

Probabilités

J-R Licois

Chapitre 1

Introduction et rappels

L'essentiel du cours est consacré au passage du discret au continu, c'est-à-dire d'un univers fini à un univers infini. La situation que vous connaissez est la suivante :

- Un *univers* Ω fini qui représente les résultats possibles d'une expérience aléatoire,
- les éléments de cet univers qui sont les *événements élémentaires*,
- les parties de cet univers qui sont appelées *événements*.

On appelle alors *probabilité* sur Ω toute application $p : \mathcal{P}(\Omega) \mapsto [0; 1]$ qui vérifie

- $p(\emptyset) = 0$, $p(\Omega) = 1$,
- $p(A \cup B) = p(A) + p(B)$ si $A \cap B = \emptyset$.

On montre facilement, si $\Omega = \{\omega_1, \dots, \omega_n\}$, que définir une probabilité p revient à se donner n réels positifs p_i tels que $\sum_{i=1}^n p_i = 1$. On a alors

$$p(\{\omega_i\}) = p_i \quad p(A) = \sum_{\{i | \omega_i \in A\}} p_i.$$

Le cas particulier de l'*équiprobabilité* correspond à $p_i = 1/n$, *i.e* tous les p_i sont égaux ; on a alors $p(A) = \text{Card}(A)/\text{Card}(\Omega)$, formule qui laisse entendre que les probabilités ne sont que du dénombrement ! Cependant vous allez devoir vous convaincre du contraire.

Dès que Ω est infini, les choses se compliquent. La loi de Poisson est un exemple où $\Omega = \mathbb{N}$; on peut encore définir les probabilités des événements élémentaires et remplacer les sommes finies par des sommes infinies, sous réserve qu'elles aient un sens (mathématique). Si Ω n'est pas dénombrable¹, la situation se bloque définitivement. Prenons $\Omega = [0, 1] \subset \mathbb{R}$: un événement élémentaire est un réel quelconque compris entre 0 et 1. Si on cherche à définir une probabilité uniforme sur Ω , c'est-à-dire que tous les réels sont équiprobables, chaque événement élémentaire doit avoir une probabilité nulle (puisque'il y en a une infinité!), mais alors tout événement aura aussi une probabilité nulle ...

Avant de définir une probabilité dans ce cas, il y a un autre problème à envisager : si Ω n'est pas dénombrable, l'ensemble des parties $\mathcal{P}(\Omega)$ est « gigantesque » et il n'est pas raisonnable de définir une probabilité sur cet ensemble. On va donc remplacer $\mathcal{P}(\Omega)$ par des ensembles plus faciles à manipuler.

1.1 Rappels et définitions

1.1.1 Ensemble des parties

Soit E un ensemble ; on note $\mathcal{P}(E)$ l'ensemble des parties de E , c'est-à-dire l'ensemble de tous les parties incluses dans E ; les éléments de $\mathcal{P}(E)$ sont donc des parties de E .

¹Un ensemble est dénombrable s'il peut être mis en bijection avec $\mathbb{N}$ (ou $\mathbb{Z}$, ou $\mathbb{Q}$), *i.e* si on peut compter ses éléments avec des entiers naturels. Par contre $\mathbb{R}$, l'intervalle $[0; 1] \subset \mathbb{R}$ ne sont pas dénombrables.

Un ensemble sera noté, soit par une lettre, soit entre accolades $\{\dots\}$. Par exemple $\{a\}$ représente l'ensemble à un élément, dont le seul élément est a et qui se lit « singleton a » ; $\mathcal{P}(E) = \{X \mid X \subset E\}$. Pour $E = \{a, b, c\}$, on a $\mathcal{P}(E) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, E\}$. Si $\text{Card}(E) = n$ alors $\text{Card}(\mathcal{P}(E)) = 2^n$.

Une *partition* d'un ensemble E est une famille de parties (A_i) deux à deux disjointes ($A_i \cap A_j = \emptyset$ si $i \neq j$) et dont la réunion est E ; par exemple, les ensembles $A_1 = \{1, 4, 7\}$, $A_2 = \{2, 5, 8\}$ et $A_3 = \{3, 6\}$ constituent une partition de $E = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

Soit $A \subset \Omega$; on appelle *fonction indicatrice* de A , notée $\mathbf{1}_A$, la fonction de $\Omega \rightarrow \{0, 1\}$ définie par $\mathbf{1}_A(x) = 1$ si $x \in A$ et $\mathbf{1}_A(x) = 0$ sinon.

1.1.2 Fonctions

Soit $f : E \rightarrow F$ et $Y \subset F$; on appelle image réciproque de Y par f , l'ensemble des $x \in E$ tels que $f(x) \in Y$, noté $f^{-1}(Y)$.

$$f^{-1}(Y) = \{x \in E \mid f(x) \in Y\}.$$

Remarque : la notation $f^{-1}(Y)$ est purement formelle et ne définit pas une certaine fonction qui serait notée f^{-1} .

Soit f une application de E vers F ; on dit que f est

- *injective* si pour tout $y \in F$ il existe au plus un $x \in E$ tel que $f(x) = y$. Il est équivalent de dire que deux éléments distincts de E ont des images distinctes.
- *surjective* si pour tout $y \in F$ il existe au moins un $x \in E$ tel que $f(x) = y$.
- *bijective* si elle est injective et surjective.

Si $f : E \rightarrow F$ est bijective, pour tout $y \in F$ il existe exactement un $x \in E$ tel que $f(x) = y$; on peut alors définir une application de F vers E qui à tout $y \in F$ associe l'unique $x \in E$ tel que $f(x) = y$. Cette application est appelée *bijection réciproque* de f et notée f^{-1} .

Remarque : si f est une bijection de E vers F alors l'image réciproque $f^{-1}(Y)$ est exactement l'image de Y par la bijection réciproque f^{-1} .

On a donc l'équivalence $y = f(x) \iff x = f^{-1}(y)$. Par exemple, pour la fonction exponentielle $\mathbb{R} \rightarrow]0; +\infty[$ la bijection réciproque est la fonction logarithme népérien $\ln$.

Exercice 1. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ définie par $f(x) = x^2$. Compléter

- $f^{-1}([1, 9]) =$
- $f^{-1}(] - 9, 9]) =$
- $f^{-1}(] - \infty, 0]) =$

Exercice 2. Soit $f : E \rightarrow F$, A et B des parties de F . Montrer que $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$.

Soient $f : E \rightarrow F$ et $g : F \rightarrow G$, l'application h définie sur E , à valeurs dans G , définie par $h(x) = g(f(x))$ est appelé composé de f et g , notée $g \circ f$.

Exercice 3.

- Avec les notations précédentes, soit $Z \subset G$; déterminer $h^{-1}(Z)$.
- Si f est une bijection de E dans F , déterminer $f \circ f^{-1}$ et $f^{-1} \circ f$.

1.1.3 Produit cartésien

On appelle produit cartésien de deux ensembles E et F , noté $E \times F$ l'ensemble des couples (x, y) avec $x \in E$ et $y \in F$.

$$E \times F = \{(x, y) \mid x \in E \text{ et } y \in F\}.$$

La définition précédente se généralise de façon immédiate à n ensembles $E_1, \dots, E_n$: le produit cartésien $\prod_{i=1}^n E_i$ est l'ensemble des n -uplets $(x_1, \dots, x_n)$ avec $x_i \in E_i$. Les définitions suivantes sont

données dans le cas d'un produit de deux ensembles mais se généralisent aisément à un produit de n ensembles.

On appelle première (resp. deuxième) projection l'application de $E \times F$ dans E (resp. F) qui à (x, y) associe x (resp. y).

Exercice 4. Soit x (resp. y) fixé et p_1 et p_2 les projections, déterminer $p_1^{-1}(\{x\})$ et $p_2^{-1}(\{y\})$.

Soit $\Omega = \Omega_1 \times \Omega_2$ et f une application de Ω dans $\mathbb{R}$. Pour ω_1 (resp. ω_2) fixé, les fonctions $f_{\omega_1} : \omega \in \Omega_2 \mapsto f(\omega_1, \omega)$ (resp. $f_{\omega_2} : \omega \in \Omega_1 \mapsto f(\omega, \omega_2)$) sont appelées *applications partielles* associées à f ; c'est la même technique qui est utilisée pour définir les dérivées partielles d'une fonction numérique qui sont simplement les dérivées des applications partielles.

Soit f une fonction définie sur Ω à valeurs dans $\mathbb{R}^2$, qui à un élément a associe $f(a) = (x, y)$; les deux fonctions $f_1 : a \mapsto x$ et $f_2 : a \mapsto y$ sont appelées *fonctions coordonnées* associées à f : en notant p_1 et p_2 les projections de $\mathbb{R}^2$ vers $\mathbb{R}$ on a

$$f_1 = p_1 \circ f, \quad f_2 = p_2 \circ f.$$

Chapitre 2

Modèle probabiliste

Dans le cas où Ω n'est pas un ensemble fini il n'est pas raisonnable de travailler avec l'ensemble des parties $\mathcal{P}(\Omega)$ c'est pourquoi on utilise des sous-ensembles de $\mathcal{P}(\Omega)$ appelés σ -algèbre (ou encore *tribu*).

Par exemple, dans $\mathbb{R}$ on considère la *tribu* des boréliens : c'est la plus petite σ -algèbre contenant les intervalles.

Définition 1 Soit $\mathcal{A}$ une σ -algèbre de Ω . Une mesure positive sur $(\Omega, \mathcal{A})$ est une fonction μ définie sur $\mathcal{A}$ à valeurs dans $[0; +\infty]$ qui vérifie la propriété de σ -additivité :

Si les $A_n \in \mathcal{A}$ sont disjoints avec $A = \bigcup_{n \in \mathbb{N}} A_n$ alors $\mu(A) = \sum_n \mu(A_n)$.

Remarque : dans la définition précédente, la série $\sum_n \mu(A_n)$ peut être divergente (la somme vaut alors $+\infty$), d'où l'intérêt que μ puisse prendre la valeur $+\infty$.

- S'il existe une partie $A \in \mathcal{A}$ telle que $\mu(A) < +\infty$, alors $\mu(\emptyset) = 0$.
- Si $A \subset B$, alors $\mu(A) \leq \mu(B)$.
- Si A est limite monotone des $A_n \in \mathcal{A}$, alors $\mu(A) = \lim_n \mu(A_n)$.

Définition 2 Une probabilité sur $(\Omega, \mathcal{A})$ est une mesure positive p telle que $p(\Omega) = 1$.

Un peu de terminologie

- Un espace *mesurable* est le couple formé d'un ensemble Ω et d'une σ -algèbre de Ω .
- Un espace *mesuré* est le triplet formé d'un ensemble Ω , d'une σ -algèbre $\mathcal{A}$ de Ω et d'une mesure μ sur $\mathcal{A}$.
- Un espace *probabilisé* est le triplet formé d'un ensemble Ω , d'une σ -algèbre $\mathcal{A}$ de Ω et d'une probabilité p sur $(\Omega, \mathcal{A})$.

Définition 3 Il existe une mesure μ , dite mesure de Lebesgue, définie sur la tribu borélienne de $\mathbb{R}$ telle que

$$\mu([a; b]) = b - a.$$

Cette mesure est aussi notée dx car c'est la mesure qui intervient lorsqu'on parle d'intégration des fonctions réelles.

Définition 4 Soient $(\Omega, \mathcal{A})$ et $(\Omega', \mathcal{A}')$ deux espaces mesurables. Une application $f : \Omega \rightarrow \Omega'$ est dite mesurable si

$$\forall A' \in \mathcal{A}', f^{-1}(A') \in \mathcal{A},$$

où $f^{-1}(A')$ désigne l'image réciproque de A' , c'est-à-dire $\{a \in \Omega \mid f(a) \in A'\}$.

Définition 5 Soit $(\Omega, \mathcal{A}, p)$ un espace probabilisé et $(\Omega', \mathcal{A}')$ un espace mesurable ; une variable aléatoire X sur $(\Omega, \mathcal{A}, p)$ à valeurs dans Ω' est une application mesurable $X : \Omega \rightarrow \Omega'$.

Théorème 1 Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, p)$ à valeurs dans $(\Omega', \mathcal{A}')$; la fonction p^X définie sur $\mathcal{A}'$ par $p^X(A') = p(X^{-1}(A'))$ est une mesure de probabilité, appelée probabilité image ou distribution de la variable aléatoire. Pour $A' \in \mathcal{A}'$ on a

$$p^X(A) = p(X^{-1}(A)) = p(\{\omega \in \Omega \mid X(\omega) \in A\}).$$

DÉMONSTRATION : Admise. □

Dans toutes les applications, si Ω' n'est pas dénombrable, ce sera un intervalle de $\mathbb{R}$ muni de la tribu borélienne.

