'y ramener.

3.2. Lois exponentielles

Soient $E = \mathbb{R}_+$ et $\lambda > 0$. Si la loi de X vérifie

$$P_X(\mathrm{d}x) = p_X(x)\,\mathrm{d}x = \lambda \mathbb{1}_{\mathbb{R}_+}(x)\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x,$$

alors la loi de X est la loi exponentielle de paramètre λ, elle est notée $\mathcal{E}(\lambda)$. On remarquera que l'ouverture ou la fermeture en 0 ne change rien à la loi considérée.

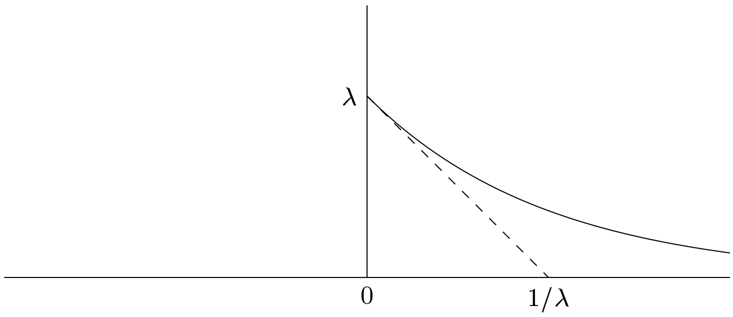


Figure 10. — *Densité de loi exponentielle.*

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{+\infty} x\,p_X(x)\,\mathrm{d}x = \int_0^{+\infty} x\lambda\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \big[x \times (-\,\mathrm{e}^{-\lambda x})\big]_0^{+\infty} + \int_0^{+\infty} \mathrm{e}^{-\lambda x}\,\mathrm{d}x = 1/\lambda \\
\mathbb{E}\big[X^2\big] &= \int_{-\infty}^{+\infty} x^2 p_X(x)\,\mathrm{d}x = \int_0^{+\infty} x^2\lambda\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \big[x^2 \times (-\,\mathrm{e}^{-\lambda x})\big]_0^{+\infty} + \int_0^{+\infty} 2x\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x \\
&= 2/\lambda \int_0^{+\infty} x\lambda\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x = 2/\lambda^2 \\
\mathrm{Var}(X) &= \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2 = 1/\lambda^2 \\
\varphi_X(\theta) &= \int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}\theta x} p_X(x)\,\mathrm{d}x = \int_0^{+\infty} \mathrm{e}^{(\mathrm{i}\theta - \lambda)x}\,\mathrm{d}x = \frac{1}{\lambda - \mathrm{i}\theta} \\
F_X(x) &= \int_{-\infty}^{x} p_X(y)\,\mathrm{d}y = \begin{cases} 0 & \text{si } x \leqslant a \\ 1 - \mathrm{e}^{-\lambda x} & \text{si } x \geqslant 0. \end{cases}
\end{aligned}$$

Remarques. — *a*) Les lois exponentielles apparaissent comme lois de temps aléatoires, par exemple le temps de désintégration d'un atome ou d'une particule instable. On y a fréquemment recours pour modéliser le temps d'attente entre deux clients dans une file d'attente.

b) le temps attendu par le r-ième client est modélisé par la somme de r variables aléatoires indépendantes toutes de loi exponentielle de paramètre λ ; pour $r \geqslant 2$, cette somme n'est pas de loi exponentielle mais de loi Gamma de paramètres r et λ donnée par

$$P_{r,\lambda}(\mathrm{d}x) = \mathbb{1}_{[0,+\infty[}(x)\,\frac{\lambda^r}{(r-1)\,!}\,x^{r-1}\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x,$$

cette mesure de probabilité est généralement notée $\mathcal{G}amma(r, \lambda)$.

c) Les lois exponentielles possèdent une propriété caractéristique d'absence de mémoire semblable à celle des lois géométriques avec ici pour temps $\mathbb{R}_+$ au lieu de $\mathbb{N}^*$.

d) Le cas $\lambda = 0$ est dégénéré : on n'observe jamais la désintégration, et donc une variable aléatoire X de loi $\mathcal{E}(0)$ vérifie $\mathbb{P}\{X = \infty\} = 1$.

Exercice 3. — Montrer que $Y : \Omega \to [0,1]$ est une variable aléatoire de loi uniforme sur l'intervalle $[0,1]$, alors pour tout $\lambda > 0$, $X = -\ln(1-Y)/\lambda$ et $X' = -\ln Y/\lambda$ sont de loi exponentielle de paramètre λ.

3.3. Lois normales

Soient $E = \mathbb{R}$, $m \in \mathbb{R}$ et $\sigma > 0$. Si la loi de $Z : \Omega \to \mathbb{R}$ vérifie

$$P_Z(\mathrm{d}x) = \mathrm{e}^{-x^2/2} \frac{\mathrm{d}x}{\sqrt{2\pi}},$$

alors la loi de Z est la loi Normale, ou loi normale standard, elle est notée $\mathcal{N}(0,1)$.

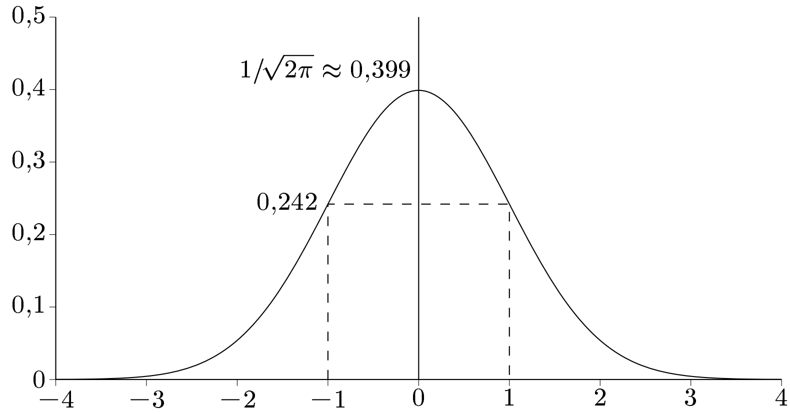


FIGURE 11. — *Densité de la loi normale* $\mathcal{N}(0,1)$.

$$\begin{aligned}
\mathbb{E}[Z] &= \int_{-\infty}^{+\infty} x\, p_Z(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} x\,\mathrm{e}^{-x^2/2} \frac{\mathrm{d}x}{\sqrt{2\pi}} = 0 && \text{(par symétrie)}\\
\mathbb{E}[Z^2] &= \int_{-\infty}^{+\infty} x^2\, p_Z(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} x^2\,\mathrm{e}^{-x^2/2} \frac{\mathrm{d}x}{\sqrt{2\pi}} = \cdots = 1 && \text{(intégration par parties)}\\
\mathrm{Var}(Z) &= \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = 1\\
\varphi_Z(\theta) &= \int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}\theta x} p_Z(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}\theta x - x^2/2} \frac{\mathrm{d}x}{\sqrt{2\pi}} = \cdots = \mathrm{e}^{-\theta^2/2} && \text{(résoudre une équa. diff.)}\\
F_Z(x) &= \int_{-\infty}^{x} p_Z(y)\,\mathrm{d}y = \int_{-\infty}^{x} \mathrm{e}^{-y^2/2} \frac{\mathrm{d}y}{\sqrt{2\pi}} = \Phi(x) && \text{(notation consacrée).}
\end{aligned}$$

Soient $m \in \mathbb{R}$, $\sigma > 0$ et $X = \sigma Z + m$. Par changement de variable linéaire $x = \sigma z + m$, la loi de X vérifie

$$P_X(\mathrm{d}x) = p_X(x)\,\mathrm{d}x = \mathrm{e}^{-\frac{1}{2}(\frac{x-m}{\sigma})^2 x} \frac{\mathrm{d}x}{\sqrt{2\pi}\,\sigma}.$$

La loi de X est la loi normale de paramètres m et σ^2, elle est notée $\mathcal{N}(m,\sigma^2)$. Réciproquement, si $X : \Omega \to \mathbb{R}$ est de loi $\mathcal{N}(m,\sigma^2)$, alors $Z = (X-m)/\sigma$ est de loi $\mathcal{N}(0,1)$. Plus généralement, on a

THÉORÈME 3. — *Si $X : \Omega \to \mathbb{R}$ est une variable aléatoire réelle de loi $\mathcal{N}(m,\sigma^2)$, alors pour tout $\alpha \geqslant 0$ et $\beta \in \mathbb{R}$, $X' = \alpha X + \beta$ est une variable aléatoire réelle de loi $\mathcal{N}(m+\beta, \alpha^2\sigma^2)$.*

On aura toléré $\alpha = 0$ dans cet énoncé en convenant que la loi $\mathcal{N}(m, 0)$ est la loi unité au point m, ce qui est cohérent avec ce que suggère la figure qui suit.

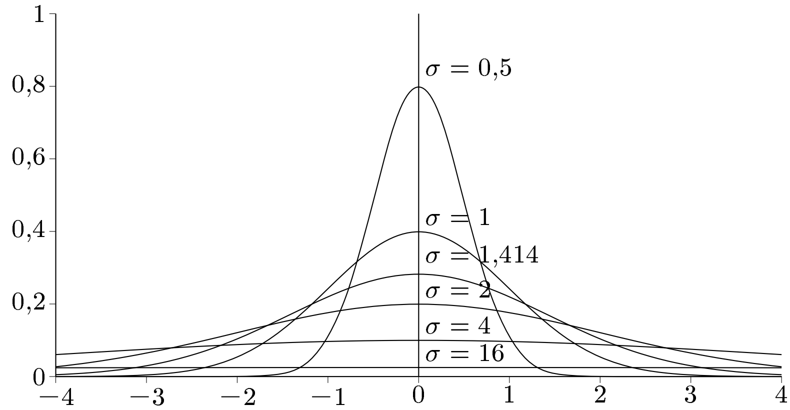


FIGURE 12. — *Densités de lois normales pour $\mu = 0$ et différentes valeurs de σ^2.*

$$\begin{aligned}
\mathbb{E}[X] \quad &= \mathbb{E}[\sigma Z + m] = \sigma\,\mathbb{E}[Z] + m = m \\
\mathbb{E}\big[X^2\big] &= \mathbb{E}\big[(\sigma Z + m)^2\big] = \sigma^2\,\mathbb{E}[Z] + 2\sigma m\,\mathbb{E}\big[Z^2\big] + m^2 = \sigma^2 + m^2 \\
\mathrm{Var}(X) &= \mathrm{Var}(\sigma Z + m) = \mathrm{Var}(\sigma Z) = \sigma^2\,\mathrm{Var}(Z) = \sigma^2 \\
\varphi_X(\theta) &= \mathbb{E}\big[\mathrm{e}^{\mathrm{i}\theta(\sigma Z + m)}\big] = \mathbb{E}\big[\mathrm{e}^{\mathrm{i}(\sigma\theta)Z}\big] \times \mathrm{e}^{\mathrm{i}m\theta} = \varphi_Z(\sigma\theta) \times \mathrm{e}^{\mathrm{i}m\theta} = \exp\big(-\sigma^2\theta^2/2 + \mathrm{i}m\theta\big) \\
F_X(x) &= \mathbb{P}\{X \leqslant x\} = \mathbb{P}\{Z \leqslant (x-m)/\sigma\} = \Phi\big((x-m)/\sigma\big).
\end{aligned}$$

La loi $\mathcal{N}(m, \sigma^2)$ est aussi appelée loi normale, ou gaussienne, de moyenne m et de variance σ^2, ou d'écart-type σ. Compte tenu de la simplicité de la relation entre une loi $\mathcal{N}(m, \sigma^2)$ et la loi $\mathcal{N}(0, 1)$, les tables ou les fonctions de calcul, ne concernent souvent que cette dernière.

THÉORÈME 4. — *Si $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ sont des variables aléatoires réelles indépendantes de lois respectives $\mathcal{N}(m_i, \sigma_i^2)$, alors $X = X_1 + \cdots + X_n$ est une variable aléatoire réelle de loi $\mathcal{N}(m, \sigma^2)$ où $m = m_1 + \cdots + m_n$ et de variance $\sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2$.*

Remarque. — Par linéarité de l'espérance, la moyenne de X est la somme des moyennes des X_i, $i = 1, \ldots, n$, et, par non corrélation des X_i, la variance de X est la somme de leurs variances : le résultat est cohérent. Pour prouver que la loi de X est normale, et en plus avoir la confirmation des valeurs de ses paramètres, il suffit de calculer sa fonction caractéristique : puisque X est une somme de variables aléatoires indépendantes, sa fonction caractéristique est égale au produit de leurs fonctions caractéristiques ; l'expression obtenue s'écrit comme la fonction caractéristique de la loi $\mathcal{N}(m, \sigma^2)$, cette dernière est donc bien la loi de X.