Remarque : Dans le cas d'une variable X à valeurs réelles, on notera $p(X = y)$ la quantité $p^X(\{y\}) = p(\{\omega \in \Omega \mid X(\omega) = y\})$ et $p(X < y)$ la quantité $p^X(]-\infty; y]) = p(\{\omega \in \Omega \mid X(\omega) < y\})$.

Dans le cas où $\Omega' = \mathbb{R}^2$, une application mesurable est appelée un *vecteur aléatoire* ; la distribution est appelée *loi conjointe* de X ; On appelle *loi marginales* de X les probabilités images de P^X par les projections p_1 et p_2 de $\mathbb{R}^2$ dans $\mathbb{R}$.

Chapitre 3

Variables discrètes

Une v.a.r X est dite *discrète* si l'univers image $X(\Omega)$ est fini ou dénombrable ; on notera $(x_i)_{i \in I}$ les éléments de $X(\Omega)$ où I est soit un ensemble fini, soit $\mathbb{N}$. La notation $\sum_{i \in I} (\dots)$ désigne une somme finie, si I est fini et sinon la somme de la série de terme général $(\dots)$ supposée convergente¹.

3.1 Espérance

Définition 6 Soit X une v.a.r discrète avec $X(\Omega) = \{x_i \mid i \in I\}$. On appelle *espérance mathématique* de la v.a.r X la quantité

$$E(X) = \sum_{i=1}^n x_i p^X(\{x_i\}) = \sum_{i=1}^n x_i p(\{X = x_i\}).$$

Remarque : Dans le cas où I est un ensemble fini, l'espérance $E(X)$ est toujours définie. Par contre si I est infini (dénombrable) l'expression de $E(X)$ fait intervenir une série ; dire que $E(X)$ existe équivaut à dire que la série considérée est convergente. Il peut donc exister des v.a.r qui n'ont pas d'espérance !

Proposition 3.1.1 (Formule de transfert) Soit X une v.a.r et f une application mesurable de $\mathbb{R}$ vers $\mathbb{R}$; $f \circ X$ est une variable aléatoire réelle et on a, avec les notations précédentes.

$$E(f(X)) = \sum_{i=1}^n f(x_i) p(\{X = x_i\}).$$

DÉMONSTRATION : Notons $X(\Omega) = \Omega' = \{x_i \mid i = 1, \dots, n\}$ et $f(\Omega') = f \circ X(\Omega) = \{y_j \mid j \in J\}$. On a : $E(f(X)) = \sum_j y_j p(\{f(X) = y_j\})$; notons $X_j = f^{-1}(\{y_j\}) = \{x_i \mid f(x_i) = y_j\}$, on a alors $p(\{f(X) = y_j\}) = \sum_{x_i \in X_j} p(X = x_i)$ d'où

$$E(f(X)) = \sum_j y_j \sum_{x_i \in X_j} p(X = x_i) = \sum_j \sum_{x_i \in X_j} y_j p(X = x_i) = \sum_j \sum_{x_i \in X_j} f(x_i) p(X = x_i).$$

puisque $f(x_i) = y_j$. Les X_j constituent une partition de l'univers image $X(\Omega)$, la somme (double) sur les x_i appartenant à tous les X_j est donc la somme sur tous les $x_i \in X(\Omega)$, d'où le résultat. $\square$

Remarque : Le calcul de $E(f(X))$ ne nécessite pas de connaître le loi de $f(X)$ mais seulement celle de X .

¹La série $\sum_{n=0}^{\infty} u_n$ est convergente si la limite ℓ lorsque $N \rightarrow \infty$ de $\sum_{n=0}^N u_n$ existe ; on pose $\sum_{n=0}^{\infty} u_n = \ell$.

Proposition 3.1.2 Soient X et Y deux v.a.r dont les espérances sont bien définies, on a $E(X + Y) = E(X) + E(Y)$

DÉMONSTRATION : On applique la formule de transfert avec la fonction $f : (x, y) \mapsto x + y$. $\square$

Une v.a.r X est dite *centrée* si $E(X) = 0$; si X est une v.a.r ayant une espérance $E(X) = m$, la v.a.r $X - m$ est la v.a.r centrée associée à X .

3.2 Couple de v.a.r

3.2.1 Indépendance

Définition 7 Soit $(\Omega, \mathcal{A}, p)$ un espace probabilisé ; deux évènements A et B sont dits indépendants si on a

$$p(A \cap B) = p(A)p(B).$$

On peut remarquer que, si $p(A) \neq 0$, alors A et B sont indépendants si la probabilité conditionnelle $p(B/A)$ est égale à $p(B)$: autrement dit B a autant de chances de se réaliser que A se soit produit ou non.

Exercice 5. Montrer que si A et B sont indépendants alors A et le complémentaire de B (noté $\overline{B}$ ou $\complement B$) sont aussi indépendants.

Des évènements $A_1, \dots, A_n$ sont dits *indépendants dans leur ensemble* si et seulement si

$$p(A_{i_1} \cap \dots \cap A_{i_k}) = p(A_{i_1}) \dots p(A_{i_k}) \quad \text{pour toute suite } i_1, \dots, i_k \text{ extraite de } \{1, \dots, n\}.$$

Définition 8 On dit que deux v.a.r. X et Y sont indépendantes si, quels que soient les boréliens B_1 et B_2 , les évènements $\{X \in B_1\}$ et $\{Y \in B_2\}$ sont indépendants, c'est-à-dire

$$p(X \in B_1 \text{ et } Y \in B_2) = p(X \in B_1)p(Y \in B_2).$$

Théorème 2 Les deux v.a.r. X et Y sont indépendantes si et seulement si la loi (conjointe) du couple $Z = (X, Y)$ est le produit des lois (marginales) de X et Y .

DÉMONSTRATION : On a les équivalences

$$p^Z = p^X \otimes p^Y \iff p^Z(B_1 \times B_2) = p^X(B_1)p^Y(B_2) \tag{3.1}$$

$$\iff p(X^{-1}(B_1) \cap Y^{-1}(B_2)) = p(X^{-1}(B_1))p(Y^{-1}(B_2)) \tag{3.2}$$

d'où la conclusion. $\square$

Proposition 3.2.1 Soient X et Y deux v.a.r indépendantes définies sur Ω , f et g deux fonctions mesurables $\mathbb{R} \rightarrow \mathbb{R}$; les v.a.r $f(X)$ et $g(Y)$ sont indépendantes.

DÉMONSTRATION : Admise $\square$

Proposition 3.2.2 Si X et Y sont deux v.a.r indépendantes ayant chacune une variance ; on a alors

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

DÉMONSTRATION : On admettra (voir plus loin) que XY admet une espérance. En posant $X(\Omega) = (x_i)_{i \in I}$ et $Y(\Omega) = (y_j)_{j \in J}$, on a

$$E(XY) = \sum_{i,j} x_i y_j p(X = x_i \text{ et } Y = y_j) \quad (3.3)$$

$$= \sum_{i,j} x_i y_j p(X = x_i) p(Y = y_j) \quad \text{indépendance de } X \text{ et } Y \quad (3.4)$$

$$= \sum_i x_i p(X = x_i) \sum_j y_j p(Y = y_j) \quad (3.5)$$

$$= E(X) E(Y). \quad (3.6)$$

□

Avec les notation de la Proposition 3.2.1 on a donc : $E(f(X) g(Y)) = E(f(X)) E(g(Y))$.

3.2.2 Covariance

Définition 9 Soient X et Y deux v.a.r ; on appelle covariance centrée, notée $\text{Cov}(X, Y)$, la quantité $E((X - E(X))(Y - E(Y)))$ si elle existe.

Définition 10 Soit X une v.a.r ; on pose $\text{Var}(X) = \text{Cov}(X, X)$ si cela existe. Si $\text{Var}(X)$ existe, la quantité $\sqrt{\text{Var}(X)}$ est appelée écart-type de X et notée $\sigma(X)$ (ou σ si aucune confusion n'est à craindre).

Proposition 3.2.3 Soit X une v.a.r admettant une variance, $m = E(X)$ et $k \in \mathbb{R}$; on a

- $\text{Var}(kX) = k^2 \text{Var}(X)$;
- $\text{Var}(X) = E(X^2) - m^2$;
- $\text{Var}(X) = E(X(X - 1)) - m^2 + m$.

Une v.a.r X telle que $\text{Var}(X) = 1$ est appelée *réduite*. Soit X une v.a.r admettant une variance, σ son écart-type et m son espérance ; la variable $Y = \frac{X - m}{\sigma}$ est appelée *variable centrée réduite* associée à X .

Définition 11 Soient X et Y deux v.a.r admettant chacune une variance ; on appelle matrice de covariance de X et Y la matrice

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

Proposition 3.2.4 Soient X et Y deux v.a.r admettant chacune une variance ; on a

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) \quad \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$$

Corollaire 3.2.5 Dans le cas où X et Y sont deux v.a.r indépendantes, on a $\text{Cov}(X, Y) = 0$ et $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

3.3 Lois discrètes usuelles

1. Loi de Bernoulli.

Une variable X suit une loi de Bernoulli, si elle est à valeurs dans $\{0, 1\}$ avec

$$p^X(1) = p, \quad p^X(0) = 1 - p = q, \quad p \in [0; 1].$$

On dit aussi que X est une variable indicatrice. On a :

$$E(X) = p, \quad \text{Var}(X) = pq.$$

2. Loi binomiale.

Une variable X suit une loi binomiale de paramètres n et p , notée $\mathcal{B}(n, p)$, si

$$\forall x \in \{0, 1, \dots, n\} \subset \mathbb{N}, \quad p^X(x) = C_n^x p^x (1-p)^{n-x}$$

Soient n v.a.r indépendantes suivant la même loi de Bernoulli de paramètre p , la variable définie par $X = \sum_{i=1}^n X_i$ suit une loi binomiale $\mathcal{B}(n, p)$; on a

$$E(X) = np, \quad \text{Var}(X) = np(1-p).$$

On en déduit que si X suit $\mathcal{B}(n, p)$ et Y suit $\mathcal{B}(m, p)$, alors $X + Y$ suit $\mathcal{B}(m+n, p)$

3. Loi hypergéométrique.

Une variable X suit une loi hypergéométrique de paramètres N, n, p si

$$p^X(x) = \frac{C_{Np}^x C_{Nq}^{n-x}}{C_N^n} \quad \text{avec } q = 1 - p.$$

On a

$$E(X) = np, \quad \text{Var}(X) = \frac{N-n}{N-1} npq.$$

Dans une population de N individus dont une proportion p possède un certain caractère C , on tire un échantillon (sans remise) de n individus : le nombre d'individus de l'échantillon suit la loi hypergéométrique de paramètres N, n, p .

4. Loi géométrique.

Une variable X suit une loi géométrique de paramètre p , si elle est à valeurs dans $\mathbb{N} \setminus \{0\}$ avec

$$\forall x \in \mathbb{N} \setminus \{0\}, \quad p^X(x) = p(1-p)^{x-1}.$$

On a

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

C'est la loi du *premier succès* dans des épreuves répétées indépendantes de probabilité p .

5. Loi de Pascal, loi de Pólya. Une variable X loi de Pascal de paramètres p et r si

$$p^X(x) = C_{k-1}^{r-1} p^r (1-p)^{k-r}.$$

La loi de Pascal est le nombre d'essais qu'il faut effectuer pour obtenir r succès dans des épreuves répétées indépendantes de probabilité p . C'est donc la somme de r loi géométriques indépendantes. On en déduit

$$E(X) = \frac{r}{p}, \quad \text{Var } X = \frac{r(1-p)}{p^2}.$$

La loi binomiale négative, ou loi de Pólya, est le nombre d'échecs qu'il faut effectuer pour obtenir r succès dans des épreuves répétées indépendantes de probabilité p .

6. Loi de Poisson.

Une variable X suit une loi de Poisson, de paramètre réel $\lambda > 0$, notée $\mathcal{P}(\lambda)$, si

$$\forall x \in \mathbb{N}, \quad p^X = e^{-\lambda} \frac{\lambda^x}{x!}.$$

On a alors

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

Soient X et Y deux v.a.r indépendantes suivant respectivement $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$; la variable $X + Y$ suit $\mathcal{P}(\lambda + \mu)$.

Exercice 6. Déterminer la loi de probabilité, l'espérance et la variance d'une loi de Pólya de paramètres r et p .

3.4 Introduction aux v.a.r continues.

Il est clair que la plupart des définitions précédentes ne peuvent s'étendre au cas des variables continues... Un des rares outils qui s'utilise avec des variables discrètes ou continues est la fonction de répartition. Une autre question est celle de la convergence de variables discrètes : nous étudierons en détail ceci au Chapitre 6 mais nous commençons ici par la plus naturelle.