On retiendra que toute combinaison linéaire (ou affine) de variables aléatoires réelles, *indépendantes*, de lois normales, est de loi normale dont l'espérance et la variance s'obtiennent immédiatement.

Si l'hypothèse d'indépendance est abandonnée, ce résultat cesse d'être vrai en général. Un affaiblissement apparent de cette hypothèse consiste à considérer ce qu'on appelle des vecteurs gaussiens : un n-uplet $(X_1, \ldots, X_n)$ de variables variables aléatoires réelles est un vecteur gaussien si toute combinaison linéaire de ses composantes est une variable aléatoire réelle gaussienne, c'est-à-dire de loi normale.

3.4. AUTRES LOIS ABSOLUMENT CONTINUES

La liste des lois absolument continues fréquemment utilisées en Statistique est assez longue, elle contient en particulier les lois de Pearson, ou du χ^2, les loi de Student, de Fisher–Snedecor, etc. Plutôt que d'en donner des descriptions semblables à ce qui précède, nous renvoyons au chapitre suivant, aux tables et à la littérature.

4. Quelques propriétés asymptotiques

4.1. Convergence presque sûre, loi forte des grands nombres

Les variables aléatoires sont des applications définies sur un espace Ω non nécessairement pourvu d'une topologie. Dans ce cadre, la notion de convergence qu'on peut envisager est la convergence simple : $(X_n)_{n\geqslant 1}$ converge simplement vers X si et seulement si, pour tout $\omega \in \Omega$, $(X_n(\omega))_{n\geqslant 1}$ converge vers $X(\omega)$. En termes probabilistes, on parle de convergence sûre. Cette notion de convergence est un peu trop forte et ne prend pas en compte la mesure de probabilité $\mathbb{P}$.

Définition 8. — Soit $(\Omega, \mathbb{P})$ un espace probabilisé et E un ensemble. Une suite de variables aléatoires $(X_n)_{n\geqslant 1}$ à valeurs dans E converge presque sûrement vers $X : \Omega \to E$ si et seulement si il existe $A \subset \Omega$, $\mathbb{P}(A) = 1$, tel que, pour tout $\omega \in A$, $(X_n(\omega))_{n\geqslant 1}$ converge vers $X(\omega)$.

Dans la pratique, l'ensemble presque sûr A n'intervient pas. Puisqu'une intersection finie ou dénombrable d'ensembles presque sûrs est presque sûre, toute suite finie ou dénombrable de convergence presque sûre a lieu simultanément sur un ensemble presque sûr.

Définition 9. — Une suite de variables aléatoires réelles $(X_n)_{n\geqslant 1}$ vérifie la loi forte des grands nombre si il existe une constante $m \in \mathbb{R}$ telle que la suite

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}, \qquad n \geqslant 1,$$

converge presque sûrement vers m.

De telles suites existent, il suffit par exemple de prendre pour un $m \in \mathbb{R}$ donné, $X_n = m$ presque sûrement pour tout $n \geqslant 1$, ou même $X_n = m_n$ presque sûrement pour tout $n \geqslant 1$, avec $(m_n)_{n\geqslant 1}$ une suite de réels de moyenne de Cesàro convergente. Un résultat fréquemment utilisé est la loi forte des grands nombres de Kolmogorov :

Théorème 5 (Loi des forte des grands nombres). — *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires de Ω dans $\mathbb{R}$, indépendantes, intégrables, de même loi. Alors, la suite $(X_n)_{n\geqslant 1}$ vérifie la loi forte des grands nombres avec $m = \mathbb{E}[X_1]$.*

La démonstration de ce résultat est loin d'être immédiate. Quelques énoncés un peu plus généraux peuvent s'en déduire en utilisant la linéarité de la limite presque sûre.

Exemple. — Considérons un dé à six faces. On veut connaître la probabilité d'obtenir 6. Pour cela, on procède à une infinité de jets indépendants du dé et on note 1 si le 6 apparaît, 0 sinon. On obtient ainsi une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires indépendantes de loi de Bernoulli de paramètre p la probabilité d'obtenir 6. Les fréquences d'apparition du 6 sont $(X_1 + \cdots + X_n)/n$, $n \geqslant 1$, et elles convergent presque sûrement vers $p = \mathbb{E}[X_1]$.

4.2. Convergence en probabilité, loi faible des grands nombres

Dans certains cas, en particulier en Statistique, il est nécessaire de considérer un mode de convergence plus faible que la convergence presque sûre. Cette sous-section ne fait qu'introduire un peu de vocabulaire et peut être omise.

DÉFINITION 10. — Soit $(\Omega, \mathbb{P})$ un espace probabilisé et (E, d) un espace métrique. Une suite de variables aléatoires $(X_n)_{n\geqslant 1}$ à valeurs dans E converge en probabilité vers $X : \Omega \to E$ si et seulement pour tout $\varepsilon > 0$, $\mathbb{P}\{d(X_n, X) \geqslant \varepsilon\} \to 0$ quand n tend vers l'infini.

Remarque. — L'ensemble E est ici un espace métrique, ce qui veut dire que pour affaiblir les conditions de convergence sur Ω, on a dû renforcer la structure de E (lorsque $E = \mathbb{R}$, on prend pour distance $d(x, y) = |x-y|$ la distance usuelle). Montrer que la convergence presque sûre implique la convergence en probabilité est assez technique, nous nous dispenserons ainsi que de donner des exemples de convergence en probabilité sans convergence presque sûre.

DÉFINITION 11. — Une suite de variables aléatoires réelles $(X_n)_{n\geqslant 1}$ vérifie la loi faible des grands nombre si il existe une constante $m \in \mathbb{R}$ telle que la suite

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}, \qquad n \geqslant 1,$$

converge en probabilité vers m.

La convergence presque sûre étant plus forte que la convergence en probabilité, il y a, *a priori*, plus de suites de variables aléatoires vérifiant la loi faible des grands nombres que de suites vérifiant la loi forte. Nous donnons un énoncé assez facile à démontrer en utilisant l'inégalité de Tchebychev.

THÉORÈME 6 (UNE LOI DES FAIBLE DES GRANDS NOMBRES). — *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires de Ω dans $\mathbb{R}$, non corrélées, de carrés intégrables, de même espérance et de variances bornées. Alors, la suite $(X_n)_{n\geqslant 1}$ vérifie la loi faible des grands nombres avec $m = \mathbb{E}[X_1]$.*

4.3. CONVERGENCE EN LOI, THÉORÈME CENTRAL LIMITE

Jusqu'à présent, nous avons considéré des convergences de variables aléatoires vues comme des applications. Si on oublie les applications, il ne reste plus que la trace de la manifestation de ces variables aléatoires, c'est-à-dire des lois de probabilité sur un ensemble E. Pour définir la convergence de lois, il faut que E soit muni d'une topologie. Si E est fini ou dénombrable ($E = \mathbb{N}$, $\mathbb{Z}$, ..., mais $E \neq \mathbb{Q}$ par exemple), on le munit de sa topologie discrète $\mathcal{P}(E)$ pour laquelle toute application de E dans $\mathbb{R}$ (par exemple) est continue. Si $E = \mathbb{R}$ ou $\mathbb{R}^n$, on considère la topologie usuelle.

DÉFINITION 12. — Soit E un espace topologique. Une suite de mesures de probabilité $(\mu_n)_{n\geqslant 1}$ converge vers une mesure de probabilité μ sur E si et seulement si pour toute application $f : E \to \mathbb{R}$, continue et bornée

$$\mu_n(f) = \int_E f(x)\, \mu_n(\mathrm{d}x) \longrightarrow \mu(f) = \int_E f(x)\, \mu(\mathrm{d}x) \qquad \text{quand } n \text{ tend vers l'infini.}$$

Remarque. — Les analystes nomment cette notion de convergence la convergence étroite. En analyse, elle est plus restrictive ou plus forte que la convergence vague pour laquelle on remplace « f continue et bornée » par « f continue à support compact ». Ces convergences sont équivalentes dans les cadres probabilistes usuels.

On dit qu'une suite de variables aléatoires $(X_n)_{n\geqslant 1}$ convergent en loi vers une loi μ ou vers la loi μ d'une variable aléatoire X si la suite de leur lois converge vers μ. Cette notion de convergence est encore plus faible que celles qui précède : si une suite de variables aléatoires convergent, presque sûrement ou en probabilité, vers une variable aléatoire X, alors la suite de leurs lois converge vers la loi de la variable aléatoire limite.

Exemple. — Nous avons vu deux types de lois uniformes : sur des ensembles finis, sur des intervalles bornés. Dans le cas où l'ensemble fini est $\{0, 1/n, 2/n, \dots, (n-1)/n, 1\}$ et l'intervalle est $[0,1]$, on est tenté de dire que la loi uniforme discrète μ_n approche la loi uniforme continue μ.

Or, si on prend $f = \mathbb{1}_{[0,1]\setminus\mathbb{Q}}$, on a $\mu_n(f) = 0$ et $\mu(f) = 1$. Cette fonction f est si discontinue qu'elle ne voit aucun des points chargés par les μ_n alors qu'elle voit presque tous les points de $[0,1]$. On comprend alors l'intérêt de ne tester la convergence que sur des classes de fonctions assez régulières, ici continues.

En revenant sur les résultats de calculs antérieurs, on constate qu'on a convergence des moments, mais aussi convergence des fonctions de répartition. Il se trouve que ces dernières convergences impliquent, voire sont équivalentes à, sous certaines conditions que nous ne préciserons pas, la convergence des lois et peuvent servir alors de critère de convergence de lois de variables aléatoires réelles. La convergence simple des fonctions caractéristiques vers une fonction caractéristique est toujours équivalente à la convergence des lois (théorème de Lévy).

THÉORÈME 7 (THÉORÈME CENTRAL LIMITE). — *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires de Ω dans $\mathbb{R}$, indépendantes, de même loi, d'espérance $m = \mathbb{E}[X_1]$ et de variance $\sigma^2 = \mathrm{Var}(X_1) = \mathbb{E}[X_1^2] - m^2$. Alors, les lois des variables*

$$Z_n = \frac{X_1 + \cdots + X_n - n \times m}{\sqrt{n} \times \sigma}$$

convergent vers la loi $\mathcal{N}(0,1)$.

Remarques. — *a*) La démonstration de ce théorème est généralement faite en estimant les fonctions caractéristiques des variables Z_n et en montrant que ces fonctions convergent simplement vers la fonction caractéristique de la loi $\mathcal{N}(0,1)$. Ainsi, si on admet le théorème de Lévy, le théorème central limite apparaît comme un résultat presque élémentaire. Il ne l'est pas. C'est un résultat profond qui tient un rôle central (d'où son nom) en Statistique.

b) Notons que par construction $\mathbb{E}[Z_n] = 0$ et $\mathrm{Var}(Z_n) = 1$, c'est-à-dire que Z_n a été obtenu en considérant d'abord $S_n = X_1 + \cdots + X_n$, puis en centrant cette somme (se ramener à une variable aléatoire d'espérance nulle en lui retranchant sa moyenne), et enfin en réduisant le tout (se ramener à une variable aléatoire de variance ou d'écart-type 1 en la divisant par son écart-type). Ces considérations sont utiles pour retrouver la normalisation de S_n utilisée dans ce théorème.

c) L'application usuelle du théorème central limite est, évidemment, d'approcher la loi d'une telle variable aléatoire Z_n par la loi normale $\mathcal{N}(0,1)$. Cependant, l'erreur commise n'est pas toujours très bien maîtrisée et les conditions d'approximation suivent certaines recettes. En particulier, si S_n est une variable aléatoire de loi binomiale $\mathcal{B}(n,p)$ (on ne suppose pas disposer nécessairement d'une écriture de la forme $S_n = X_1 + \cdots + X_n$) telle que $n \geqslant 50$, $np \geqslant 10$, $n(1-p) \geqslant 10$, on s'autorise à « approcher la loi de S_n par une loi normale », plus exactement, à approcher la loi de Z_n ainsi obtenu par la loi $\mathcal{N}(0,1)$.