3.4.1 Fonction de répartition

Définition 12 Soit X une v.a.r; on appelle fonction de répartition de X , notée F_X , la fonction définie par

$$F_X(x) = p(X \leq x) = p^X([-\infty; x]).$$

Proposition 3.4.1 La fonction de répartition F_X d'une v.a.r X

- est croissante (au sens large),
- est continue à droite,
- vérifie $\lim_{-\infty} F = 0$ et $\lim_{+\infty} F = 1$.

DÉMONSTRATION : Ces propriétés sont bien connues dans le cas discret et se démontrent de manière analogue dans le cas continu. $\square$

Remarque : certains auteurs définissent la fonction de répartition par $p^X([-\infty; x])$, la fonction est alors continue à gauche.

Exercice 7. On jette un dé à 6 faces numérotées de 1 à 6, supposé bien équilibré; définir la fonction de répartition de la variable X égale au numéro de la face supérieure.

Remarque : dans le cas discret, la fonction de répartition F_X définit entièrement la variable X ; en effet les points de discontinuités de F_X définissent l'univers image et les probabilités correspondantes sont égales au « saut » de f , différence entre la limite à droite et la limite à gauche de F_X .

Exercice 8. Montrer que, pour $a, b \in \mathbb{R}$, $a < b$, $p(a < X \leq b) = F_X(b) - F_X(a)$.

Proposition 3.4.2 On a, pour $a \in \mathbb{R}$, $p(X < a) = \lim_{x \rightarrow a^-} F_X(x)$.

DÉMONSTRATION : Admise $\square$

Corollaire 3.4.3 *Pour $a \in \mathbb{R}$, on a $P(X = a) = F_X(a) - \lim_{x \rightarrow a^-} F_X(x)$ et en particulier $P(X = a) = 0$ si F_X est continue en a .*

Si $Z = (X, Y)$ est un vecteur aléatoire à valeurs dans $\mathbb{R}^2$ la fonction de répartition de Z est par définition la fonction $F_Z : (x, y) \mapsto p(X \leq x \text{ et } Y \leq y) = p^Z([-\infty, x] \times [-\infty, y])$

Exercice 9. Avec les notations précédentes, montrer que X et Y sont indépendantes si et seulement si la fonction de répartition F_Z est le produit des fonctions de répartition F_X et F_Y .

Exemple : jet de deux dés et loi du sup.

Nous étudierons au Chapitre 6 la convergence des v.a.r ; on peut néanmoins donner dès maintenant la définition suivante :

Définition 13 *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r discrètes à valeurs dans le même ensemble $\Omega' = \{x_1, \dots, x_k, \dots\}$; on dira que $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si et seulement si,*

$$\lim_{n \rightarrow \infty} p(X_n = x_k) = P(X = x_k).$$

Exemple : Approximation de la loi binomiale par la loi de Poisson : pour n grand ($n > 50$), p petit ($p < 0,1$) et $np < 10$.

Chapitre 4

V.a.r continues

4.1 Densité

Soit X une v.a.r et F_X sa fonction de répartition ; on rappelle que pour $a, b \in \mathbb{R}$, $a < b$,

$$p(a < X \leq b) = F_X(b) - F_X(a).$$

Proposition 4.1.1 *On a, pour $a \in \mathbb{R}$,*

$$\begin{aligned} p(X < a) &= \lim_{x \rightarrow a^-} F_X(x), \\ P(X = a) &= F_X(a) - \lim_{x \rightarrow a^-} F_X(x). \end{aligned}$$

En particulier $P(X = a) = 0$ si F_X est continue en a .

DÉMONSTRATION : Admise

□

Ceci montre que si F_X est « raisonnable », il n'y a aucun espoir de pouvoir définir les probabilités des événements élémentaires : la bonne généralisation vient de la notion de densité, celle que l'on utilise lorsqu'on trace un histogramme !

4.1.1 Introduction

Étudions d'abord un exemple ; on a tiré 100 réels au hasard dans l'intervalle $[0; 1[$ et trié ces nombres par ordre croissant :

.0251993610, .0323872300, .0351350660, .0412616320, .052718860, .0547581865, .0597122722, .0657896384, .0670605410, .0704602411, .0841258422, .0870767577, .0932995624, .1349121969, .1549126680, .1687044139, .1872552414, .1946404552, .2087353384, .2221661095, .2235759057, .2236093056, .2344502692, .2491072884, .2550506145, .2568350616, .2605821940, .2735740995, .2972255555, .3169416632, .3220572293, .3244371666, .3263527636, .3292874498, .3312914882, .3331280990, .3478777048, .3512521904, .3642533782, .3729093675, .4104243813, .4225672762, .4252373834, .4335992295, .4648379030, .4689144526, .4781081141, .4855318020, .5214101451, .5327170429, .5485271105, .5495528887, .5589958833, .5642745237, .5680466826, .5838920174, .6027955658, .6063862735, .6123201931, .6207123021, .6217577345, .6311640406, .6325516909, .6420653296, .6511981067, .6595012473, .6627421854, .6777903170, .6814076415, .6935178771, .6981468984, .7371619734, .7454273113, .7772214376, .7841174551, .7885538678, .8105571403, .8253910356, .8268041852, .8326442705, .8357984960, .8524973108, .8528295062, .8560694385, .8634897942, .8716516347, .8741669464, .8767291126, .8858054838, .8879748579, .8972161318, .8987248808, .9059502929, .9081688982, .9401000606, .9476354357, .9529224743, .9629177911, .9854799213, .9962856220

On va réaliser plusieurs histogrammes (des fréquences) de cette même série, en calculant à chaque fois la densité correspondant à chaque rectangle (*i.e* la hauteur de celui-ci).

Cas 1

Classes	n_i	$d_i = f_i/a_i$
$[0; 1/2[$	48	0,96
$[1/2; 1[$	52	1,04

Cas 2

Classes	n_i	$d_i = f_i/a_i$
$[0; 1/4[$	24	0,96
$[1/4; 1/2[$	24	0,96
$[1/2; 1[$	52	1,04

Cas 3

Classes	n_i	$d_i = f_i/a_i$
$[0; 1/4[$	24	0,96
$[1/4; 0,35[$	13	1,3
$[0,35; 0,40[$	3	0,6
$[0,40; 1/2[$	8	0,8
$[1/2; 1[$	52	1,04

On constate que la densité est à peu près constante et ne dépend ni du découpage en classes ni de la classe choisie. De même si l'on considère un histogramme particulier, la somme des aires de tous les rectangles vaut 1 ; en outre cette aire représente aussi l'intégrale de la fonction en escalier qui représente la densité d .

Intéressons-nous maintenant à la fonction de répartition associée $F : x \mapsto p(X < x)$: on sait que c'est une fonction en escalier (*i.e* constante par morceaux) qui correspond aux fréquences cumulées. Par exemple

- Cas 1 : $F(0,5) = 0,48 = 0,96 \times 0,5$.
- Cas 2 : $F(0,5) = 0,48 = 0,24 + 0,24 = 0,96 \times 0,25 + 0,96 \times 0,25$.
- Cas 3 : $F(0,5) = 0,48 = 0,24 + 0,13 + 0,03 + 0,08 = 0,96 \times 0,25 + 0,6 \times 0,05 + 0,8 \times 0,1$.

On remarque que dans chaque cas, $F(x)$ est une somme de termes $d(x)(x_{i+1} - x_i)$ où les x_i sont les limites de classes contenues dans $[0; x[$. On a donc montré la propriété importante suivante

La valeur de $F(x)$ est égale à **l'intégrale** de 0 à x de la **densité** d .

4.1.2 Définition

Parmi toutes les v.a.r nous allons étudier celles dont la fonction de répartition possède des propriétés particulières permettant de définir une densité.

Définition 14 Soit X une v.a.r, F_X sa fonction de répartition. On dit que X est une v.a.r absolument continue¹ s'il existe une fonction f définie sur $\mathbb{R}$ vérifiant

1. $\forall x \in \mathbb{R}, f(x) \geq 0$,
2. f est continue sur $\mathbb{R}$, sauf peut-être en un nombre fini de points où elle admet une limite à droite et une limite à gauche,
3. $\int_{-\infty}^{+\infty} f(t)dt = 1$,
4. $\forall x \in \mathbb{R}, F_X(x) = \int_{-\infty}^x f(t)dt$.

On dit alors que f est une densité de X .

Remarques

- f détermine entièrement F .
- F détermine f à un ensemble de mesure nulle près.
- Si F est dérivable en un point x , alors $f(x) = F'_X(x)$.

¹En toute rigueur, on devrait ajouter « par rapport à la mesure de Lebesgue », car c'est elle qui intervient dans les items 3 et 4.

Exemple : Soit X la v.a.r dont la fonction de répartition F est
$$\begin{cases} F(x) = 0 & \text{si } x \leq 0, \\ F(x) = x & \text{si } x \in]0; 1[, \\ F(x) = 1 & \text{si } x \geq 1. \end{cases}$$

Une densité pour X est, par exemple, la fonction f définie par $f(x) = 1$ si $x \in [0; 1]$ et $f(x) = 0$ sinon ; mais la fonction f_1 définie par $f_1(x) = 1$ si $x \in]0; 1[$ et $f_1(x) = 0$ sinon, est aussi une densité pour F .

Historiquement, et contrairement à ce que pourrait laisser penser le Chapitre 2, la notion de mesure a été inventée avant le calcul des probabilités. Les mathématiciens cherchaient alors à généraliser la notion d'intégrale et à étudier les propriétés (en particulier la dérivabilité) de $\int_{-\infty}^x g(t) d\mu$ avec le minimum d'hypothèses sur g .

Exercice 10. Soit X une v.a.r absolument continue ayant pour densité f ; montrer que la v.a.r définie par $Y = aX + b$ ($(a, b) \in \mathbb{R} * \times \mathbb{R}$) est absolument continue avec pour densité $g = \frac{1}{|a|} f\left(\frac{x-b}{a}\right)$.

Mode d'emploi De façon un peu caricaturale, pour passer du discret au continu, on remplacera les expressions du type $\sum(\dots)p(X=x)$ par $\int(\dots)f(t)dt$ où f est la densité.

Exercice 11. Soit X une v.a.r possédant la densité f et $x_0 \in \mathbb{R}$. Montrer que si f est continue,

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} p(x_0 < X < x_0 + \varepsilon) = f(x_0).$$

4.2 Exemples

4.2.1 La loi uniforme

Soit $I = [a; b]$ un intervalle de $\mathbb{R}$. La fonction f définie par

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

est la densité d'une v.a.r X appelée *loi uniforme sur $[a; b]$* . Vérifions : f est positive, f est continue en tout point sauf en $x = a$ et $x = b$ où elle admet bien une limite à droite et une limite à gauche. Enfin

$$\int_{-\infty}^{\infty} f(t)dt = \int_a^b f(t)dt = \int_a^b \frac{dt}{b-a} = \frac{t}{b-a} \Big|_a^b = 1.$$

La fonction de répartition est alors définie par

$$\begin{cases} F(x) = 0 & \text{si } x < a, \\ F(x) = \frac{x-a}{b-a} & \text{si } x \in [a; b], \\ F(x) = 1 & \text{si } x > b. \end{cases}$$

Autrement dit la probabilité $p(X < x)$ est proportionnelle à la longueur de l'intervalle $[a; x]$ ce qui est bien conforme à l'intuition.

Exercice 12. On tire au hasard (selon la loi uniforme et de façon indépendante) deux nombres x et y dans l'intervalle $[0; 2]$.

1) a) Exprimer les fonctions de répartition des v.a.r correspondantes.

b) En déduire la fonction de répartition, puis la densité, de la v.a.r S égale à $\sup\{x, y\}$.

2) On considère la v.a.r Z égale à $\sup\{x, 1\}$; calculer $p(Z = 1)$ et en déduire que Z n'est pas absolument continue.