Exercices

PROBABILITÉS DISCRÈTES

Exercice 4. — Soient $n \in \mathbb{N}^*$ et $p \in [0,1]$. Nous considérons la loi binomiale $\mathcal{B}(n,p)$ de paramètres n et p, c'est-à-dire la mesure de probabilité discrète μ sur $\{0, 1, \dots, n\}$ définie par

$$\mu\{k\} = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{si } k \in \{0, 1, \dots, n\} \\ 0 & \text{sinon,} \end{cases} \qquad \text{où} \qquad \binom{n}{k} = C_n^k = \frac{n\,!}{k\,! \times (n-k)\,!}$$

(i) Vérifier que cette définition correspond à une mesure de probabilité discrète.

(ii) Calculer son espérance, ou moyenne, ainsi que sa variance de manière directe.

(iii) Soient $\xi_1, \dots, \xi_n$ des variables aléatoires indépendantes de loi $\mathcal{B}(1, p)$ ($\mathbb{P}\{\xi_i = 1\} = 1 - \mathbb{P}\{\xi_i = 0\} = p$). Montrer que $X = \xi_1 + \cdots + \xi_n$ est une variable de loi $\mathcal{B}(n, p)$.

(iv) En déduire, par linéarité de l'espérance, celle de X. En développant la variance de X, et en éliminant les termes nuls, déterminer la variance de X.

Correction. — Non disponible.

Exercice 5. — Soient X et Y deux variables aléatoires indépendantes de lois $\mathcal{B}(m, p)$ et $\mathcal{B}(n, p)$ respectivement, où $m, n \in \mathbb{N}^*$ et $p \in [0, 1]$.

(i) Montrer par un calcul direct que $Z = X + Y$ qui est une variable aléatoire à valeurs dans $\{0, 1, \dots, m + n\}$ est de loi $\mathcal{B}(m + n, p)$.

(ii) En déduire que si $\xi_1, \dots, \xi_n$ est une suite de variables aléatoires indépendantes toutes de loi $\mathcal{B}(1, p)$, alors $X = \xi_1 + \cdots + \xi_n$ est de loi $\mathcal{B}(n, p)$.

Remarque. — Au lieu d'un calcul direct, on aurait pu utiliser les fonctions caractéristiques ou génératrices. Ce développement, bien qu'intéressant, sort du cadre restreint dans lequel nous sommes confinés.

Correction. — Non disponible.

Exercice 6. — (i) Soit $p \in]0, 1]$. Considérons la suite de nombres suivante :

$$\mu\{n\} = \begin{cases} p(1-p)^{n-1} & \text{si } n \in \{1, 2, \dots\} = \mathbb{N}^* \\ 0 & \text{sinon.} \end{cases}$$

Montrer qu'elle définit une mesure de probabilité et calculer sa moyenne. Éventuellement, on pourra représenter graphiquement cette mesure (diagramme en bâtons) et calculer sa variance. Qu'en est-il si $p = 0$? Déterminer la fonction de répartition pour $p \in [0, 1]$. Faire le lien avec une expérience de Bernoulli infinie.

(ii) Soit $\lambda > 0$. Considérons la suite de nombres suivante :

$$\mu\{n\} = \begin{cases} \mathrm{e}^{-\lambda} \times \dfrac{\lambda^n}{n\,!} & \text{si } n \in \{0, 1, 2, \dots\} = \mathbb{N} \\ 0 & \text{sinon.} \end{cases}$$

Montrer qu'elle définit une mesure de probabilité et calculer sa moyenne. Éventuellement, on pourra représenter graphiquement cette mesure (diagramme en bâtons) et calculer sa variance.

Correction. — Non disponible.

Probabilités (absolument) continues

Exercice 7. — Nous considérons la fonction positive p définie par

$$\begin{aligned} p : \mathbb{R} &\longrightarrow \mathbb{R}_+ \\ x &\longmapsto C \times \mathbb{1}_{\mathbb{R}_+}(x)\,\mathrm{e}^{-\lambda x} \end{aligned}$$

où C est une constante strictement positive.

(i) Déterminer C pour que la fonction p soit une densité de mesure de probabilité.

(ii) Calculer la moyenne associée à cette distribution de probabilité. Représenter cette fonction de densité ainsi que sa moyenne.

(iii) Après calcul, la variance de cette loi de probabilité apporte-t-elle quelque information ?

(iv) Considérons X une variable aléatoire de loi exponentielle de paramètre λ. Soit Y la partie entière supérieure de X, $Y : \Omega \to \mathbb{N}^*$. Déterminer la loi de Y.

Correction. — Non disponible.

Exercice 8. — (i) Calculer l'intégrale de Gauss

$$I = \int_{-\infty}^{+\infty} \mathrm{e}^{-x^2/2}\,\mathrm{d}x.$$

Pour cela, on pourra calculer I^2 et utiliser un changement de variables polaires.

(ii) En déduire que la loi $\mathcal{N}(0,1)$ est une mesure de probabilité sur $\mathbb{R}$.

(iii) Calculer la moyenne (espérance) de cette loi de probabilité ainsi que sa variance.

(iv) Montrer que si $m \in \mathbb{R}$, $\sigma > 0$, $\mathcal{N}(m,\sigma^2)$ a pour espérance m et pour variance σ^2.

Correction. — Non disponible.

Exercice 9. — Soient $m \in \mathbb{R}$, $\sigma > 0$ et $X : \Omega \to \mathbb{R}$ une variable aléatoire de loi $\mathcal{N}(m,\sigma^2)$.

(i) Déterminer $\mathbb{P}\{X \leqslant m\}$, $\mathbb{P}\{X > m\}$, $\mathbb{P}\{X = m\}$ et $\mathbb{P}\{X \geqslant m\}$.

(ii) Déterminer à l'aide des tables $\mathbb{P}\{m - \sigma \leqslant X \leqslant m + \sigma\}$. Comparer $\mathbb{P}\{m - \sigma \leqslant X\}$ et $\mathbb{P}\{X \leqslant m + \sigma\}$, faire le lien avec le calcul précédent.

(iii) Exprimer les quantiles de la la loi de X en fonction de ceux de la loi normale $\mathcal{N}(0,1)$.

Correction. — Non disponible.

Chapitre II

ÉLÉMENTS DE STATISTIQUE

1. Problématique de la Statistique

Nous considérons un ensemble Ω, souvent appelé *population* ; des éléments $\omega \in \Omega$, souvent appelés *individus* ; et une variable $X : \Omega \to E$, souvent appelée *variable statistique*. Lorsque Ω est muni d'une mesure de probabilité — ce qui est implicitement le cas même si cela n'est pas toujours précisé —, la loi de X est notée $\mu = P_X$, elle est souvent appelée *loi statistique*. Ce qui nous intéresse est de déterminer μ.

Exemples. — Revenus des foyers français ; taux de plomb dans l'eau du robinet ; résultats à des élections futures, ...

Définition 13. — Soit μ une mesure de probabilité sur un ensemble E. On appelle *paramètre* de la loi μ toute quantité numérique π, réelle, complexe ou vectorielle calculée à partir de μ.

Lorsque $E = \mathbb{R}$, ou $E = \mathbb{R}^n$, un paramètre $\pi \in E$ de μ est un *paramètre de position* si, en translatant la mesure μ, on translate d'autant le paramètre π ; un paramètre π réel, vectoriel ou matriciel, de μ est un *paramètre de dispersion* si il est inchangé par toute translation de μ et s'il est homogène d'ordre 1 (une homothétie de rapport $\alpha \geqslant 0$ de E multiplie π par α).

S'il existe une unité de mesure sous-jacente, les paramètres de position et de dispersion, ou leurs coefficients, s'expriment dans la même unité.

Exemples. — Supposons que μ est une mesure de probabilité sur $\mathbb{R}$. Alors ses quantiles, en particulier sa médiane, et sa moyenne — si elle est définie — sont des paramètres de position de μ. Les différences entre quantiles, en particulier l'étendue interquartile $q_3 - q_1$, et sa variance — si elle est définie — sont des paramètres de dispersion de μ.

Remarque. — Si μ est une mesure de probabilité sur $E = \mathbb{R}^n$, son espérance est le vecteur m de coordonnées

$$m_i = \int_{\mathbb{R}^n} x_i \, \mu(\mathrm{d}x), \qquad i = 1, \dots, n.$$

Il s'agit ci-dessus bien d'une intégrale multiple puisqu'il n'y a aucune raison que $\mu(\mathrm{d}x) = \mu(\mathrm{d}x_1 \dots \mathrm{d}x_n)$ s'écrive $\mu_i(\mathrm{d}x_i) \otimes \nu_i(\mathrm{d}x_1 \dots \mathrm{d}x_{i-1}\, \mathrm{d}x_{i+1} \dots \mathrm{d}x_n)$ (cas pour lequel on aurait $m_i = \int_{\mathbb{R}} x \, \mu_i(\mathrm{d}x)$). Lorsque l'espérance de μ est finie, la matrice de covariance est la matrice symétrique positive $\mathrm{Cov}(\mu)$ de coefficients

$$\mathrm{Cov}(\mu)_{i,j} = \int_{\mathbb{R}^n} (x_i - m_i)(x_j - m_j)\, \mu(\mathrm{d}x), \qquad i, j = 1, \dots, n.$$

Les mêmes remarques quant aux intégrales multiples s'appliquent aussi.

Si on considère un unique individu $\omega \in \Omega$, une *observation* $x = X(\omega)$ de X ne suffit généralement pas à déterminer μ. On a besoin de plus d'observations.

Définition 14. — Soit $n \in \mathbb{N}$.

(i) Un *échantillon de la population* est une suite d'individus $(\omega_1, \dots, \omega_n)$.

(ii) Un *échantillon de la variable* X est une suite de variables $(X_1, \ldots, X_n)$ de même loi $\mu = P_X$.

(iii) Un *échantillon observé* est une suite d'observations

$$(x_1, \ldots, x_n) = \big(X(\omega_1), \ldots, X(\omega_n)\big).$$

L'entier $n \in \mathbb{N}$ est la *taille* de l'échantillon.

(iv) Une variable Λ de la forme $\Lambda = f_n(X_1, \ldots, X_n)$ est appelée *statistique* de l'échantillon $(X_1, \ldots, X_n)$, lui correspond des valeurs observées $\ell = f_n(x_1, \ldots, x_n)$.

La méthode choisie pour former un échantillon s'appelle *échantillonnage.* Parmi toutes les méthodes envisageables, la plus simple est *l'échantillonnage représentatif* : l'échantillon de la population $(\omega_1, \ldots, \omega_n)$ est obtenu par un tirage avec remise dans la population (boules dans une urne) ; de manière plus précise cela revient au niveau de l'échantillon de la variable X à supposer $X_1, \ldots, X_n$ indépendantes (et de même loi).

D'autres modes d'échantillonnage sont fréquemment pratiqués : échantillonnage exhaustif (tirage sans remise), échantillonnages stratifiés (on découpe la population en sous-populations dont on connaît assez bien les proportions relatives, et, sur chacune, on pratique un échantillonnage représentatif ou exhaustif), ...

2. Estimation

2.1. Estimation ponctuelle

Définition 15. — Soient μ une mesure de probabilité sur un ensemble E, π un paramètre de μ, et $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires de loi μ.

(i) On appelle *estimateur* de π toute suite de variables aléatoires $(Z_n)_{n\geqslant 1}$ de la forme

$$Z_n = f_n(X_1, \ldots, X_n), \qquad \text{avec } f_n \text{ une application sur } E^n,$$

à valeurs dans le même espace ($\mathbb{R}$, $\mathbb{C}$ ou $\mathbb{R}^n$) que π.

(ii) Un estimateur $(Z_n)_{n\geqslant 1}$ de π est *consistant* si la suite $(Z_n)_{n\geqslant 1}$ converge en probabilité vers π lorsque n tend vers l'infini.