4.2.2 Loi de Laplace

Une v.a.r X suit une loi de Laplace si sa densité est donnée par $f(x) = \frac{1}{2}e^{-|x|}$. Vérifions : f est positive, continue car la fonction valeur absolue est continue.

$$\int_{-\infty}^{\infty} f(x)dx = 2 \int_0^{\infty} f(x)dx \quad \text{car } f \text{ est paire} \quad (4.1)$$

$$= \int_0^{\infty} e^{-x}dx = \left[-e^{-x}\right]_0^{\infty} = 1. \quad (4.2)$$

4.2.3 Loi exponentielle

Une v.a.r suit une loi exponentielle de paramètre $\lambda > 0$ si sa densité est donnée par

$$\begin{cases} f(x) = 0 & \text{si } x < 0, \\ f(x) = \lambda e^{-\lambda x} & \text{si } x \geq 0. \end{cases}$$

Vérifions : f est positive, continue sauf en $x = 0$ où elle admet des limites à droite et à gauche.

$$\int_{-\infty}^{\infty} f(t)dt = \int_0^{\infty} f(t)dt = \lambda \int_0^{\infty} e^{-\lambda t}dt = \left[-\frac{1}{\lambda}e^{-\lambda t}\right]_0^{\infty} = 1.$$

La fonction de répartition est alors donnée par

$$\begin{cases} F_X(x) = 0 & \text{si } x < 0, \\ F_X(x) = 1 - e^{-\lambda x} & \text{si } x \geq 0. \end{cases}$$

Exercice 13. Soit X une v.a.r suivant une loi exponentielle de paramètre λ .

1) Montrer que, pour $h > 0$,

$$p(X > t + h | X > t) = p(X > h)$$

2) Réciproquement, soit X une v.a.r absolument continue vérifiant, pour $y > 0$,

$$p(X > x + y | X > x) = p(X > y).$$

Montrer que X suit une loi exponentielle.

Remarque : Si on interprète X comme la durée de vie d'une particule, la probabilité qu'une particule existante à l'instant t existe encore au bout de h unités est indépendante de l'instant t de départ : on dit qu'on a un phénomène *sans mémoire*.

Exercice 14. Supposons que le nombre d'appels à un standard pendant une durée t suive une loi de Poisson de paramètre λt où λ est un réel strictement positif. Montrer que le temps d'attente du premier appel suit une loi exponentielle.

4.2.4 Loi normale

Une v.a.r X suit une loi normale (on dit aussi loi de Gauss, loi de Laplace-Gauss) de paramètres² m et $\sigma > 0$, notée $\mathcal{N}(m, \sigma)$, si sa densité est définie par

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

La densité f est la célèbre « courbe en cloche ». f est positive et continue sur $\mathbb{R}$; on vérifiera plus loin que $\int_{-\infty}^{\infty} f(t)dt = 1$.

²Certains auteurs et certaines calculatrices prennent comme paramètres m et σ^2 !

4.3 Moments

4.3.1 Rappels d'intégration

Soit f intégrable sur $\mathbb{R}$ et a, b, c trois réels ; on a

$$\int_a^c f(t) dt = \int_a^b f(t) dt + \int_b^c f(t) dt. \quad \text{Relation de Chasles}$$

Soient u et v deux fonctions dérivables sur $[a; b]$, u' et v' étant continues. On a

$$\int_a^b u'(t)v(t) dt = u(b)v(b) - u(a)v(a) - \int_a^b u(t)v'(t) dt. \quad \text{Intégration par parties}$$

La formule reste vraie si a et/ou b sont infinis, en remplaçant $u(b)v(b)$ par la limite lorsque t tends vers b de $u(t)v(t)$.

Proposition 4.3.1 *Soit u une fonction dérivable sur l'intervalle $[a; b]$ et f une fonction continue sur $[a; b]$. On a*

$$\int_a^b f(u(x))u'(x) dx = \int_{u(a)}^{u(b)} f(u) du. \quad \text{Changement de variable}$$

DÉMONSTRATION : Soit F une primitive de f ; la formule de dérivation des fonctions composées donne

$$(F(u(x)))' = f(u(x))u'(x).$$

En intégrant entre a et b on obtient $\int_a^b f(u(x))u'(x) dx = \int_a^b (F(u(x)))' dx$; la dernière intégrale est par définition égale à $F(u(b)) - F(u(a))$, c'est-à-dire $\int_{u(a)}^{u(b)} f(u) du$. $\square$

En pratique dans le terme de gauche, on remplace $u(x)$ par u , $u'(x) dx$ par du (i.e $u'(x) = \frac{du}{dx}$) et on ajuste les bornes en conséquence : pour $x = a$, $u(x) = u(a)$ et pour $x = b$, $u(x) = u(b)$.

On admettra le résultat suivant

$$\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}.$$

Exercice 15. En utilisant le changement de variable $t = \frac{x - m}{\sigma\sqrt{2}}$ montrer que $\int_{-\infty}^{\infty} f(t)dt = 1$ où f est la densité de la loi normale $\mathcal{N}(m, \sigma)$.

Exercice 16. Pour p entier naturel on pose $I_p = \int_0^{\infty} t^p e^{-t^2} dt$. Montrer, à l'aide d'une intégration par parties, la relation $I_p = \frac{2}{p+1} I_{p+2}$. En déduire les valeurs

$I_1 = \frac{1}{2}, \quad I_2 = \frac{\sqrt{\pi}}{4}, \quad I_3 = \frac{1}{2}, \quad I_4 = \frac{3\sqrt{\pi}}{8}.$
--

4.3.2 Espérance et variance

Proposition 4.3.2 *La formule de transfert s'applique aux variables absolument continues sous la forme suivante : soit X une v.a.r absolument continue de densité f et φ une fonction dérivable ; on pose $Y = \varphi(X)$.*

$$E(Y) = \int_{-\infty}^{\infty} \phi(t) f(t) dt.$$

DÉMONSTRATION : Admise. □

En particulier $m_2(X) = E(X^2)$ et $\text{Var}(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$.

Définition 15 *Soit $r \geq 1$ un entier et X une v.a.r absolument continue de densité f . On appelle moment d'ordre r de X , si cela existe, la quantité*

$$m_r(X) = \int_{-\infty}^{\infty} t^r f(t) dt$$

Par définition, l'espérance $E(X)$ est égale au moment d'ordre 1, noté $m_1(X)$; on appelle *moment centré d'ordre r* le moment d'ordre r de $X - E(X)$

$$\mu_r = m_r(X - E(X)).$$

En particulier, si cela existe, on pose $\text{Var}(X) = \mu_2(X)$ et $\sigma(X) = \sqrt{\text{Var}(X)}$, respectivement *variance* et *écart-type* de X .

Remarque : Si X admet un moment d'ordre 2, alors X admet une espérance. En effet

$$\forall t, \quad -t^2 - 1 \leq t \leq t^2 + 1$$

car $t \leq t^2$ pour $t > 1$ et $t \leq 1$ pour $t \in [0; 1]$.

$$\int_{-\infty}^{\infty} t f(t) dt \leq \int_{-\infty}^{\infty} (t^2 + 1) f(t) dt \leq m_2(X) + \int_{-\infty}^{\infty} f(t) dt,$$

cette dernière intégrale valant 1 par définition. On montre de la même façon l'inégalité

$$\int_{-\infty}^{\infty} t f(t) dt \geq -m_2(X) - 1.$$

Exercice 17. On considère la v.a.r X dont la densité est donnée par $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ (*Loi de Cauchy*). Montrer que l'espérance de X n'existe pas.

4.3.3 Exemples

– **Loi uniforme** : Si X suit la loi uniforme sur $[a; b]$ on a :

$$\boxed{E(X) = \frac{a+b}{2} \quad \text{Var}(X) = \frac{(b-a)^2}{12}}$$

En effet, $f(t) = 1/(b-a)$, donc

$$\begin{aligned} E(X) &= \int_a^b t f(t) dt = \int_a^b \frac{t}{b-a} dt = \frac{t^2}{2(b-a)} \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2} \\ E(X^2) &= \int_a^b t^2 f(t) dt = \int_a^b \frac{t^2}{b-a} dt = \frac{t^3}{3(b-a)} \Big|_a^b = \frac{a^2 + ab + b^2}{3} \\ \text{Var}(X) &= \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2}{12} = \frac{(b-a)^2}{12}. \end{aligned}$$

- **Loi exponentielle** : Si X suit une loi exponentielle de paramètre λ , on a :

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

En effet, la densité est donnée par $f(t) = e^{-\lambda t} \lambda$ et $E(X) = \int_0^\infty t e^{-\lambda t} \lambda dt$. En intégrant par parties, avec $u(t) = t$, $u'(t) = 1$, $v'(t) = e^{-\lambda t} \lambda$, $v(t) = -e^{-\lambda t}$, on obtient

$$E(X) = t(-e^{-\lambda t}) \Big|_0^\infty + \int_0^\infty e^{-\lambda t} dt = \frac{-e^{-\lambda t}}{\lambda} \Big|_0^\infty = \frac{1}{\lambda}.$$

De même, en intégrant par parties,

$$\begin{aligned} E(X^2) &= \int_0^\infty t^2 e^{-\lambda t} \lambda dt \\ &= 2t e^{-\lambda t} \Big|_0^\infty + 2 \int_0^\infty t e^{-\lambda t} dt \\ &= \frac{2}{\lambda} \int_0^\infty t e^{-\lambda t} \lambda dt \\ &= \frac{2}{\lambda} E(X) = \frac{2}{\lambda^2}. \\ \text{Var}(X) &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \end{aligned}$$

- **Loi normale** : Si X suit une loi normale de paramètres m et σ on a

$$\boxed{E(X) = m \quad \text{Var}(X) = \sigma^2.}$$

La démonstration est laissée au lecteur.

4.4 Loi dérivées de la loi normale

Exercice 18. Soit X une v.a.r suivant la loi $\mathcal{N}(0,1)$; déterminer la densité de probabilité de la variable $Y = X^2$. Calculer $E(Y)$ et $\text{Var}(Y)$.

Les deux lois ci-dessus sont utilisées pour estimer la variance d'une population à partir d'un échantillon.

Définition 16 Soit $(X_k)_{k \in [1;n]}$ une suite de v.a.r indépendantes de même loi $\mathcal{N}(0,1)$. On pose

$$V_n = \sum_{k=1}^n X_k^2.$$

La loi de la variable V_n est appelée loi du χ^2 à n degrés de liberté ; en bon français : loi du khi-deux.

Définition 17 Soient X suivant une loi normale $\mathcal{N}(0,1)$, V indépendante de X suivant une loi du χ^2 à n degrés de liberté. On pose

$$S = \frac{X}{\sqrt{\frac{V}{n}}}.$$

On dit que S suit une loi de Student.

Corollaire 4.4.1 Si (Y_k) est une suite (finie) de v.a.r indépendantes suivant la loi $\mathcal{N}(m, \sigma)$, alors

$$V_n = \sum_{k=1}^n \left(\frac{Y_k - m}{\sigma} \right)^2$$

suit une loi du χ^2 à n degrés de liberté.

DÉMONSTRATION : $\frac{Y_k - m}{\sigma}$ est la variable centrée réduite associée à Y_k , elle suit donc $\mathcal{N}(0,1)$. $\square$

Proposition 4.4.2 Si X suit une loi du χ^2 à n degrés de liberté, alors

$$E(X) = n, \quad \text{Var}(X) = 2n.$$

Si T suit une loi de Student à n degrés de liberté, avec $n > 2$, alors

$$E(T) = 0, \quad \text{Var}(T) = \frac{n}{n-2}.$$

DÉMONSTRATION : La première affirmation est une conséquence directe de l'exercice 18, puisque les X_k^2 sont indépendantes. On admettra la deuxième. $\square$

Exemple : Une machine fabrique des pièces dont le poids suit une loi normale de moyenne m et de variance σ . Le fabricant de la machine affirme que σ vaut 0,05 ; peut-on lui faire confiance ? Pour estimer σ , on prélève n pièces, on mesure leur poids p_k et on calcule la variance empirique

$$V = \frac{1}{n} \sum (p_k - m)^2.$$

On peut considérer que chaque p_k est la valeur d'une v.a.r Y_k suivant $\mathcal{N}(m, \sigma)$ et considérer la v.a.r

$$V_n = \frac{1}{n} \sum (Y_k - m)^2.$$

Si le prélèvement est fait dans de bonnes conditions, on admettra que les Y_k sont indépendantes. Un échantillon de 20 pièces a donné les poids suivants :

2.059, 1.972, 2.012, 1.928, 1.946, 1.999, 1.871, 1.978, 1.950, 1.999,
2.076, 1.970, 2.008, 2.033, 1.973, 2.107, 2.009, 1.969, 1.978, 2.041 .

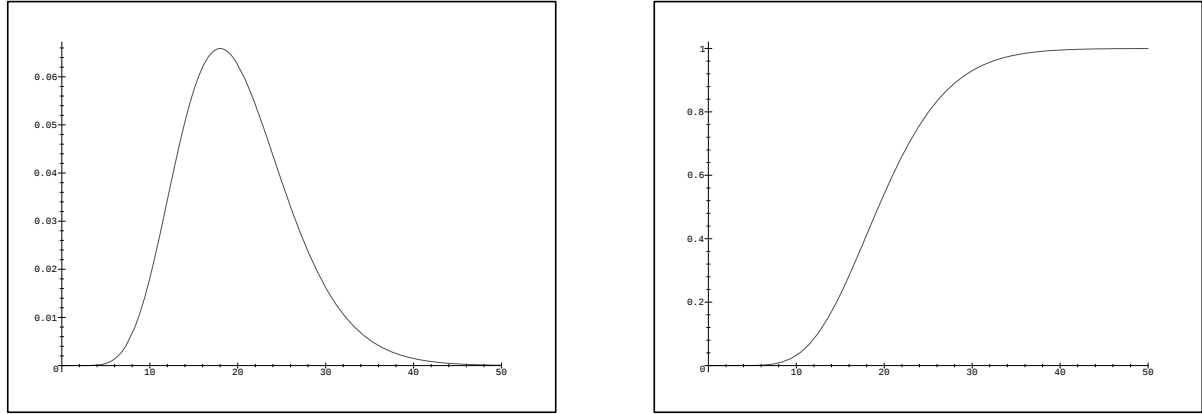
Remarque : La moyenne de l'échantillon ci-dessus est 1,9939 ; elle est différente de 2 et c'est parfaitement logique ! Il n'y a aucune raison que la moyenne d'un échantillon soit systématiquement égale à celle de la population.

PREMIÈRE APPROCHE :

On suppose donc $m = 2$ connue³. On calcule la variance empirique $V_n = \frac{1}{20} \sum (p_k - 2)^2 = 0,05535/20$ (soit un écart-type empirique égal à 0,0526). D'après ce qui précède $n V_n / \sigma^2$ suit une loi du χ^2 à 20 degrés de liberté. Si on suppose $\sigma = 0,05$, la valeur observée $0,05535/0,0025 = 22,14$ est-elle compatible avec l'hypothèse ? On ne peut pas vraiment répondre à cette question ; mais on peut raisonner différemment.

Reprenons le même calcul avec l'affirmation $\sigma \geq 0,075$, la valeur observée est alors *inférieure* à 9,84 et la fonction de répartition du χ^2 à 20 degrés de liberté montre qu'elle est assez peu probable : $F(9,84) = 2,9\%$; de même l'affirmation $\sigma < 0,025$ conduit à une valeur observée *supérieure* à 88,56 qui est quasi-impossible. On peut en conclure que σ est compris entre 0,025 et 0,075 au seuil de 97%.

³Il est illogique de connaître m précisément et pas σ , mais c'est à titre d'exemple !

FIG. 4.1 – Densité et répartition du χ^2 à 20 degrés de liberté.

Abandonnons provisoirement cet exemple et, dans un cas général, calculons l'espérance de V_n . Supposons d'abord m connue; on a alors $E(V_n) = \sigma$; en effet

$$E(V_n) = \frac{1}{n} \sum_{k=1}^n (Y_k - m)^2 = \frac{1}{n} \sum_{k=1}^n \text{Var}(Y_k) = \sigma.$$

Dans ce cas, on dira que V_n est un estimateur *sans biais* de σ .

Supposons maintenant m et σ inconnus. Posons alors

$$M_n = \frac{1}{n} \sum Y_k \quad W_n = \frac{1}{n} \sum (Y_k - M_n)^2$$

Proposition 4.4.3 *On a $E(M_n) = m$, i.e M_n est un estimateur sans biais de m . Par contre $E(W_n) = \frac{n-1}{n} \sigma^2$; un estimateur sans biais⁴ de σ est donc $W'_n = \frac{n}{n-1} W_n$.*

DÉMONSTRATION : La première affirmation est triviale : $E(M_n) = \frac{1}{n} \sum_{k=1}^n E(Y_k) = \frac{1}{n} n m = m$. Considérons la v.a.r $R = Y_1 - M_n$: c'est une variable centrée qui a même loi que toutes les $Y_k - M_n$. On a donc, pour tout k ,

$$E(Y_k - M_n)^2 = E(R^2) \quad \text{et donc} \quad E(W_n) = E(R^2).$$

Comme R est centrée, $\text{Var}(R) = E(R^2) = E(W_n)$. Le problème vient du fait que Y_1 et M_n ne sont pas indépendantes, puisque M_n dépend explicitement de Y_1 . On peut écrire

$$R = Y_1 - \frac{1}{n} \sum_{k=1}^n Y_k = Y_1 - \frac{1}{n} Y_1 - \frac{1}{n} \sum_{k=2}^n Y_k = \frac{n-1}{n} Y_1 - \frac{1}{n} \sum_{k=2}^n Y_k,$$

avec cette fois-ci une somme de variables indépendantes. On a donc

$$\begin{aligned} \text{Var}(R) &= \left(\frac{n-1}{n} \right)^2 \text{Var}(Y_1) + \frac{1}{n^2} \text{Var}(Y_k) \\ &= \left(\frac{n-1}{n} \right)^2 \sigma^2 + \frac{n-1}{n^2} \sigma^2 = \frac{n-1}{n} \sigma^2. \end{aligned}$$

D'où on tire $E(W_n) = \frac{n-1}{n} \sigma^2$ et $E(W'_n) = \sigma^2$. □

⁴C'est pour cela que les calculatrices donnent généralement deux écarts-types : l'un correspond à W_n , l'autre à W'_n .

Théorème 3 (Fisher) Supposons que $X_1, \dots, X_n$ sont des v.a.r indépendantes suivant $\mathcal{N}(0,1)$; on pose $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ et on définit S_n par $S_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2$.

- la v.a.r $\sqrt{n}\bar{X}$ suit $\mathcal{N}(0,1)$.
- la v.a.r $n S_n^2$ suit une loi du χ^2 à $(n-1)$ degrés de liberté.
- les v.a.r $\bar{X}$ et S_n^2 sont indépendantes.

DÉMONSTRATION : Puisque les X_k sont centrées, $\bar{X}$ est centrée; $\text{Var}(\bar{X}) = \frac{1}{n^2} \sum \text{Var}(X_k) = \frac{n}{n^2} = \frac{1}{n}$, donc $\sqrt{n}\bar{X}$ est réduite centrée. On montrera plus tard qu'une somme de lois normales est normale. On admettra les deux autres propriétés. $\square$

Corollaire 4.4.4 La variable $T = \sqrt{n-1} \frac{\bar{X}}{S_n}$ suit une loi de Student. Plus généralement si les X_k suivent $\mathcal{N}(m, \sigma)$ alors $\frac{\bar{X} - m}{S_n} \sqrt{n-1}$ suit une loi de Student à $(n-1)$ degrés de liberté.

DÉMONSTRATION : D'après ce qui précède, $\frac{\bar{X} - m}{\sigma} \sqrt{n}$ suit $\mathcal{N}(0,1)$ et $\frac{n S_n^2}{\sigma^2}$ suit une loi du χ^2 à $(n-1)$ degrés de liberté : le quotient

$$\frac{\frac{\bar{X} - m}{\sigma} \sqrt{n}}{\sqrt{\frac{n S_n^2}{(n-1)\sigma^2}}} = \frac{\bar{X} - m}{S_n} \sqrt{n-1}$$

suit donc une loi de Student à $(n-1)$ degrés de liberté. $\square$

Remarque : Si, au lieu de calculer S_n , on calcule $S_n^* = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$, alors $\frac{\bar{X} - m}{S_n^*} \sqrt{n}$ suit une loi de Student à $(n-1)$ degrés de liberté.

Remarque : L'expression obtenue pour T ne dépend que de m et ne dépend plus de σ .

DEUXIÈME APPROCHE :

Reprenons les données ci-dessus : les hypothèses $m = 2$ et $\sigma = 0,05$ sont-elles raisonnables ? La valeur observée de $\bar{X}$ est alors $\sqrt{20} \frac{39,878 - 40}{20 \times 0,05} = -0,5456$. Une table de la loi normale montre qu'il n'y a pas lieu de rejeter cette hypothèse.

Même question avec $m = 1,9$? La valeur observée est alors $\sqrt{20} (39,878 - 38) = 8,39$ tout à fait invraisemblable. Remarquez que l'on teste simultanément la valeur de m et la valeur de σ , ce qui peut paraître absurde.

Le corollaire et la dernière remarque nous permettent de tester uniquement l'hypothèse $m = 1,9$. Après calcul on obtient avec les données ci-dessus $\bar{X} = 1,9939$, $S_n = 0,0523$. Pour $m = 2$ la valeur de T est -0,5084 et il n'y a pas de raison de rejeter l'hypothèse; pour $m = 1,9$ la valeur de T est 7,826 et les tables de la loi de Student montrent que c'est peu vraisemblable.

4.5 Compléments

Proposition 4.5.1 Soit X une v.a.r dont la fonction de répartition F_X est continue et strictement croissante. La variable $Y = F_X(X)$ suit la loi uniforme sur l'intervalle $[0; 1]$.

DÉMONSTRATION : Par définition F_X , et aussi Y , sont à valeurs dans $[0; 1]$; comme F_X est continue et strictement monotone elle réalise une bijection de $\mathbb{R}$ dans $]0; 1[$. On notera F_X^{-1} la bijection réciproque.

Cherchons la fonction de répartition de Y ; pour $y \in]0; 1[$ on a

$$p(Y \leq y) = p(F_X(X) < y) = p(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y,$$

ce qui montre que Y suit la loi uniforme. □

Il existe des tables fournissant des nombres au hasard qui peuvent être utilisées pour générer des nombres distribués de manière uniforme ; les calculatrices (ou les tableurs) peuvent aussi engendrer des séquences pseudo-aléatoires uniformément distribuées.

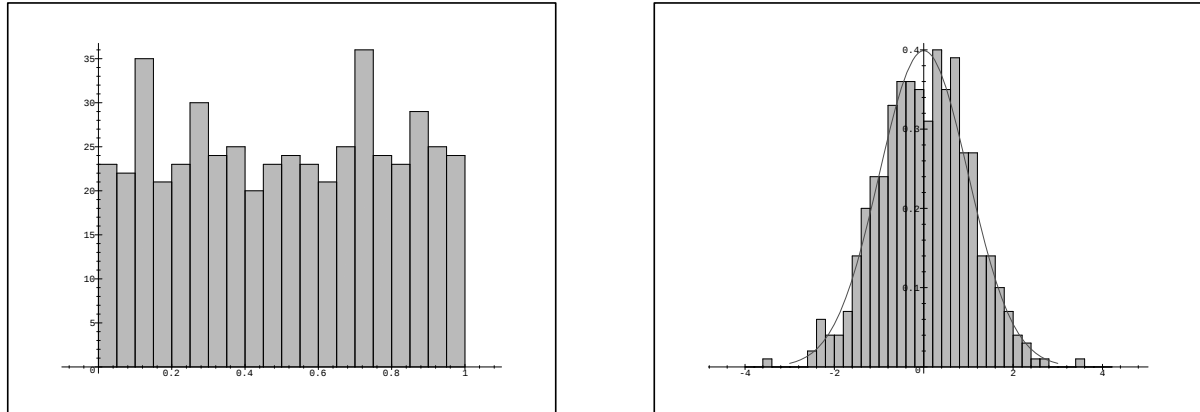


FIG. 4.2 – Génération d'une distribution normale à partir d'une distribution uniforme.

Si on dispose d'une telle distribution, on peut en déduire moyennant le résultat suivant, une suite distribuée selon une loi arbitraire dont on connaît la fonction de répartition.

Corollaire 4.5.2 Soient $y_1, \dots, y_n$ n réels uniformément distribués dans $[0; 1]$; on pose $x_i = \sup\{x \mid F_X(x) < y_i\}$. Les x_i sont distribués selon la loi de X .

Remarque : La formulation ci-dessus est adaptée à une loi dont la fonction de répartition est tabulée (*i.e* donnée dans une table) ; si F_X^{-1} est explicitement connue, on peut utiliser directement $x_i = F_X^{-1}(y_i)$.

Exemple : On a tiré 500 réels y_i au hasard dans l'intervalle $[0; 1]$, puis appliqué la méthode ci-dessus à partir d'une table de la loi normale pour obtenir les x_i . La figure ci-dessus montre l'histogramme de la série des y_i et celui de la série des x_i superposée à la distribution théorique.

Remarque : La loi du χ^2 est aussi utilisée pour tester la nature d'une distribution. Sur l'exemple ci-dessus, on s'attend à ce que chaque classe de l'histogramme des y_i (qui suivent une loi uniforme) contienne 25 éléments. En réalité les effectifs observés sont

24, 30, 21, 35, 24, 23, 23, 24, 25, 21, 24, 29, 25, 36, 23, 23, 23, 25, 20, 22.

Si on note N_i les effectifs observés et n_i les effectifs théoriques des k catégories, la quantité

$$D = \sum_{i=1}^k \frac{(N_i - n_i)^2}{n_i}$$

suit approximativement une loi du χ^2 à $(k - 1)$ degrés de liberté. Sur l'exemple, on a $D = 14,08$ ce qui est tout à fait acceptable.