(iii) Un estimateur $(Z_n)_{n\geqslant 1}$ de π est *sans biais* si, pour tout $n \geqslant 1$, Z_n est intégrable et $\mathbb{E}[Z_n] = \pi$. Si cette dernière condition n'est pas satisfaite, l'estimateur est dit *biaisé.*

(iv) Un estimateur $(Z_n)_{n\geqslant 1}$ de π est *asymptotiquement sans biais* si, pour tout $n \geqslant 1$, Z_n est intégrable et $\mathbb{E}[Z_n]$ tend vers π lorsque n tend vers l'infini.

(v) Une observation $z_n = Z_n(\omega)$ d'un estimateur $(Z_n)_{n\geqslant 1}$ de π est une *estimation* de π.

Il est fréquent et abusif d'appeler estimateur une variable Z_n plutôt que toute la suite $(Z_n)_{n\geqslant 1}$. Ceci est certainement dû au fait que dans la pratique n est la taille de l'échantillon d'individus ou d'observations effectivement collecté, et se retrouve par là même fixé.

Théorème 8, définitions et notations. — *Soient μ une mesure de probabilité sur $E = \mathbb{R}$ d'espérance m et de variance σ^2, et $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes de loi μ.*

(i) *La moyenne échantillon (ou empirique)*

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n},$$

qui est une variable, est un estimateur consistant et sans biais de l'espérance m. La moyenne observée d'un échantillon

$$\bar{x}_n = \frac{x_1 + \cdots + x_n}{n},$$

qui est un réel, est une estimation de m.

(ii) *La variance (empirique) d'un échantillon*

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\big(X_i - \bar{X}_n\big)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \big(\bar{X}_n\big)^2,$$

qui est une variable, est un estimateur consistant, biaisé, mais asymptotiquement sans biais, de la variance σ^2. La variance observée d'un échantillon

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x}_n)^2,$$

qui est un réel, est une estimation de σ^2.

(iii) *La variance (empirique) corrigée ou « échantillon » d'un échantillon*

$$S_{c,n}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\big(X_i - \bar{X}_n\big)^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1} \times \big(\bar{X}_n\big)^2,$$

qui est une variable, est un estimateur consistant et sans biais de la variance σ^2. La variance corrigée d'un échantillon observé

$$s_{c,n}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1} \times (\bar{x}_n)^2,$$

qui est un réel, est une estimation de σ^2.

Ce théorème résulte assez immédiatement de la loi des grands nombres et de calculs d'espérances :

$$\mathbb{E}\big[\bar{X}_n\big] = m, \qquad \mathbb{E}\big[S_n^2\big] = \frac{n-1}{n} \times \sigma^2, \qquad \mathbb{E}\big[S_{c,n}^2\big] = \sigma^2.$$

On doit noter qu'une estimation n'est que ponctuelle et rien ne permet d'affirmer de manière sûre si elle est proche ou non de la valeur à estimer. Lorsque l'estimateur est non biaisé et que sa variance, ou son écart-type, est petit, alors il est plus probable d'observer des valeurs proches du paramètre à estimer que des valeurs éloignées (ce qui reste possible). Par exemple,

$$\begin{aligned}\operatorname{Var}\big(\bar{X}_n\big) = \operatorname{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) &= \frac{1}{n^2}\operatorname{Var}(X_1 + \cdots + X_n) \\ \text{(indépendance)} &= \frac{1}{n^2}\big(\operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n)\big) = \frac{\sigma^2}{n}\end{aligned}$$

permet d'avoir un contrôle sur l'ensemble des valeurs qu'on peut observer de $\bar{X}_n$ mais pas sur une observation, ou estimation, ponctuelle.

Exercice 1. — Calculer la variance de la variance empirique S_n^2 lorsque $X_1, \ldots, X_n$ sont des variables aléatoires réelles indépendantes admettant des moments d'ordre 4 et de même loi μ.

Correction. — Soient $X_1, \dots, X_n$ des variables aléatoires réelles indépendantes admettant des moments d'ordre 4. Notons qu'elles admettent alors des moments d'ordre 1, 2, 3, et 4, qui sont égaux puisque les variables ont même loi. Nous notons $m = \mathbb{E}[X_1]$ et $\sigma^2 = \mathrm{Var}(X_1)$ et omettrons l'indice n pour la moyenne et la variance empirique.

Remarquons tout d'abord que la variance empirique S^2 ne dépend pas de la valeur de m : si on ajoute à chaque variable $X_1, \dots, X_n$ un réel ν, la moyenne empirique est incrémentée d'autant et la variance empirique est inchangée (ce qui est normal puisqu'elle est liée à la dispersion des variables). Nous pouvons donc supposer $m = 0$ dans la suite. Nous notons

$$S^4 = (S^2)^2, \qquad \sigma^4 = (\sigma^2)^2 = \mathrm{Var}(X_1)^2 = \mathbb{E}[X_1^2]^2 \quad \text{et} \quad \tau^4 = \mathbb{E}[X_1^4].$$

On a

$$\begin{aligned}
n^2\,\mathbb{E}\big[S^4\big] &= \mathbb{E}\Big[\sum_{k,\ell=1}^n (X_k-\bar X)^2(X_\ell-\bar X)^2\Big] = \sum_{k,\ell=1}^n \mathbb{E}\big[(X_k-\bar X)^2(X_\ell-\bar X)^2\big] \\
&= \sum_{k,\ell=1}^n \mathbb{E}\big[X_k^2X_\ell^2 + X_k^2\bar X^2 + X_\ell^2\bar X^2 + \bar X^4 \\
&\qquad\qquad - 2X_k^2X_\ell\bar X - 2X_kX_\ell^2\bar X + 2X_kX_\ell\bar X^2 + 2X_kX_\ell\bar X^2 - 2X_k\bar X^3 - 2X_\ell\bar X^3\big] \\
&= \sum_{k,\ell=1}^n \mathbb{E}\big[X_k^2X_\ell^2 + 2X_k^2\bar X^2 + \bar X^4 - 4X_k^2X_\ell\bar X + 4X_kX_\ell\bar X^2 - 4X_k\bar X^3\big]
\end{aligned}$$

la dernière ligne ayant été obtenue en permutant les indices k et ℓ sur certains termes puisqu'ils jouent des rôles symétriques. En utilisant simplement la linéarité de l'espérance,

$$\begin{aligned}
n^2\,\mathbb{E}\big[S^4\big] &= \sum_{k,\ell=1}^n \mathbb{E}\big[X_k^2X_\ell^2\big] + 2\sum_{k,\ell=1}^n \mathbb{E}\big[X_k^2\bar X^2\big] + \sum_{k,\ell=1}^n \mathbb{E}\big[\bar X^4\big] \\
&\quad - 4\sum_{k,\ell=1}^n \mathbb{E}\big[X_k^2X_\ell\bar X\big] + 4\sum_{k,\ell=1}^n \mathbb{E}\big[X_kX_\ell\bar X^2\big] - 4\sum_{k,\ell=1}^n \mathbb{E}\big[X_k\bar X^3\big] \\
&= \sum_{k,\ell=1}^n \mathbb{E}\big[X_k^2X_\ell^2\big] + 2n\sum_{k=1}^n \mathbb{E}\big[X_k^2\bar X^2\big] + n^2\,\mathbb{E}\big[\bar X^4\big] \\
&\quad - 4\sum_{k=1}^n \mathbb{E}\Big[X_k^2\Big(\sum_{\ell=1}^n X_\ell\Big)\bar X\Big] + 4\,\mathbb{E}\Big[\Big(\sum_{k=1}^n X_k\Big)\Big(\sum_{\ell=1}^n X_\ell\Big)\bar X^2\Big] - 4\sum_{\ell=1}^n \mathbb{E}\Big[\Big(\sum_{k=1}^n X_k\Big)\bar X^3\Big] \\
&= \sum_{k,\ell=1}^n \mathbb{E}\big[X_k^2X_\ell^2\big] + 2n\sum_{k=1}^n \mathbb{E}\big[X_k^2\bar X^2\big] + n^2\,\mathbb{E}\big[\bar X^4\big] \\
&\quad - 4\sum_{k=1}^n \mathbb{E}\big[X_k^2\big(n\bar X\big)\bar X\big] + 4\,\mathbb{E}\big[\big(n\bar X\big)\big(n\bar X\big)\bar X^2\big] - 4\sum_{\ell=1}^n \mathbb{E}\big[\big(n\bar X\big)\bar X^3\big] \\
&= \sum_{k,\ell=1}^n \mathbb{E}\big[X_k^2X_\ell^2\big] + 2n\sum_{k=1}^n \mathbb{E}\big[X_k^2\bar X^2\big] + n^2\,\mathbb{E}\big[\bar X^4\big] \\
&\quad - 4n\sum_{k=1}^n \mathbb{E}\big[X_k^2\bar X^2\big] + 4n^2\,\mathbb{E}\big[\bar X^4\big] - 4n^2\,\mathbb{E}\big[\bar X^4\big] \\
&= \sum_{k,\ell=1}^n \mathbb{E}\big[X_k^2X_\ell^2\big] - 2n\sum_{k=1}^n \mathbb{E}\big[X_k^2\bar X^2\big] + n^2\,\mathbb{E}\big[\bar X^4\big],
\end{aligned}$$

soit

$$n^2\,\mathbb{E}\big[S^4\big] = \sum_{k,\ell=1}^{n} \mathbb{E}\big[X_k^2 X_\ell^2\big] - 2n\sum_{k=1}^{n}\mathbb{E}\big[X_k^2\bar{X}^2\big] + n^2\,\mathbb{E}\big[\bar{X}^4\big]$$

qui est une identité remarquable reposant simplement sur les symétries de sommation. Pour le premier terme : lorsque $k=\ell$, $\mathbb{E}[X_k^2X_k^2] = \mathbb{E}[X_k^4] = \tau^4$, et lorsque $k\neq\ell$, $\mathbb{E}[X_k^2X_\ell^2] = \mathbb{E}[X_k^2]\,\mathbb{E}[X_\ell^2] = \sigma^2\sigma^2 = \sigma^4$ par indépendance de X_k^2 et X_ℓ^2 ; on a donc

$$\sum_{k,\ell=1}^{n} \mathbb{E}\big[X_k^2X_\ell^2\big] = n\tau^4 + n(n-1)\sigma^4.$$

Pour le second terme, on a

$$\mathbb{E}\big[X_k^2\bar{X}^2\big] = \frac{1}{n^2}\sum_{i,\ell=j}^{n}\mathbb{E}\big[X_k^2X_iX_j\big].$$

Si $i\neq j$, l'un de ces deux indices est différent de k, par exemple $i\neq k$, alors X_i est indépendant de $X_k^2X_j$ et $\mathbb{E}[X_k^2X_iX_j] = \mathbb{E}[X_k^2X_j]\,\mathbb{E}[X_i] = 0$ car on a supposé $m=0$. Si $i=j\neq k$, alors X_i^2 est indépendant de X_k^2 et $\mathbb{E}[X_k^2X_i^2] = \mathbb{E}[X_k^2]\,\mathbb{E}[X_i^2] = \sigma^4$. Finalement (pour $i=j=k$), on a $\mathbb{E}[X_k^4] = \tau^4$. D'où,

$$\mathbb{E}\big[X_k^2\bar{X}^2\big] = \frac{\tau^4}{n^2} + \frac{n-1}{n^2}\,\sigma^4 \qquad \text{et} \qquad 2n\sum_{k=1}^{n}\mathbb{E}\big[X_k^2\bar{X}^2\big] = 2\tau^4 + 2(n-1)\sigma^4.$$

Pour le troisième terme,

$$n^2\,\mathbb{E}\big[\bar{X}^4\big] = \frac{1}{n^2}\sum_{i,j,k,\ell=1}^{n}\mathbb{E}[X_iX_jX_kX_\ell].$$