Chapitre 5

Couple de v.a.r

5.1 Intégrale double

5.1.1 Définition

Soit f une fonction continue¹ définie sur le pavé $P = [a; b] \times [c; d]$ à valeurs dans $\mathbb{R}$; on appelle intégrale (double) de f sur P la quantité définie par

$$\iint_P f(x, y) \, dx dy = \int_a^b \left(\int_c^d f(x, y) \, dy \right) dx = \int_c^d \left(\int_a^b f(x, y) \, dx \right) dy.$$

Exemple : Pour $P = [0; 1] \times [1; 2]$ et $f(x, y) = x^2 + xy + y^2$ on a

$$\begin{aligned} \iint_P f(x, y) \, dx dy &= \int_0^1 \left(\int_1^2 (x^2 + xy + y^2) \, dy \right) dx \\ &= \int_0^1 \left([x^2 y + \frac{1}{2} x y^2 + \frac{1}{3} y^3]_{y=1}^{y=2} \right) dx \\ &= \int_0^1 \left(x^2 + \frac{3}{2} x + \frac{7}{3} \right) dx \\ &= \left[\frac{1}{3} x^3 + \frac{3}{4} x^2 + \frac{7}{3} x \right]_0^1 = \frac{41}{12}. \end{aligned}$$

Plus généralement si D est un domaine défini par les conditions $a \leq x \leq b$ et $\phi_1(x) \leq y \leq \phi_2(x)$ où ϕ_1 et ϕ_2 sont des fonctions continues, alors

$$\iint_D f(x, y) \, dx dy = \int_a^b \left(\int_{\phi_1(x)}^{\phi_2(x)} f(x, y) \, dy \right) dx.$$

Si le domaine D peut aussi être défini par les conditions $c \leq y \leq d$ et $\psi_1(y) \leq x \leq \psi_2(y)$, l'intégrale de f sur D est aussi égale

$$\iint_D f(x, y) \, dx dy = \int_c^d \left(\int_{\psi_1(y)}^{\psi_2(y)} f(x, y) \, dx \right) dy.$$

Remarque : Dans le cas particulier où f est la fonction indicatrice de D (qui vaut 1 sur D et 0 en dehors), l'intégrale double de f sur D est, par définition, l'aire de D .

Remarque : Les définitions précédentes s'étendent sans difficulté au cas où $a, b, c, d, \phi_1, \phi_2, \psi_1, \psi_2$ peuvent être infinis, pourvu que les intégrales (impropres) considérées existent.

¹On suppose f continue pour simplifier, mais on peut affaiblir les hypothèses sur f .

5.1.2 Changement de variables

On admettra le théorème suivant

Théorème 4 Soit $\Phi = (\Phi_1, \Phi_2)$ une application dérivable, à dérivées continues, d'un domaine U dans un domaine V dont le Jacobien

$$J(x, y) = \det \begin{pmatrix} \frac{\partial \Phi_1}{\partial x} & \frac{\partial \Phi_1}{\partial y} \\ \frac{\partial \Phi_2}{\partial x} & \frac{\partial \Phi_2}{\partial y} \end{pmatrix}$$

est supposé borné sur U . Si g est une fonction $V \longrightarrow \mathbb{R}$ intégrable, on a

$$\iint_V g(u, v) du dv = \iint_U g \circ \Phi(x, y) |J(x, y)| dx dy$$

Dans la suite on oubliera parfois de vérifier certaines hypothèses dans l'application de ce théorème ! En pratique, on procède comme pour le changement de variable à une variable

- on pose $u = \Phi_1(x, y)$ et $v = \Phi_2(x, y)$ et on exprime la fonction à intégrer en fonction de u et v ,
- on exprime le domaine dans les nouvelles variables,
- on remplace $du dv$ par $|J(x, y)| dx dy$.

Exemple : Le changement de variable en coordonnées polaires. On pose $x = r \cos \theta$ et $y = r \sin \theta$; la matrice Jacobienne est alors

$$\begin{pmatrix} \cos \theta & \sin \theta \\ -r \sin \theta & r \cos \theta \end{pmatrix}$$

et la Jacobien de la transformation est $|r| = r$ car r est positif ou nul. On a alors

$$\iint_V f(x, y) dx dy = \iint_U f(r \cos \theta, r \sin \theta) r dr d\theta.$$

Exemple : Soit D le demi-cercle de centre $(0, 0)$, de rayon 1 défini par $y \geq 0$ et $f(x, y) = y$.

- Calcul en coordonnées cartésiennes. D est défini par $-1 \leq x \leq 1$ et $0 \leq y \leq \sqrt{1-x^2}$; on a donc

$$\begin{aligned} J &= \iint_D f(x, y) dx dy = \int_{-1}^1 \left(\int_0^{\sqrt{1-x^2}} y dy \right) dx \\ &= \int_{-1}^1 \left[\frac{y^2}{2} \right]_0^{\sqrt{1-x^2}} dx \\ &= \int_{-1}^1 \frac{1-x^2}{2} dx \\ &= \left[\frac{x}{2} - \frac{x^3}{6} \right]_{-1}^1 = \frac{2}{3}. \end{aligned}$$

- Calcul en coordonnées polaires. Le demi-cercle est alors défini comme le pavé $P = \{0 \leq r \leq 1\} \times \{0 \leq \theta \leq \pi\}$.

$$\begin{aligned} J &= \iint_P r \sin \theta r dr d\theta \\ &= \int_0^\pi \left(\int_0^1 r^2 \sin \theta dr \right) d\theta \\ &= \int_0^\pi \left[\frac{r^3}{3} \right]_0^1 \sin \theta d\theta \\ &= \int_0^\pi \frac{1}{3} \sin \theta d\theta = \frac{2}{3} \end{aligned}$$

Exercice 19. On pose $J = \iint_{\mathbb{R}^2} e^{-x^2-y^2} dx dy$.

- 1) Montrer, en utilisant les coordonnées cartésiennes, que $J = \left(\int_{\mathbb{R}} e^{-x^2} dx \right)^2$.
- 2) Calculer J en utilisant les coordonnées polaires.
- 3) En déduire que $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$.

5.2 Densité

Définition 18 Une application $\mu : \mathbb{R}^2 \longrightarrow \mathbb{R}$ est une densité de probabilité si et seulement si

- μ est à valeurs positives
- $\iint_{\mathbb{R}^2} \mu(x, y) dx dy = 1$.

On rappelle que la fonction de répartition du couple (X, Y) est la fonction $F : \mathbb{R}^2 \longrightarrow \mathbb{R}$ définie par $F(x, y) = p(\{X \leq x\} \cap \{Y \leq y\}) = p(X \leq x \text{ et } Y \leq y)$.

Définition 19 Un couple (X, Y) de v.a.r admet une densité s'il existe une densité μ telle que, pour tout $x, y \in \mathbb{R}$, on ait

$$F(x, y) = p(\{X \leq x\} \cap \{Y \leq y\}) = \int_{-\infty}^x \int_{-\infty}^y \mu(s, t) dt ds.$$

Proposition 5.2.1 Soit $D \subset \mathbb{R}^2$ un borélien et (X, Y) un couple de v.a.r ayant pour densité μ . On considère l'évènement $(x, y) \in D$; sa probabilité est alors définie par

$$p((x, y) \in D) = \iint_D \mu(x, y) dx dy$$

DÉMONSTRATION : Par définition la propriété est vérifiée si D est de la forme $] -\infty; x] \times] -\infty; y]$ et s'étend immédiatement aux pavés $]a; b] \times]c; d]$. On passe ensuite à la limite. $\square$

Remarque : L'existence d'une densité pour un couple de v.a.r est une propriété très contraignante : si on considère deux v.a.r X et Y quelconques, il y a peu de chances que le couple (X, Y) admette une densité même si on peut déterminer la loi conjointe.

Exemple : Soit X une v.a.r; le couple (X, X^2) ne possède pas de densité. Posons $Y = X^2$ et considérons l'évènement $\{Y - X^2 = 0\}$; c'est un évènement certain et sa probabilité est donc 1. Si le couple (X, Y) possède une densité μ on doit avoir

$$p\{Y - X^2 = 0\} = \iint_P \mu(x, y) dx dy,$$

où P désigne la parabole définie par $y = x^2$: mais cette dernière intégrale est nulle car la parabole P est un ensemble de mesure nulle (i.e. elle n'a pas d'« épaisseur »!).

Remarque : Si le couple admet une densité μ , on a

$$\mu(x, y) = \frac{\partial^2 F}{\partial x \partial y}(x, y),$$

où F est la fonction de répartition du couple.

Proposition 5.2.2 Soit (X, Y) un couple de v.a.r ayant pour densité μ , alors X et Y ont pour densités marginales

$$f(s) = \int_{-\infty}^{+\infty} \mu(s, t) dt \quad g(t) = \int_{-\infty}^{+\infty} \mu(s, t) ds$$

DÉMONSTRATION : Considérons l'évènement $\{X \leq a\}$ et soit $D = \{(x, y) \mid x \leq a\} =]-\infty, a] \times \mathbb{R}$. On a

$$p(X \leq a) = \iint_D \mu(x, y) dx dy = \int_{-\infty}^a \left(\int_{\mathbb{R}} \mu(x, y) dy \right) dx.$$

En posant $f(x) = \int_{\mathbb{R}} \mu(x, y) dy$ on voit que $p(X \leq a) = \int_{-\infty}^a f(x) dx$ ce qui montre bien que f est une densité pour X . $\square$

Définition 20 Soit (X, Y) un couple de v.a.r ayant pour densité μ ; on suppose que X admet une densité f . La densité de Y sachant $\{X = x\}$, notée g_x vérifie

$$\mu(x, y) = f(x) g_x(y).$$

En particulier, on a la propriété suivante

Proposition 5.2.3 Soit (X, Y) un couple de v.a.r ayant pour densité μ , f et g les densités marginales de X et Y . X et Y sont indépendantes si et seulement si $\mu(x, y) = f(x) g(y)$ pour tout $(x, y) \in \mathbb{R}^2$.

DÉMONSTRATION : Si $\mu(x, y) = f(x) g(y)$ alors $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y \mu(x, y) dx dy$ ce qui se traduit par

$$F(x, y) = \int_{-\infty}^x \left(\int_{-\infty}^y f(x) g(y) dy \right) dx = \int_{-\infty}^x \left(f(x) \int_{-\infty}^y g(y) dy \right) dx = \left(\int_{-\infty}^x f(x) dx \right) \int_{-\infty}^y g(y) dy.$$

ce qui montre que la fonction de répartition de (X, Y) est le produit des fonctions de répartition de X et Y d'où l'indépendance des variables.

On admettra la réciproque. $\square$

Proposition 5.2.4 (Formule de transfert) Soit Φ une fonction continue de $\mathbb{R}^2$ dans $\mathbb{R}$ et (X, Y) un couple de v.a.r ayant pour densité μ . On a

$$E(\Phi(X, Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Phi(x, y) \mu(x, y) dx dy.$$

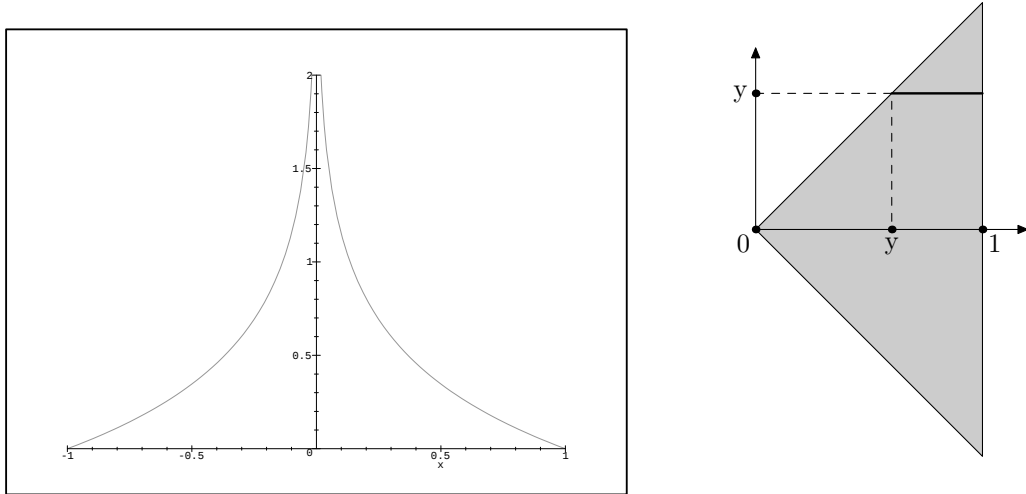
En particulier $E(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy \mu(x, y) dx dy.$

DÉMONSTRATION : Admise. $\square$

5.3 Exemples

Exercice 20. On pose $\mu(x, y) = K e^{-x^2 + xy - y^2}$.

- 1) Déterminer K pour que μ soit une densité de probabilité d'un couple de v.a.r (X, Y) . On pourra remarquer que $x^2 - xy + y^2 = (x - \frac{1}{2}y)^2 + \frac{3}{4}y^2$ et faire le changement de variable $u = x - \frac{1}{2}y$ et $v = y$.
- 2) Déterminer les lois marginales de X et Y .

FIG. 5.1 – Densité de Y et domaine T .

Exemple : On tire au hasard (*i.e* selon la loi uniforme) un réel X au hasard dans $]0; 1[$, puis un réel Y au hasard (*i.e* selon la loi uniforme) dans l'intervalle $[-X; X]$.

- 1) Calculer la densité de X .
- 2) Calculer la densité du couple (X, Y)
- 3) En déduire la densité de Y et calculer $E(Y)$.

DÉMONSTRATION : Cherchons la densité de Y sachant $X = x$: c'est la densité d'une loi uniforme sur l'intervalle $[-x; x]$; on a donc

$$\begin{cases} f(x) = 1 & \text{si } x \in [0; 1] \\ f(x) = 0 & \text{sinon} \end{cases} \quad \begin{cases} g_x(y) = \frac{1}{2x} & \text{si } y \in [-x; x] \\ g_x(y) = 0 & \text{sinon.} \end{cases}$$

La densité μ du couple (X, Y) est égale au produit $f g_x$: elle est définie sur le triangle défini par $T = \{(x, y) \mid 0 < x < 1, -x \leq y \leq x\}$ par $\mu(x, y) = \frac{1}{2x}$ et $\mu(x, y) = 0$ en dehors de T .

On peut vérifier que μ est bien une densité de probabilité :

$$\iint_T \mu(x, y) dx dy = \int_0^1 \left(\int_{-x}^x \frac{1}{2x} dy \right) dx = \int_0^1 1 dx = 1.$$

La densité de Y est donnée par

$$g(y) = \int_{-\infty}^{\infty} \mu(x, y) dx = \int_{|y|}^1 \frac{1}{2x} dx = \left[\frac{1}{2} \ln |x| \right]_{|y|}^1 = -\frac{1}{2} \ln |y| \quad \text{pour } y \in [-1, 1].$$

Remarquez que la densité devient infinie pour $y = 0$ et que le signe « moins » devant le logarithme est bien justifié. On peut vérifier que g est une densité de probabilité

$$\int_{-1}^1 g(y) dy = \int_{-1}^1 -\frac{1}{2} \ln |y| dy = -\frac{1}{2} \left[y \ln |y| - y \right]_{-1}^1 = 1.$$

Il reste à calculer l'espérance de Y , $E(Y) = \int_{-1}^1 -\frac{1}{2} y \ln |y| dy$. Une intégration par parties montre que $E(Y) = 0$. □

Exercice 21. X suit une loi uniforme sur $[0; 1]$ et Y suit une loi uniforme sur $[0; 2]$. Sachant que X et Y sont indépendantes (i.e. le couple (X, Y) suit la loi uniforme), déterminer la probabilité de l'évènement $\{X + Y < 2\}$; généraliser à un évènement $\{X + Y < a\}$. En déduire la densité de $X + Y$.

Théorème 5 Soient X et Y deux v.a.r indépendantes ayant pour densités respectives f et g . La densité de la v.a.r $X + Y$ est donnée par

$$h(u) = \int_{-\infty}^{\infty} f(u - y) g(y) dy$$

La fonction h est appelée produit de convolution de f et g .

DÉMONSTRATION : Notons H la fonction de répartition de $X + Y$. Soit $z \in \mathbb{R}$; $H(z)$ est la probabilité de l'évènement $(X + Y \leq z)$, c'est-à-dire la probabilité que le couple (X, Y) appartienne au demi-plan $D = \{x + y \leq z\}$ et d'après la Proposition 5.2.1 vaut

$$H(z) = \iint_D \mu(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} \mu(x, y) dx dy.$$

Un calcul simple, utilisant le changement de variable $u = x + y$ montre que

$$H(z) = \int_{-\infty}^z \int_{-\infty}^{\infty} \mu(u - y, y) dy du$$

Par définition de la densité, on a $H(z) = \int_{-\infty}^z h(u) du$ d'où par identification

$$h(u) = \int_{-\infty}^{\infty} \mu(u - y, y) dy.$$

La formule annoncée découle de $\mu(x, y) = f(x)g(y)$ d'après l'indépendance de X et Y . □

Exercice 22. Montrer, à l'aide d'un changement de variables, que l'on a aussi

$$h(u) = \int_{-\infty}^{\infty} f(y) g(u - y) dy.$$

Exercice 23. X suit une loi uniforme sur $[0; 2]$ et Y , indépendante de X , suit une loi uniforme sur $[0; 1]$. Déterminer la densité de $X + Y$ et vérifier que $E(X + Y) = 1,5$. Comparer avec l'exercice précédent.

Compte tenu de l'expression de $g(y)$, la formule précédente donne $h(z) = \int_0^1 f(z - y) dy$. Comme f est une fonction nulle en dehors de $[0; 2]$ toute la difficulté vient de savoir comment est placé $z - y$ par rapport à 0 et 2. Faisons le changement de variable $u = z - y$ on alors

$$h(z) = \int_z^{z-1} f(u) (-du) = \int_{z-1}^z f(u) du.$$

- pour $z \leq 0$ et pour $z > 3$, l'intégrale est nulle car l'intervalle $[z - 1; z]$ ne rencontre pas $[0; 2]$.
- pour $0 < z < 1$, l'intersection $[z - 1; z] \cap [0; 2]$ est égale à $[0; z]$ et $h(z) = \int_0^z \frac{1}{2} du = z/2$.
- pour $1 < z < 2$, l'intersection $[z - 1; z] \cap [0; 2]$ est égale à $[z - 1, z]$ et $h(z) = \int_{z-1}^z \frac{1}{2} du = 1/2$.
- pour $2 < z < 3$, l'intersection $[z - 1; z] \cap [0; 2]$ est égale à $[z - 1, 2]$ et $h(z) = \int_{z-1}^2 \frac{1}{2} du = (3 - z)/2$.

Proposition 5.3.1 *Si X suit $\mathcal{N}(m, \sigma)$ et Y , indépendante de X , suit $\mathcal{N}(m', \sigma')$ alors $X + Y$ suit une loi normale.*

DÉMONSTRATION : On va faire la démonstration dans le cas où X et Y suivent $\mathcal{N}(0, 1)$, le cas général se traitant de la même manière par changement de variable. La densité de $X + Y$ est donnée par

$$h(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{(x-y)^2}{2}} e^{-\frac{y^2}{2}} dy = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{(x-y)^2}{2} - \frac{y^2}{2}} dy.$$

En remarquant que $-\frac{(x-y)^2}{2} - \frac{y^2}{2} = -y^2 + xy - \frac{x^2}{2} = -(y - \frac{x}{2})^2 - \frac{x^2}{4}$ et à l'aide du changement de variable $u = y - \frac{x}{2}$ on obtient

$$\begin{aligned} h(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-u^2 - \frac{x^2}{4}} du = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{x^2}{4}} e^{-u^2} du \\ &= \frac{1}{2\pi} e^{-\frac{x^2}{4}} \int_{-\infty}^{\infty} e^{-u^2} du \\ &= \frac{1}{2\sqrt{\pi}} e^{-\frac{x^2}{4}}, \end{aligned}$$

qui est la densité d'une loi $\mathcal{N}(0, \sqrt{2})$. □

Exercice 24. En s'inspirant de la démonstration du théorème ci-dessus, donner une formule permettant de calculer la loi du quotient de deux v.a.r X et Y .

En déduire que si X et Y suivent $\mathcal{N}(0, 1)$, le quotient suit une loi de Cauchy.

Chapitre 6

Convergences

En probabilités il existe plusieurs types de convergence, chacune adaptée à un certain type de résultats. On peut avoir une convergence ponctuelle (p.s), ou en terme d'évènements (en probabilité), ou en terme de densité (en loi).

Dans toute la suite $(X_n)_{n \in \mathbb{N}}$ désigne une suite de v.a.r définie sur Ω .

6.1 Définitions

6.1.1 Convergence presque sûre

On dit que $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement vers X , si $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ pour presque tout $\omega \in \Omega$, c'est-à-dire pour tout ω sauf pour un ensemble de probabilité nulle. Notation : $X_n \longrightarrow X$ p.s .

6.1.2 Convergence en probabilité

On dit que $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X si

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

Cette convergence s'exprime en terme d'évènements et nécessite de connaître X et les X_n , et pas seulement leurs lois.

6.1.3 Convergence en loi

On note respectivement F_n et F les fonctions de répartition de X_n et X . On dit que $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad \text{en tout point } x \text{ où } F \text{ est continue.}$$

Remarque : La convergence en probabilité entraîne la convergence en loi. La réciproque est fausse.

Proposition 6.1.1 (Définition équivalente) $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si et seulement si, pour tout couple (a, b) , $a < b$ avec F continue en a et b , on a

$$\lim_{n \rightarrow \infty} p(\{a < X_n \leq b\}) = p(\{a < X \leq b\}).$$

DÉMONSTRATION : Admise ; elle repose sur le fait que $p(\{a < X_n \leq b\}) = F_n(b) - F_n(a)$. $\square$

Remarque : Cette définition s'applique aussi bien à des variables discrètes qu'à des variables continues. Dans le cas de variables discrètes X_n et X à valeurs dans le même ensemble $\{x_1, \dots, x_k, \dots\}$, la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si et seulement si

$$\forall k, \quad \lim_{n \rightarrow \infty} p(X_n = x_k) = p(X = x_k).$$

Proposition 6.1.2 *On suppose que X_n suit une loi binomiale $\mathcal{B}(n, p_n)$ avec $\lim_{n \rightarrow \infty} n.p_n = \lambda$ fixé ; alors X_n converge en loi vers une loi de Poisson de paramètre λ .*

En pratique, on confondra X_n et sa limite pour $n \geq 30$ et $n.p_n < 5$.

DÉMONSTRATION : On part de $p(X_n = k) = C_n^k p_n^k (1 - p_n)^{n-k}$. On peut alors écrire $(1 - p_n)^{n-k} = e^{(n-k) \ln(1-p_n)}$ et approximer p_n par λ/n et $n(n-1)(n-2)\dots(n-k+1)$ par n^k lorsque $n \rightarrow \infty$, k étant fixé. On obtient alors

$$\begin{aligned} p(X_n = k) &= \frac{n(n-1)\dots(n-k+1)}{k!} p_n^k e^{(n-k) \ln(1-p_n)} \\ &\approx \frac{n^k}{k!} \left(\frac{\lambda}{n}\right)^k e^{(n-k) \ln(1-p_n)} \\ &\approx \frac{\lambda^k}{k!} e^{(n-k) \ln(1-p_n)}. \end{aligned}$$

Enfin, comme p_n et k/n tendent vers 0, on a $\ln(1 - p_n) \approx -p_n$ et

$$\lim_{n \rightarrow \infty} (n - k) \ln(1 - p_n) = \lim_{n \rightarrow \infty} -n(1 - \frac{k}{n})p_n = - \lim_{n \rightarrow \infty} np_n = -\lambda.$$

$\square$

6.2 Théorèmes limites

6.2.1 Inégalité de Bienaymé-Tchebicheff

Proposition 6.2.1 *Soit X une v.a.r positive ayant une espérance $E(X)$ vérifiant $0 < E(X) < \infty$. Alors, pour tout réel $z > 0$, on a*

$$E(X) \geq z p(X \geq z).$$

DÉMONSTRATION :

- Cas discret. On suppose que X prend les valeurs x_i avec les probabilités p_i ; on a $E(X) = \sum x_i p_i$. Décomposons cette somme en deux : d'un côté les termes où $x_i \geq z$ de l'autre ceux où $x_i < z$. On a

$$\begin{aligned} E(X) &= \sum_{x_i \geq z} x_i p_i + \sum_{x_i < z} x_i p_i \\ &\geq \sum_{x_i \geq z} x_i p_i \quad \text{les } x_i p_i \text{ sont positifs} \\ &\geq z \sum_{x_i \geq z} p_i \quad \text{en minorant } x_i \text{ par } z \\ &\geq z p(X \geq z). \end{aligned}$$

– Cas continu. Par hypothèse, la densité f de X est nulle sur $] -\infty; 0]$ On a alors

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} t f(t) dt = \int_0^{\infty} t f(t) dt \geq \int_z^{\infty} t f(t) dt \\ &\geq z \int_z^{\infty} f(t) dt \geq z p(X \geq z). \end{aligned}$$

En prenant $z = a E(X)$ on obtient $E(X) \geq a E(X) p(X \geq a E(X))$ et en divisant par $E(X)$ on montre :

Corollaire 6.2.2 (Inégalité de Markov) Soit X une v.a.r positive ayant une espérance $E(X)$ vérifiant $0 < E(X) < \infty$; on a, pour tout réel $a > 0$, $p\{X \geq a E(X)\} \leq \frac{1}{a}$.