Lorsque les indices forment un carré (n possibilités), $\mathbb{E}[X_iX_jX_kX_\ell] = \tau^4$; lorsqu'ils forment un brelan ($4n(n-1)$ possibilités), $\mathbb{E}[X_iX_jX_kX_\ell] = \mathbb{E}[X_1^3]\,\mathbb{E}[X_1] = 0$; lorsqu'ils forment deux paires ($3n(n-1)$ possibilités), $\mathbb{E}[X_iX_jX_kX_\ell] = \mathbb{E}[X_1^2]\,\mathbb{E}[X_1^2] = \sigma^4$; lorsqu'ils forment une paire simple ($6n(n-1)(n-2)$ possibilités), $\mathbb{E}[X_iX_jX_kX_\ell] = \mathbb{E}[X_1^2]\,\mathbb{E}[X_1]\,\mathbb{E}[X_1] = 0$; lorsqu'ils sont tous différents ($n(n-1)(n-2)(n-3)$ possibilités), $\mathbb{E}[X_iX_jX_kX_\ell] = \mathbb{E}[X_i]\,\mathbb{E}[X_j]\,\mathbb{E}[X_k]\,\mathbb{E}[X_\ell] = 0$. D'où,

$$n^2\,\mathbb{E}\big[\bar{X}^4\big] = \frac{1}{n^2}\big(n\tau^4 + 3n(n-1)\sigma^4\big) = \frac{\tau^4}{n} + 3\,\frac{n-1}{n}\,\sigma^4.$$

Finalement

$$\begin{aligned}
n^2\,\mathbb{E}\big[S^4\big] &= n\tau^4 + n(n-1)\sigma^4 - 2\tau^4 - 2(n-1)\sigma^4 + \frac{\tau^4}{n} + 3\,\frac{n-1}{n}\,\sigma^4\\
&= \frac{n^2-2n+1}{n}\,\tau^4 + \frac{n^2-2n+3}{n}(n-1)\sigma^4\\
&= \frac{(n-1)^2}{n}\,\tau^4 + \frac{n^2-2n+3}{n}(n-1)\sigma^4.
\end{aligned}$$

D'où

$$\mathrm{Var}\big(S^2\big) = \mathbb{E}\big[S^4\big] - \mathbb{E}\big[S^2\big]^2 = \frac{(n-1)^2}{n^3}\,\tau^4 - \frac{(n-3)(n-1)}{n^3}\,\sigma^4$$

qui tend bien vers 0 lorsque $\tau^4 = \mathbb{E}[X_1^4]$ est fini, ce qui confirme la convergence en probabilité vers σ^2. Remarquons au passage la force de la loi des grands nombres de Kolmogorov : celle-ci permet rapidement de voir que $(S_n^2)_{n\geqslant 1}$ converge vers σ^2 presque sûrement, et donc en probabilité, lorsqu'on a seulement $\sigma^2 < \infty$. Le cas de la variance corrigé s'obtient immédiatement en multipliant les résultats par $n/(n-1)$ ou $(n/(n-1))^2$.

2.2. Estimation par intervalle

Définition 16 (intervalles de confiance). — Soient μ une mesure de probabilité sur un ensemble E, $\pi \in \mathbb{R}$ un paramètre réel de μ, et $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires de loi μ. Soit $\alpha \in]0,1[$. On appelle *intervalle de confiance* de niveau de confiance $1-\alpha$ (ou de seuil α) un intervalle aléatoire $[\Lambda_n^{\min}, \Lambda_n^{\max}]$, où les extrémités $\Lambda_n^{\min} = f_n(X_1, \ldots, X_n)$ et $\Lambda_n^{\max} = g_n(X_1, \ldots, X_n)$ sont des statistiques de l'échantillon $(X_1, \ldots, X_n)$, vérifiant

$$\mathbb{P}\{\pi \in [\Lambda_n^{\min}, \Lambda_n^{\max}]\} = \mathbb{P}\{\Lambda_n^{\min} \leqslant \pi \leqslant \Lambda_n^{\max}\} \geqslant 1-\alpha.$$

Lorsqu'il y a égalité, on parle d'intervalle de confiance exact. Si pour avoir égalité, on doit faire tendre n vers l'infini, on parle d'intervalle de confiance asymptotique, ou approximatif.

Généralement, on part d'un estimateur Λ_n du paramètre π. La valeur observée ℓ_n peut être proche ou éloignée de π. On se donne alors une fourchette $[\ell_n^{\mathrm{m}}, \ell_n^{\max}]$ autour de ℓ_n dont les bornes sont calculées de sorte que pour une proportion $1-\alpha$ des échantillons qu'on peut choisir, π appartient à l'intervalle correspondant. Il faut noter qu'une fois l'échantillon réellement choisi, on n'a plus le choix : ou bien l'estimation par intervalle est satisfaisante, ou bien elle ne l'est pas. Évidemment plus α est petit, plus la proportion des échantillons conduisant à une estimation correcte est grande ; la contreprtie est que pour se faire, la largeur de l'intervalle grandit et l'estimation perd de sa précision supposée.

Exercice 2. — Une jeune chercheuse en Biologie a obtenu 3 souris mutantes sur une portée de 12 et souhaite conclure que la probabilité d'obtenir une souris mutante est 1/4.

(i) Préciser le cadre statistique de ce problème.

(ii) À l'aide de l'abaque, ou de la table correspondante, donnée en annexe, déterminer la valeur observée de l'intervalle de confiance exact au seuil 5 % de cette probabilité.

(iii) Calculer les intervalles de confiance asymptotiques. Comparer.

(iv) Conclure.

3. Notion de test statistique

Nous considérons (E, μ) un espace probabilisé. Ce que nous savons est que $\mu = \mu_\theta$ pour un certain $\theta \in \Theta$ ensemble indexant une famille $\{\mu_t : t \in \Theta\}$ de mesures de probabilité sur E. L'ensemble Θ est appelé *ensemble des états de la nature*. On se demande si $\theta \in \Theta_0$ pour un sous-ensemble Θ_0 de Θ, autrement dit si $\mu = \mu_\theta$ possède certaines caractéristiques particulières exprimées par l'appartenance ou non de θ à Θ_0. Pour répondre à une telle question, on utilise une procédure de test adaptée à la situation. Une telle procédure se présente ainsi :

- On teste

$$H_0 : \theta \in \Theta_0 \qquad \text{contre} \qquad H_1 : \theta \notin \Theta_0 \quad (\text{ou } (\theta \in \Theta_1 = \Theta \backslash \Theta_0).$$

- On se donne un seuil $\alpha \in]0,1[$ (un niveau de confiance $1-\alpha \in]0,1[$) de test qui correspond à la probabilité de rejeter H_0 alors que H_0 est vraie.

- On se donne une règle de décision au seuil α :

$$f_n : E^n \longrightarrow \mathbb{R}, \qquad A_{\alpha,n} \subset \mathbb{R} \text{ (un intervalle)} \quad \text{et} \quad R_{\alpha,n} = \mathbb{R} \backslash A_{\alpha,n},$$

  c'est-à-dire une fonction de décision f_n, une région d'acceptation $A_{\alpha,n}$ et la région de rejet associée $R_{\alpha,n}$ qui dépendent du test, du seuil α et de la taille d'échantillon n qui sont employés.

- On se donne une réalisation $(x_1, \ldots, x_n)$ d'un échantillon de variables $(X_1, \ldots, X_n)$ indépendantes de même loi μ.

- On calcule $\ell_n = f_n(x_1, \ldots, x_n)$ qui est la valeur observée de la statistique de décision du test $\Lambda_n = f_n(X_1, \ldots, X_n)$, et qui correspond à l'échantillon observé.

- Si $\ell_n \in A_{\alpha,n}$, on conserve H_0, autrement dit, les observations faites ne sont pas en flagrante contradiction avec la possibilité que H_0 soit vérifiée. Si $\ell_n \in R_{\alpha,n}$, on rejette H_0, autrement dit, les observations contredisent significativement la possibilité que H_0 soit vérifiée.

Remarque. — *a*) La construction d'un bon test est du ressort du statisticien. Les utilisateurs ne se préoccupent que d'appliquer la procédure qui convient au problème rencontré. Certaines contraintes sont à respecter pour qu'une procédure de test ait un sens. En particulier :

- si H_0 est vraie, alors $\mathbb{P}\{\Lambda_n \in R_{\alpha,n}\} \leqslant \alpha$, c'est-à-dire que la probabilité de rejeter à tort l'hypothèse H_0 (risque de première espèce) doit être faible ;

- si H_0 est fausse, alors $\mathbb{P}\{\Lambda_n \in A_{\alpha,n}\}$ doit être aussi petit que possible, c'est-à-dire que la probabilité de conserver à tort l'hypothèse H_0 (risque de seconde espèce) doit être faible ;

Si le risque de première espèce est généralement bien contrôlé par la procédure et souvent égal à α, le risque de seconde espèce β est moins bien défini et peut être difficile à évaluer (la puissance du test est définie par $1 - \beta$).

Tout ceci est illustré et approfondi sur des exemples. On y trouvera des répétitions, l'introduction de notions qui n'auront pas été présentées formellement, etc.

4. Test du paramètre p d'une loi de Bernoulli

4.1. Cadre

Nous considérons E un ensemble à deux éléments, par exemple $\{0, 1\}$, $\{\text{pile}, \text{face}\}$, etc. L'ensemble des mesures de probabilité, qui sont des lois de Bernoulli, sur E est paramétré par $\Theta = [0, 1]$, le paramètre étant la probabilité affectée à un point privilégié dans E. Soit μ une mesure de probabilité sur E, p son paramètre. Nous cherchons des informations sur p, par exemple, s'il peut être considéré comme égal à une certaine valeur p_0, ou bien s'il peut être considéré supérieur ou inférieur à ce p_0.

Nous disposons au départ de la valeur de référence p_0, d'un seuil de test α et d'observations $(x_1, \ldots, x_n)$ d'un échantillon $(X_1, \ldots, X_n)$ de la loi $\mu = \mathcal{B}(1, p)$. Rappelons que

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

est un estimateur consistant et sans biais de p et que $\bar{x} = (x_1 + \cdots + x_n)/n$ en est une estimation.

4.2. Test bilatéral

Soient $(X_1, \ldots, X_n)$ un échantillon de la loi $\mathcal{B}(1,p)$, $p \in]0,1[$ inconnu, et $\alpha \in]0,1[$ le seuil de test. Soit $p_0 \in]0,1[$. On teste

$$H_0 : p = p_0 \qquad \text{contre} \qquad H_1 : p \neq p_0.$$

On introduit la statistique

$$\Lambda = \frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{\bar{X}_n - p}{\sqrt{p(1-p)/n}} \times \frac{\sqrt{p(1-p)}}{\sqrt{p_0(1-p_0)}} + \frac{p - p_0}{\sqrt{p_0(1-p_0)/n}}$$

D'après le théorème central limite, lorsque n est grand ($n \geqslant 50$, $np \geqslant 10$, $n(1-p) \geqslant 10$), le premier facteur du premier terme est de loi proche de $\mathcal{N}(0,1)$. Ainsi, si H_0 est fausse, Λ a tendance à prendre des valeurs éloignées de 0. Aussi est-on amené à définir la région d'acceptation de la forme

$$A = [-z_{1-\alpha/2}, z_{1-\alpha/2}],$$

où $z_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi normale $\mathcal{N}(0,1)$. La décision se fait par le calcul de la valeur observée de Λ :

$$\ell = \frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)/n}}\,.$$

Si $\ell \in [-z_{1-\alpha/2}, z_{1-\alpha/2}]$, on accepte H_0, sinon, on rejette H_0.

4.3. Tests unilatéraux

Dans le cas du test

$$H_0 : p \geqslant p_0 \qquad \text{contre} \qquad H_1 : p < p_0,$$

on constate que si H_0 est fausse, la statistique Λ a tendance à prendre des valeurs éloignées de 0 par valeurs négatives. On est amené à définir la région d'acceptation de la forme

$$A = [-z_{1-\alpha}, +\infty[,$$

où $z_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi normale $\mathcal{N}(0,1)$. Si $\ell \in [-z_{1-\alpha}, +\infty[$, on accepte H_0, sinon, on rejette H_0.

Dans le cas du test

$$H_0 : p \leqslant p_0 \qquad \text{contre} \qquad H_1 : p > p_0,$$

on constate que si H_0 est fausse, la statistique Λ a tendance à prendre des valeurs éloignées de 0 par valeurs positives. On est amené à définir la région d'acceptation de la forme

$$A =]-\infty, z_{1-\alpha}],$$

où $z_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi normale $\mathcal{N}(0,1)$. Si $\ell \in]-\infty, z_{1-\alpha}]$, on accepte H_0, sinon, on rejette H_0.

Exercice 3. — Considérons la population des personnes atteintes d'un cancer des poumons. Un échantillon de taille 30 de tels personnes comprend une proportion égale à 0,8 de fumeurs. Cette proportion observée est-elle en accord avec l'affirmation selon laquelle 90 % des malades atteints d'un tel cancer sont fumeurs ?

Correction. — Soit p la proportion des fumeurs dans cette population. Nous testons au seuil $\alpha = 0{,}05$

$$H_0 : p = 0{,}9 \qquad \text{contre} \qquad H_1 : p \neq 0{,}9.$$

La statistique de décision observée est

$$\ell = \frac{0{,}8 - 0{,}9}{\sqrt{0{,}9(1-0{,}9)/30}} \approx -1{,}824 \qquad (\ell' \approx -1{,}368)$$

qui est dans l'intervalle d'acceptation $[-1{,}96\,;\ 1{,}96]$ au seuil $\alpha = 0{,}05$. Avec ce seuil, nous sommes amené à conclure que les observations ne sont pas en contradiction significative avec l'hypothèse $p = 0{,}9$.

En revanche, si le seuil est $\alpha = 0{,}1$, alors l'intervalle d'acceptation devient $[-1{,}645\,;\ 1{,}645]$ et on conclut alors que les observations sont en contradiction significative avec l'hypothèse $p = 0{,}9$.

Exercice 4 (expérience de Weldon). — Afin d'étudier l'équilibrage de dés, le biologiste W. F. R. Weldon lança 26 306 fois un groupe de 12 dés. Il considéra comme « succès » l'apparition d'un 5 ou d'un 6, et obtint les résultats suivants :

x	0	1	2	3	4	5	6	7	8	9	10	11	12
n_x	185	1149	3265	5475	6114	5194	3067	1331	403	105	14	4	0

(i) Indiquer la population d'étude, la variable X étudiée, son type, ses valeurs possibles (ou *modalités*).

(ii) Sous des hypothèses raisonnables la loi de X est une loi classique dont l'un des paramètres n'est pas précisément connu, paramètre que nous noterons π. Préciser quelles sont les hypothèses en question et quelle est la loi de X. Indiquer quelles sont l'espérance et la variance de X.

(iii) Que vaudrait π si les dés étaient parfaitement équilibrés ?

(iv) Les relevés nous permettent d'avoir une estimation de π. Définir un *intervalle de confiance* autour de cette valeur en tenant compte du modèle. La valeur « idéale » y appartient-elle ?

Correction. — *Cadre d'étude.* — (i) La population est l'ensemble théorique des lancers de 12 dés, ensemble infini, ou le modèle probabiliste fini correspondant qui n'est pas nécessairement un modèle classique puisqu'il n'y a aucune raison de supposer que toutes les faces de chaque dé soient équiprobables. La variable statistique est le nombre de « succès » (nombre d'apparitions d'un 5 ou d'un 6) lors d'un lancer des 12 dés. Elle est quantitative discrète et ses valeurs possibles, ou modalités, sont $\{0, 1, \ldots, 11, 12\}$.

(ii) Les hypothèses usuelles sont que les dés sont identiques et que les chiffres qu'ils marquent après un lancer sont indépendants. Dans ce cas la variable X a pour loi la loi binomiale de paramètres 12 (le nombre de dés) et π la probabilité pour qu'un dé marque 5 ou 6. Il est bien connu qu'alors $\mathbb{E}[X] = 12\pi$ et $\mathrm{Var}(X) = 12\pi(1-\pi)$.

(iii) Si les dés étaient parfaitement équilibrés, nous aurions bien entendu $\pi = 1/6 + 1/6 = 1/3$.

Statistiques descriptives. — (i) La moyenne de la série d'observations est

$$\frac{1}{n}\sum_{x=0}^{12} n_x x \approx 4{,}053,$$

où $n = 26\,306$ est la taille totale de l'échantillon, sa variance est

$$s^2 = \frac{1}{n} \sum_{x=0}^{12} n_x (x - \bar{x})^2 \approx 2{,}698,$$

sa variance corrigée est

$$s_c^2 = \frac{1}{n-1} \sum_{x=0}^{12} n_x (x - \bar{x})^2 = \frac{n}{n-1}\, s^2 \approx 2{,}698.$$

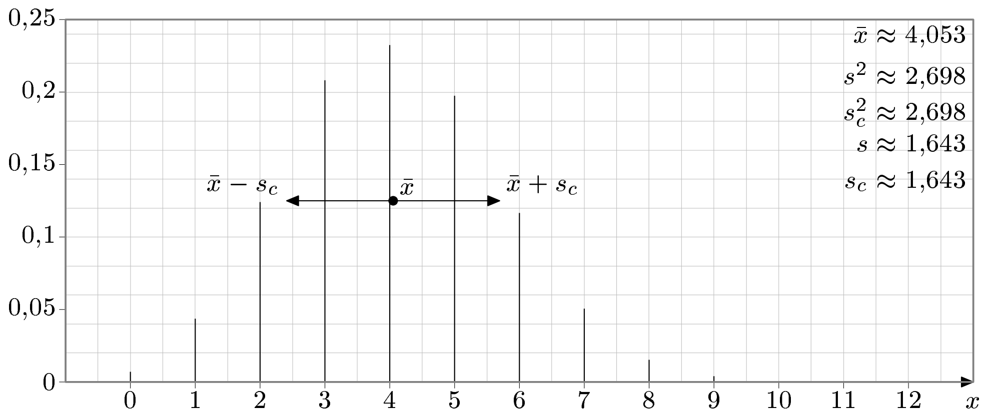
Puisque la taille totale $n = 26\,306$ est grande nous avons $s_c^2 \approx s^2$ (la correction de la variance est approximativement sans effet). De plus on a $\bar{x}(1-\bar{x}/12) \approx 2{,}684 \approx s^2$, ce qui est cohérent avec la relation liant la moyenne et la variance de la variable X.

(ii) Le tableau de distribution des fréquences est :

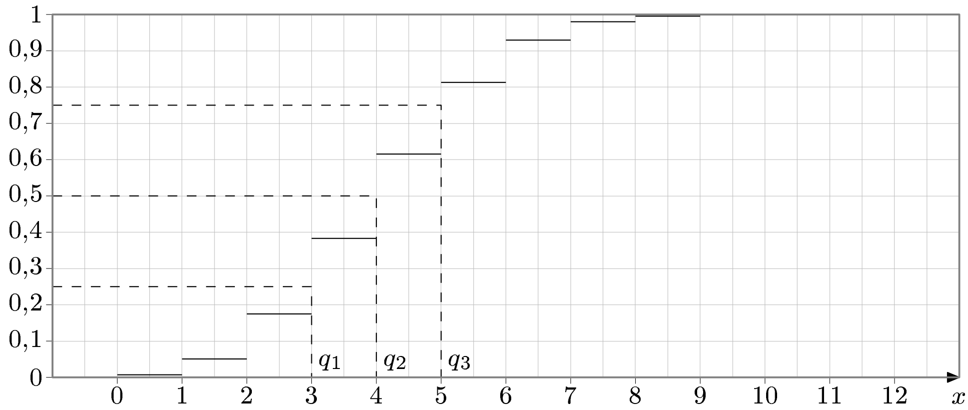
x	0	1	2	3	4	5	6
n_x	185	1149	3265	5475	6114	5194	3067
nc_x	185	1334	4599	10074	16188	21382	24449
f_x	0,0703	0,0437	0,1241	0,2081	0,2324	0,1974	0,1166
fc_x	0,0703	0,0507	0,1748	0,383	0,615	0,8128	0,9294

x	7	8	9	10	11	12	
n_x	1331	403	105	14	4	0	
nc_x	25780	26183	26288	26302	26306	26306	
f_x	0,0506	0,0153	0,004	0,0005	0,00015	0	
fc_x	0,98	0,9953	0,9993	0,9998	1	1	

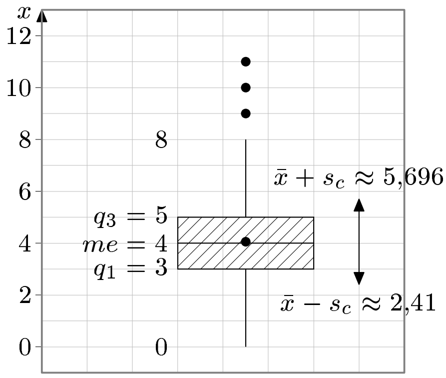
La variable X étant discrète et comportant un nombre faible (13) de modalités, la distribution des fréquences observée se représente avec un diagramme en bâtons :



La fonction de répartition observée est une fonction en escalier dont la représentation graphique est la suivante :



(iii) Sur le graphique précédent nous lisons les trois quartiles de la distribution observée : le premier quartile $q_1 = 3$, la médiane $me = 4$ et le troisième quartile $q_3 = 5$. Ils sont obtenus en déterminant une « image réciproque » des fractions $1/4$, $1/2$, $3/4$ par la fonction de répartition. Pour tracer le diagramme en boîte de la distribution observée, il nous faut déterminer les extrémités des moustaches : l'étendue interquartile vaut $q_3 - q_1 = 2$; multipliée par 1,5 on obtient une amplitude maximale pour les moustaches égale à 3 ; la plus valeur observée supérieure ou égale à $q_1 - 3 = 0$ est 0 qui est donc l'extrémité inférieure du diagramme ; la plus grande valeur observée inférieure ou égale à $q_3 + 3 = 8$ est 8 qui est donc l'extrémité supérieure du diagramme. Ainsi, le diagramme en boîte de la série d'observations est :



Des valeurs sont exclues du diagramme : 9, 8, 10, 11, et évidemment 12 ; elles sont considérées comme rares voire exceptionnelles. Nous avons adjoint la moyenne et représenté l'écart-type corrigé. Les paramètres de position représentés dans ce diagramme sont les trois quartiles, la moyenne et d'une certaine manière les maximum et minimum de la série d'observations représentés par les extrémités des moustaches. Les paramètres de dispersion représentés sont l'étendue interquartile, l'écart-type corrigé, et d'une certaine manière l'étendue par l'écart entre les extrémités des moustaches.

Estimations ponctuelle et par intervalle. — (i) Un n-uplet $(X_1, \ldots, X_n)$ de variables aléatoires est un échantillon de la variable X si ce sont des variables aléatoires définies sur un même espace probabilisé, indépendantes, et de même loi que X.

(ii) Les estimateurs usuels consistants et sans biais de $\mathbb{E}[X]$ et $\mathrm{Var}(X)$ sont respectivement :

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \qquad \text{et} \qquad S_c^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

c'est-à-dire qu'ils sont respectivement de moyenne $\mathbb{E}[X]$ et $\mathrm{Var}(X)$ et que lorsque n tend vers l'infini ils convergent en probabilité respectivement vers $\mathbb{E}[X]$ et $\mathrm{Var}(X)$.

(iii) Dans ce problème, $n = 26\,306$, et les valeurs observées de ces estimateurs sont $\bar{x} \approx 4{,}053$ et $s_c^2 \approx 2{,}698$ et ce sont des estimations ponctuelles des paramètres $\mathbb{E}[X]$ et $\mathrm{Var}(X)$.

(iv) Un intervalle de confiance approximatif de niveau de confiance $1-\alpha$ d'un paramètre λ est un intervalle aléatoire $[\Lambda_1, \Lambda_2]$ où $\Lambda_i = f_i(X_1, \ldots, X_n)$ sont deux variables statistiques telles que

$$\mathbb{P}\{\Lambda_1 \leqslant \lambda \leqslant \Lambda_2\} \approx 1 - \alpha.$$

Puisque n est grand ($n \geqslant 50$) un intervalle de confiance de $\mathbb{E}[X]$ peut être défini par

$$\Big[\bar{X} - u_{1-\alpha/2}\sqrt{S^2/n}\,,\ \bar{X} + u_{1+\alpha/2}\sqrt{S^2/n}\Big]$$

où $u_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi normale $\mathcal{N}(0,1)$. Puisqu'ici $\pi = \mathbb{E}[X]/12$, on en déduit l'intervalle de confiance pour π :

$$\Big[\bar{X}/12 - u_{1-\alpha/2}/12 \times \sqrt{S^2/n}\,,\ \bar{X}/12 + u_{1+\alpha/2}/12 \times \sqrt{S^2/n}\Big].$$

(v) On a $u_{0{,}975} \approx 1{,}96$. Ainsi, l'intervalle de confiance observé de $\mathbb{E}[X]$ au niveau de confiance 0,95 est

$$\big[\bar{x} - u_{1-\alpha/2} \times \sqrt{s^2/n}\,,\ \bar{x} + u_{1+\alpha/2} \times \sqrt{s^2/n}\big] \approx [4{,}033\,;\ 4{,}073],$$

et pour π, c'est

$$\big[\bar{x}/12 - u_{1-\alpha/2}/12 \times \sqrt{s^2/n}\,,\ \bar{x}/12 + u_{1+\alpha/2}/12 \times \sqrt{s^2/n}\big] \approx [0{,}3361\,;\ 0{,}3394].$$

(vi) Ces intervalles de confiance suggèrent que $\mathbb{E}[X]$ est légèrement supérieur à 4 et que π est légèrement supérieur à 1/3. Cependant, nous ne pouvons pas avoir de conclusion définitive puisque les estimations ponctuelles et par voie de conséquence, par intervalles peuvent différer notablement des quantités estimées du fait qu'elles dépendent des observations particulières qui peuvent éventuellement ne pas refléter ce qu'il se passe au niveau de la population. Les estimations de π laissent à penser qu'il est strictement supérieur à 1/3. Ceci n'est guère surprenant si on tient compte du fait que les trous dans les faces des dés conduisent à ce que leurs centres de gravité soient excentrés et plus éloignés des faces les plus trouées 5 et 6.

Test statistique sur la moyenne. — (i) Le test portant sur π consiste à tester l'hypothèse $H_0 : \pi = 1/3$ contre $H_1 : \pi \neq 1/3$. En multipliant par 12, ceci revient à tester $H_0 : 12\pi = 4$ contre $H_1 : 12\pi \neq 4$, c'est-à-dire que nous testerons

$$H_0 : \mathbb{E}[X] = 4 \qquad \text{contre} \qquad H_1 : \mathbb{E}[X] \neq 4.$$

(ii) La statistique du test est

$$Z = \frac{\bar{X} - 4}{\sqrt{S^2/n}},$$

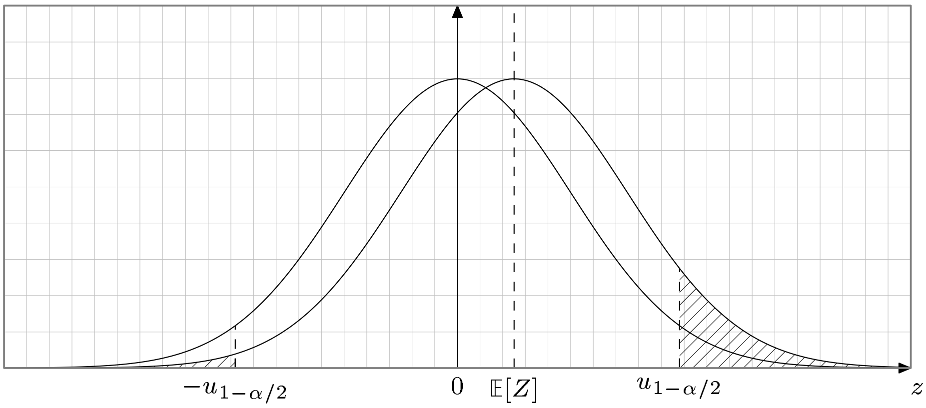
sa valeur observée est

$$z = \frac{\bar{x} - 4}{\sqrt{s^2/n}} \approx \frac{4{,}053 - 4}{\sqrt{2{,}698/26\,306}} \approx 5{,}233.$$

(iii) Si l'hypothèse nulle est vraie, la variable Z est approximativement — compte tenu de la grande valeur de $n = 26\,306$ et du théorème central limite... — de loi la loi normale $\mathcal{N}(0,1)$. La région d'acceptation du test est de la forme $[u_{\alpha/2}, u_{1-\alpha/2}] = [-u_{1-\alpha/2}, u_{1-\alpha/2}]$: les valeurs de Z trop éloignées de 0 conduisent à rejeter H_0. Le risque de première espèce est la probabilité, si H_0 est satisfaite, que Z n'appartienne pas à la région d'acceptation.

Puisque Z est sous cette hypothèse approximativement de loi $\mathcal{N}(0,1)$, par construction cette probabilité est approximativement α.

(iv) Si l'hypothèse H_0 n'est pas satisfaite, la loi de Z est grossièrement centrée autour d'une valeur distincte de 0. Le graphique suivant précise ce que sont alors la puissance du test et l'erreur de seconde espèce :



Hachurée sous la courbe centrée en $\mathbb{E}[Z]$ — courbe représentant une densité approximative de la loi de Z lorsque $\mathbb{E}[Z] \neq 0$, c'est-à-dire lorsque H_0 n'est pas satisfaite —, la surface est égale à la puissance du test, c'est-à-dire la probabilité de rejeter, à raison, H_0 au cours du test. La surface complémentaire représente l'erreur de seconde espèce, c'est-à-dire la probabilité d'accepter, à tort, H_0 au cours du test. On remarque que la puissance est approximativement supérieure à α.

(v) Au seuil $\alpha = 0{,}05$, on a pour intervalle d'acceptation $[-1{,}96\,;\,1{,}96]$. La valeur observée de la statistique du test $z \approx 5{,}233$ n'appartient pas à la région d'acceptation. Nous sommes donc amené à rejeter H_0 : le test est significatif au sens où la série d'observations conduit à rejeter l'hypothèse *a priori* selon laquelle π puisse être égal à $1/3$. Un test unilatéral $H_0 : \mathbb{E}[X] \leqslant 4$ contre $H_1 : \mathbb{E}[X] > 4$ nous aurait de même amené à rejeter H_0. La série d'observations suggère donc que $\pi > 1/3$ ce qui est cohérent avec les justifications concrètes faites lors de l'interprétation des intervalles de confiance.

5. **Test de la moyenne d'une loi normale**

5.1. Cadre

Nous considérons $E = \mathbb{R}$ muni d'une mesure de probabilité μ normale de moyenne m et de variance σ^2. Lorsque σ est connu, le seul paramèter inconnu est la moenne m, aussi l'ensemble des mesures de probabilité qu'on peut envisager est paramétré par $\Theta = \mathbb{R}$, l'ensemble des valeur que peuvent prendre des moyennes de lois normales. Lorsque σ est inconnu, la variance entre alors dans l'ensemble des paramètre et on a alors $\Theta = \mathbb{R} \times \mathbb{R}_+^*$ (une variance nulle est généralement exclue).

Nous cherchons des informations sur m, par exemple, si elle peut être considérée comme égal à une valeur nominale m_0, ou bien si elle peut être considérée supérieure ou inférieure à cette m_0.

Nous disposons au départ de la valeur de référence m_0, d'un seuil de test α et d'observations $(x_1, \ldots, x_n)$ d'un échantillon $(X_1, \ldots, X_n)$ de la loi $\mu = \mathcal{B}(1,p)$. Rappelons que

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

est un estimateur consistant et sans biais de m et que $\bar{x} = (x_1 + \cdots + x_n)/n$ en est une estimation.

Considérons le test bilatéral (les tests unilatéraux étant semblables). Soient $(X_1, \ldots, X_n)$ un échantillon de la loi $\mathcal{N}(m, \sigma^2)$, $m \in \mathbb{R}$ inconnu, $\sigma \in \mathbb{R}_+^*$, et $\alpha \in]0,1[$ le seuil de test. Soit $m_0 \in \mathbb{R}$. On considère le test unilatéral

$$H_0 : m = m_0 \qquad \text{contre} \qquad H_1 : m \neq m_0.$$

5.2. Cas où σ est connu

On introduit la statistique

$$\Lambda = \frac{\bar{X}_n - m_0}{\sigma/\sqrt{n}} = \frac{\bar{X}_n - m}{\sigma/\sqrt{n}} + \frac{m - m_0}{\sigma/\sqrt{n}}$$

Le premier terme a pour loi la loi $\mathcal{N}(0,1)$, ce qui résulte du fait que $\bar{X}_n$ a pour loi $\mathcal{N}(m, \sigma^2/n)$. Ainsi, si H_0 est fausse, Λ a tendance à prendre des valeurs éloignées de 0. Aussi définit-on la région d'acceptation par

$$A = [-z_{1-\alpha/2}, z_{1-\alpha/2}],$$

où $z_{1-\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi normale $\mathcal{N}(0,1)$. La décision se fait par le calcul de la valeur observée de Λ :

$$\ell = \frac{\bar{x}_n - m_0}{\sigma/\sqrt{n}}.$$

Si $\ell \in [-z_{1-\alpha/2}, z_{1-\alpha/2}]$, on accepte H_0, sinon, on rejette H_0.

Exercice 5. — Le diamètre d'une certaine pièce manufacturée suit une loi normale dont la moyenne m n'est pas connue mais d'écart-type σ connu égal à 0,1 cm. Sur 100 pièces issues de la chaîne de fabrication, on a relevé un diamètre moyen de 10,02 cm. Le diamètre normal étant de 10 cm, peut-on penser qu'il y ait un défaut dans la chaîne de fabrication ?

Correction. — Nous testons la moyenne d'une loi normale d'écart-type connu :

$$H_0 : m = 10\,\text{cm} \qquad \text{contre} \qquad H_1 : m \neq 10\,\text{cm}.$$

La statistique de décision observée est alors

$$\ell = \frac{10{,}02 - 10}{0{,}1/\sqrt{100}} = 2.$$

La région d'acceptation au seuil $\alpha = 0{,}05$ étant $[-1{,}96\,;\ 1{,}96]$ et puisque $\ell = 2$ n'y appartient pas, nous rejetons H_0 : les observations conduisent significativement à penser qu'il y a un défaut de fabrication.

5.3. Si σ est inconnu

On introduit la statistique

$$\Lambda = \frac{\bar{X}_n - m_0}{S_n/\sqrt{n-1}}.$$

Lorsque $m = m_0$, sa loi est la loi de Student à $n - 1$ degrés de liberté (*voir* le formulaire sur les lois classiques). De même qu'auparavant, l'intervalle d'acceptation du test bilatéral est de la forme

$$A = [-t_{1-\alpha/2}, t_{1-\alpha/2}],$$

où $t_{1-\alpha/2}$ est le quantile d'odre $1-\alpha/2$ de la loi de Student à $n-1$ degrés de liberté. La décision se fait par le calcul de la valeur observée

$$\ell = \frac{\bar{x}_n - m_0}{s_n/\sqrt{n-1}}\,.$$

Si la valeur observée est notablement différente de 0, c'est-à-dire qu'elle se trouve en dehors de l'intervalle d'acceptation, on rejette H_0, sinon on conserve cette hypothèse.

Exercice 6. — Entre 1851 et 1860, L.-Ad. Bertillon releva la taille de 1 101 178 conscrits. Il put observer que la distribution des tailles était indubitablement normale et observa une moyenne de 163,814 cm et un écart-type corrigé de 6,158 cm. On relève dans un journal que les jeunes gens français durant cette période avait une taille moyenne de 1,65 m. Peut-on accepter cette affirmation, voir émettre un avis plus précis ?

Correction. — Nous devons estimer la moyenne d'une loi normale d'écart-type inconnu. Comme $n = 1\,101\,178$ est très grand, la loi de Student à $n-1$ degrés de liberté peut être remplacée par la loi normale $\mathcal{N}(0,1)$. On obtient ainsi pour intervalle de confiance de niveau de confiance $\alpha = 95\,\%$

$$\left[\bar{x} - z_{0{,}975} \times s_{c,n}/\sqrt{n}\,,\ \bar{x} + z_{0{,}975} \times s_{c,n}/\sqrt{n}\right] \approx [163{,}802\,;\ 163{,}826].$$

Le niveau de confiance étant assez large, il semble bien qu'on puisse préciser que la moyenne devait être de 1,638 m. Néanmoins, ça ne fait guère de différence.

EXAMEN, 20 JANVIER 2009,
AMPHI J, 16H–18H

Tout document hormis le formulaire des lois de probabilité classiques et les tables de Probabilité et Statistique est interdit. L'usage d'une calculatrice est autorisé. Les différents exercices sont indépendants. La qualité de la rédaction sera prise en compte dans l'évaluation des copies.

Exercice 1. — Nous considérons une variable aléatoire $X : \Omega \to \{m, \dots, n\}$ uniformément répartie sur les entiers relatifs compris entre m et n inclus, $m \leqslant n \in \mathbb{Z}$.

(i) L'espérance (moyenne) de X est-elle définie et pourquoi (sans calcul) ?

(ii) Si oui, que vaut-elle ?

Correction. — Non disponible.

Exercice 2. — Nous considérons sur $\mathbb{R}$ une fonction f définie par

$$x \in \mathbb{R}, \qquad f(x) = a \times \frac{1}{x^4} \times \mathbb{1}_{[1,\infty[}(x)$$

où on rappelle que la fonction indicatrice d'un ensemble vaut 1 sur celui-ci et 0 ailleurs.

(i) Déterminer la constante a pour que la fonction f soit la densité d'une mesure de probabilité sur $\mathbb{R}$.

(ii) On considère $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle dont la loi, absolument continue, admet pour densité f. Calculer l'espérance $\mathbb{E}[X]$ et la variance $\mathrm{Var}(X)$ de la variable aléatoire X.

(iii) Que valent les probabilités $\mathbb{P}\{X < 1\}$, $\mathbb{P}\{X \geqslant 2\}$, $\mathbb{P}\{2 \leqslant X \leqslant 3\}$?

Correction. — Non disponible.

Exercice 3. — Une pièce de monnaie est lancée 600 fois de suite. À 275 reprises, on obtient pile.

(i) Qu'étudie-t-on ?

(ii) Quel est l'échantillon ?

(iii) Que vaut la valeur estimée $\bar{x}$ du paramètre d'intérêt ?

(iv) On veut tester si la pièce est équilibrée, c'est-à-dire que la probabilité d'obtenir pile est $p = 1/2$. Formuler les hypothèses H_0 et H_1 associées.

(v) Usuellement, la statistique de test utilisée dans cette situation est de la forme

$$\Lambda = \frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)/n}},$$

où ici $n = 600$, qui lorsque $p_0 = p$ est approximativement distribuée selon la loi $\mathcal{N}(0,1)$. Calculer la valeur observée ℓ de la statistique de test.

(vi) Déterminer la région d'acceptation associée au test pour $\alpha = 0.05$ (*voir* tables).

(vii) Conclure.

Correction. — Non disponible.

$$\mathbb{E}[S^2] = \mathbb{E}\Big[\frac{1}{n}\sum_{k=1}^{n}\big(X_k-\bar{X}\big)^2\Big] = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\Big[\big(X_k-\bar{X}\big)^2\Big].$$

Pour $1 \leqslant k \leqslant n$,

$$\begin{aligned}\mathbb{E}\Big[\big(X_k-\bar{X}\big)^2\Big] &= \mathbb{E}\big[X_k^2\big] - 2\,\mathbb{E}\big[X_k\bar{X}\big] + \mathbb{E}\big[\bar{X}^2\big] = \sigma^2 - \frac{2}{n}\sum_{\ell=1}^{n}\mathbb{E}[X_kX_\ell] + \frac{\sigma^2}{n}\\ &= \frac{n+1}{n}\,\sigma^2 - \frac{2}{n}\sum_{\ell=1}^{n}\mathrm{cov}(X_k,X_\ell) = \frac{n+1}{n}\,\sigma^2 - \frac{2}{n}\,\sigma^2 \sum_{\ell=1,\ell\neq k}^{n}\mathrm{cov}(X_k,X_\ell)\\ &= \frac{n-1}{n}\,\sigma^2 - \frac{2}{n}\sum_{\ell=1,\ell\neq k}^{n} 0 = \frac{n-1}{n}\,\sigma^2.\end{aligned}$$

Ainsi, $\mathbb{E}[S^2] = \frac{n-1}{n}\,\sigma^2$ comme annoncé. Passons au calcul de $\mathbb{E}[S^4]$. Le calcul est plus long et peut être conduit de plusieurs manières en fonction des développements de puissances choisis. On a

$$\begin{aligned}\mathbb{E}\big[S^4\big] &= \mathbb{E}\Bigg[\Big(\frac{1}{n}\sum_{k=1}^{n}X_k^2 - \bar{X}^2\Big)^2\Bigg] = \mathbb{E}\Bigg[\Big(\frac{1}{n}\sum_{k=1}^{n}X_k^2\Big)^2\Bigg] - \frac{2}{n}\,\mathbb{E}\Big[\sum_{k=1}^{n}X_k^2\bar{X}^2\Big] + \mathbb{E}\big[\bar{X}^4\big]\\ &= \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}\big[X_k^4\big] + \frac{1}{n^2}\sum_{k,\ell=1,\,k\neq\ell}^{n}\mathbb{E}\big[X_k^2X_\ell^2\big] - \frac{2}{n}\sum_{k=1}^{n}\mathbb{E}\big[X_k^2\bar{X}^2\big] + \mathbb{E}\big[\bar{X}^4\big].\\ &= \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}\big[X_k^4\big] + \frac{1}{n^2}\sum_{k,\ell=1,\,k\neq\ell}^{n}\mathbb{E}\big[X_k^2\big]\,\mathbb{E}\big[X_\ell^2\big] - \frac{2}{n}\sum_{k=1}^{n}\mathbb{E}\big[X_k^2\bar{X}^2\big] + \mathbb{E}\big[\bar{X}^4\big]\\ &= \frac{\mathbb{E}[X_1^4]}{n} + \frac{n-1}{n}\,\mathbb{E}[X_1^2]\,\mathbb{E}[X_1^2] - \frac{2}{n}\sum_{k=1}^{n}\mathbb{E}\big[X_k^2\bar{X}^2\big] + \mathbb{E}\big[\bar{X}^4\big]\\ &= \frac{\tau^4}{n} + \frac{n-1}{n}\,\sigma^4 - \frac{2}{n}\sum_{k=1}^{n}\mathbb{E}\big[X_k^2\bar{X}^2\big] + \mathbb{E}\big[\bar{X}^4\big],\end{aligned}$$

où, pour $k \neq \ell$, $\mathbb{E}[X_k^2X_\ell^2] = \mathbb{E}[X_k^2]\,\mathbb{E}[X_\ell^2] = \sigma^4$ par indépendance de X_k^2 et X_ℓ^2, et $\tau^4 = \mathbb{E}[X_1^4]$.

Pour $1 \leqslant k \leqslant n$, on a

$$\begin{aligned}
\mathbb{E}\big[X_k^2 \bar{X}^2\big] &= \frac{1}{n^2} \sum_{\ell,m=1}^{n} \mathbb{E}\big[X_k^2 X_\ell X_m\big] \\
&= \frac{1}{n^2}\,\mathbb{E}\big[X_k^4\big] + \frac{1}{n^2} \sum_{\substack{m=1\\ m\neq k}}^{n} \mathbb{E}\big[X_k^3 X_m\big] + \frac{1}{n^2} \sum_{\substack{\ell=1\\ \ell\neq k}}^{n} \mathbb{E}\big[X_k^3 X_\ell\big] + \frac{1}{n^2} \sum_{\substack{\ell,k=1\\ \ell\neq k, m\neq k}}^{n} \mathbb{E}\big[X_k^2 X_\ell X_m\big] \\
&= \frac{\tau^4}{n^2} + \frac{2}{n^2} \sum_{\substack{\ell=1\\ \ell\neq k}}^{n} \mathbb{E}\big[X_k^3 X_\ell\big] + \frac{1}{n^2} \sum_{\substack{\ell,k=1\\ \ell\neq k, m\neq k}}^{n} \mathbb{E}\big[X_k^2 X_\ell X_m\big] \\
&= \frac{\tau^4}{n^2} + \frac{2}{n^2} \sum_{\substack{\ell=1\\ \ell\neq k}}^{n} \mathbb{E}\big[X_k^3\big]\,\mathbb{E}\big[X_\ell\big] + \frac{1}{n^2} \sum_{\substack{\ell,k=1\\ \ell\neq k, m\neq k}}^{n} \mathbb{E}\big[X_k^2\big]\,\mathbb{E}\big[X_\ell X_m\big] \\
&= \frac{\tau^4}{n^2} + \frac{2}{n^2} \sum_{\substack{\ell=1\\ \ell\neq k}}^{n} \mathbb{E}\big[X_k^3\big] \times 0 + \frac{1}{n^2} \sum_{\substack{\ell=1\\ \ell\neq k}}^{n} \mathbb{E}\big[X_k^2\big]\,\mathbb{E}\big[X_\ell^2\big] + \frac{1}{n^2} \sum_{\substack{\ell,k=1\\ \ell\neq k, m\neq k, \ell\neq m}}^{n} \mathbb{E}\big[X_k^2\big]\,\mathbb{E}\big[X_\ell X_m\big] \\
&= \frac{\tau^4}{n^2} + 0 + \frac{n-1}{n^2}\,\sigma^4 + \frac{1}{n^2} \sum_{\substack{\ell,k=1\\ \ell\neq k, m\neq k, \ell\neq m}}^{n} \mathbb{E}\big[X_k^2\big]\,\mathbb{E}\big[X_\ell\big]\,\mathbb{E}\big[X_m\big] \\
&= \frac{\tau^4}{n^2} + \frac{n-1}{n^2}\,\sigma^4 + \frac{1}{n^2} \sum_{\substack{\ell,k=1\\ \ell\neq k, m\neq k, \ell\neq m}}^{n} \mathbb{E}\big[X_k^2\big] \times 0 \times 0 \\
&= \frac{\tau^4}{n^2} + \frac{n-1}{n^2}\,\sigma^4.
\end{aligned}$$

Puis, on a

$$\mathbb{E}\big[\bar{X}^4\big] = \frac{1}{n^4} \sum_{i,j,k,\ell} \mathbb{E}[X_i X_j X_k X_\ell]$$

avec $\mathbb{E}[X_1^4] = \tau^4$, $\mathbb{E}[X_1^3 X_2] = 0$, $\mathbb{E}[X_1^2 X_2^2] = \sigma^4$, $\mathbb{E}[X_1^2 X_2 X_3] = 0$, $\mathbb{E}[X_1 X_2 X_3 X_4] = 0$, le choix des indices $\{1,2,3,4\}$ n'étant qu'illustratif, donc

$$\mathbb{E}\big[\bar{X}^4\big] = \frac{1}{n^4}(n\tau^4 + C_4^2 C_n^2\,\sigma^4) = \frac{1}{n^4}\left(n\tau^4 + 6\,\frac{n(n-1)}{2}\,\sigma^4\right) = \frac{\tau}{n^3} + \frac{3(n-1)}{n^3}\,\sigma^4.$$

Finalement,

$$\begin{aligned}
\mathbb{E}\big[S^4\big] &= \frac{\tau^4}{n} + \frac{n-1}{n}\,\sigma^4 - \frac{2}{n} \times n\left(\frac{\tau^4}{n^2} + \frac{n-1}{n^2}\,\sigma^4\right) + \frac{\tau^4}{n^3} + \frac{3(n-1)}{n^3}\,\sigma^4 \\
&= \frac{(n-1)^2}{n^3}\,\tau^4 + \frac{(n^2-2n+3)(n-1)}{n^3}\,\sigma^4,
\end{aligned}$$

d'où

$$\begin{aligned}
\operatorname{Var}\big(S^2\big) &= \frac{(n-1)^2}{n^3}\,\tau^4 + \frac{(n^2-2n+3)(n-1)}{n^3}\,\sigma^4 - \Big(\frac{n-1}{n}\,\sigma^2\Big)^2 \\
&= \frac{(n-1)^2}{n^3}\,\tau^4 - \frac{(n-3)(n-1)}{n^3}\,\sigma^4
\end{aligned}$$

qui tend bien vers 0 lorsque $\tau^4 = \mathbb{E}[X_1^4]$ est fini, ce qui confirme la convergence en probabilité vers σ^2. Remarquons au passage la force de la loi des grands nombres de Kolmogorov : celle-ci permet rapidement de voir que $(S_n^2)_{n\geqslant 1}$ converge vers σ^2 presque sûrement, et donc en probabilité, lorsqu'on a seulement $\sigma^2 < \infty$. Le cas de la variance corrigé s'obtient immédiatement en multipliant les résultats par $n/(n-1)$ ou $(n/(n-1))^2$.