En appliquant l'inégalité de Markov à la variable $Y = |X - E(X)|^2$ on obtient $p\{Y \geq a E(Y)\} \leq \frac{1}{a}$. Par définition $E(Y) = \text{Var}(X) = \sigma^2$; en posant $a = b^2$ on obtient alors en prenant la racine carrée :

Proposition 6.2.3 (Inégalité de Bienaymé-Tchebicheff)

$$p\{|X - E(X)| \geq b\sigma\} \leq \frac{1}{b^2}$$

Remarque : L'intérêt de cette formule est qu'elle ne nécessite aucune connaissance de la loi de X ; son utilité est surtout théorique : en pratique mieux vaut utiliser les tables de la loi de X si elle est connue. Par exemple, soit X telle que $E(X) = 0$ et $\sigma = 1$; l'inégalité de Bienaymé-Tchebicheff avec $b = 2$ donne $p\{|X| \geq 2\} \leq 0,25$. Si on sait que X suit $\mathcal{N}(0,1)$ les tables donnent $p\{|X| \geq 2\} = 0,0456$. Sans commentaire !

6.2.2 Loi faible des grands nombres

Considérons une suite de v.a.r X_k , $1 \leq k \leq n$, indépendantes de même loi de moyenne m et de variance σ . On considère la moyenne¹

$$M_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

On a $E(M_n) = m$ et, d'après l'indépendance,

$$\text{Var}(M_n) = \frac{1}{n^2} \sum \text{Var}(X_k) = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}.$$

En appliquant l'inégalité de Bienaymé-Tchebicheff on obtient

Proposition 6.2.4 On a : $p\{|M_n - m| \geq \frac{b\sigma}{\sqrt{n}}\} \leq \frac{1}{b^2}$

Si on se donne $\varepsilon > 0$, en posant $b = \frac{\varepsilon\sqrt{n}}{\sigma}$ on obtient

$$p\{|M_n - m| \geq \varepsilon\} \leq \frac{1}{n} \frac{\sigma^2}{\varepsilon^2},$$

ce qui démontre la proposition suivante :

¹En statistiques ce serait la moyenne d'un échantillon.

Proposition 6.2.5 (Loi faible des grands nombres) M_n converge en probabilité vers m .

Ce résultat est d'un grand intérêt théorique : il permet de justifier l'idée intuitive qu'une probabilité correspond à une fréquence, ou une moyenne : si on répète un grand nombre de fois une expérience d'espérance m , la moyenne des valeurs observées sera approximativement égale à m .

Exemple : Supposons que les X_n suivent tous une loi de Bernoulli de paramètre p , $0 < p < 1$. On a alors $E(X_n) = p$ et $\text{Var}(X_n) = p(1-p) \leq \frac{1}{4}$ ce qui conduit à

$$p\{|M_n - p| \geq \varepsilon\} \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

Autrement dit, la fréquence des succès tends vers la probabilité p du succès d'une épreuve élémentaire, et ceci indépendamment de p .

Remarque : On peut montrer un résultat plus précis, à savoir que M_n converge vers m presque sûrement. Cette dernière affirmation est appelée *Loi forte des grands nombres*.

6.2.3 Théorème central limite

Le premier théorème limite, historiquement parlant, est dû à De Moivre et Laplace et c'est ce résultat qui justifie l'expression de la densité de la loi normale.

Proposition 6.2.6 Soient S_n une suite de v.a.r telles que S_n suit une loi binomiale $\mathcal{B}(n, p)$ où p est fixé ($q = 1 - p$). Alors

$$\lim_{n \rightarrow \infty} p\{np + a\sqrt{npq} \leq S_n \leq np + b\sqrt{npq}\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

DÉMONSTRATION : Admise. Pour une démonstration « élémentaire » on pourra consulter l'ouvrage d' E. Lesigne *Pile ou Face Une introduction aux théorèmes limites du Calcul des Probabilités* dans la collection *Opuscules* aux Editions Ellipses (2001). $\square$

La formulation ci-dessus est la formulation originale, mais il est facile de voir qu'elle équivaut à dire que la variable centrée réduite $\frac{S_n - m}{\sigma}$ converge en loi vers $\mathcal{N}(0;1)$.

Théorème 6 (Théorème central-limite de P. Levy) Soit X_n une suite de v.a.r deux à deux indépendantes de même loi, de moyenne m et de variance σ^2 ; on pose $M_n = \frac{1}{n} \sum X_n$; la variable $Z_n = \frac{M_n - m}{\sigma/\sqrt{n}}$ converge en loi vers $\mathcal{N}(0;1)$.

DÉMONSTRATION : Admise. $\square$

Remarque : Ce théorème suppose connus m et σ^2 ce qui est assez contraignant. Par contre, il explique l'omniprésence de la loi normale : si on répète des épreuves indépendantes un grand nombre de fois, la distribution asymptotique sera normale.

Remarque : Si on considère qu'une loi binomiale $\mathcal{B}(n, p)$ est une somme de n lois de Bernoulli de paramètre p , le théorème de de Moivre-Laplace est une conséquence du Théorème central-limite.

Exemple : Une proportion inconnue p d'une certaine population est constituée de fumeurs. On utilise un tirage avec remise de n individus dans cette population pour estimer p . Combien de tirages doit-on effectuer pour connaître p avec une erreur inférieure à 0,005 ?

Aucun tirage non exhaustif ne peut apporter la certitude que la proportion observée p' vérifie $|p - p'| < 0,0005$; par contre on peut rendre cet événement (très) improbable. Cherchons donc n de façon que la probabilité de l'évènement ($|p - p'| > 0,005$) soit supérieure à 0,95. Soit X_i la variable de Bernoulli valant 1 si le i -ème individu tiré est fumeur.

La suite (X_i) est une suite de v.a.r indépendantes de même loi et $S_n = X_1 + \dots + X_n$ est le nombre de fumeurs de l'échantillon ($S_n = np'$). En utilisant le théorème de Moivre-Laplace

$$p(|S_n - np| < 0,005n) = p\left(\frac{|S_n - np|}{\sqrt{npq}} < 0,005 \frac{\sqrt{n}}{\sqrt{pq}}\right).$$

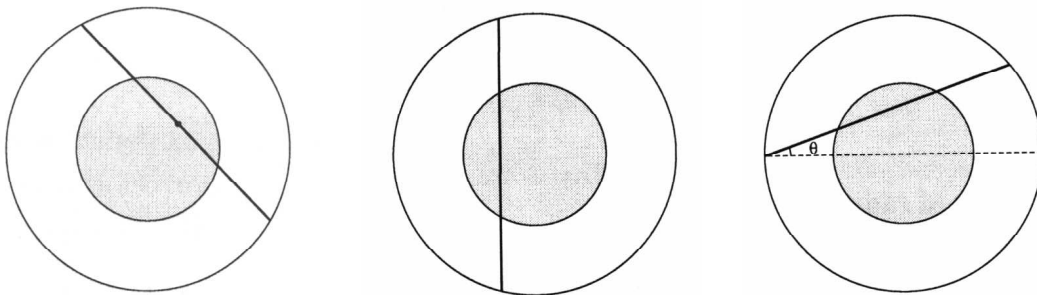
En majorant $\sqrt{pq}$ par 1/2, on est amené à chercher n tel que $p(Z < 0,01\sqrt{n}) > 0.95$ pour Z suivant $\mathcal{N}(0;1)$. Un calcul classique donne $0,01\sqrt{n} > 1,96$, soit $n > 38416$. Inutile de préciser que ce n'est pas une méthode appliquée par les instituts de sondage!

Annexe A

Introduction

L'exemple suivant est une version simplifiée du paradoxe de Bertrand¹ proposée par L.C. Evans dans *An introduction to Stochastic Differential equations* disponible sur le Web à l'adresse suivante : math.berkeley.edu/~evans/SDE.course.pdf

Considérons un cercle $\mathcal{C}$ de centre 0 et de rayon 2 ; on choisit « au hasard » une corde de ce cercle (*i.e* un segment de droite dont les extrémités appartiennent à $\mathcal{C}$). Quelle est la probabilité que cette corde recoupe le cercle $\mathcal{C}'$ de centre 0 et de rayon 1 ?



Solution 1 Une telle corde est entièrement déterminée par son milieu ; la probabilité cherchée est donc égale à la probabilité que le milieu appartienne à $\mathcal{C}'$, c'est-à-dire au rapport $(\text{aire } \mathcal{C}')/(\text{aire } \mathcal{C}) = \frac{1}{4}$.

Solution 2 Quitte à faire une rotation autour de 0, on peut toujours supposer que la corde est verticale. La distance de la corde au point 0 est inférieure à 2 et elle rencontre $\mathcal{C}'$ si cette distance est inférieure à 1 : la probabilité est donc $\frac{1}{2}$.

Solution 3 Quitte à faire une rotation autour de 0, on peut toujours supposer qu'une des extrémités du segment est situé sur le point le plus à gauche de $\mathcal{C}$. L'angle θ entre la corde et l'horizontale appartient à $[-\pi/2; \pi/2]$ et la corde rencontre $\mathcal{C}'$ si θ appartient à $[-\pi/6; \pi/6]$: la probabilité cherchée est donc $\frac{\pi/6}{\pi/2} = \frac{1}{3}$.

Cet exemple montre qu'on doit définir soigneusement ce qu'on entend par « au hasard » dès que l'ensemble des événements (*i.e* des possibilités) est infini non dénombrable ... La bonne manière de procéder est de définir une structure mathématique appelée *espace de probabilité* ou *espace probabilisé*.

Le paradoxe de Bertrand s'explique par le fait que chaque « solution » correspond à un espace probabilisé différent : le choix entre les trois modèles n'est plus du ressort des mathématiques ...

¹Joseph Bertrand (1822-1900) : Mathématicien français né à Paris.

On part d'un ensemble Ω dont certaines parties seront interprétées comme des *événements* i.e des résultats possibles d'une expérience aléatoire; les éléments de Ω , appelés *événements élémentaires*, interviennent peu dans les calculs.

La première difficulté est qu'on ne peut choisir comme ensemble des événements l'ensemble des parties de Ω : il faut se restreindre à des sous-ensembles plus petits nommés *σ -algèbre*; on définit alors une application sur cette *σ -algèbre* qu'on appellera *probabilité*. La deuxième difficulté consiste à définir la notion de *variable aléatoire* : plutôt que de travailler avec des parties (abstraites) de Ω on préfère travailler sur $\mathbb{R}$. Quelles propriétés doit-on imposer à une application de $\Omega \rightarrow \mathbb{R}$ pour pouvoir définir une *probabilité image* ?

Annexe B

Modèle probabiliste

B.1 σ -algèbre

Définition 21 Soit Ω un ensemble ; une partie $\mathcal{A} \subset \mathcal{P}(\Omega)$ est appelée une algèbre de Boole si elle est stable par réunion, intersection et passage au complémentaire.

$$\forall A, B \in \mathcal{A}, \quad A \cup B \in \mathcal{A}, A \cap B \in \mathcal{A}, \complement A \in \mathcal{A}.$$

Si Ω est un ensemble fini, on montre que toute algèbre de Boole sur Ω est de la forme $\mathcal{P}(\Omega')$ pour un certain Ω' .

Exemple : L'ensemble des parties de $\mathbb{R}$ qui peuvent s'écrire comme une réunion *finie* d'intervalles de $\mathbb{R}$ constitue une algèbre de Boole de parties de $\mathbb{R}$.

Nous allons généraliser cette définition pour prendre en compte des ensembles infinis

Définition 22 Soit Ω un ensemble ; une partie $\mathcal{A} \subset \mathcal{P}(\Omega)$ est appelée une σ -algèbre (ou encore tribu) si elle est stable par intersection, passage au complémentaire et réunion dénombrable. Cette dernière propriété signifie que si I est un ensemble dénombrable, et les A_n , pour $n \in I$, sont des parties de $\mathcal{A}$, alors $\bigcup_{n \in I} A_n$ appartient aussi à $\mathcal{A}$.

Exemple : si on considère $I = \mathbb{N}$ et $A_n = [0; 2 - \frac{1}{n+1}] \subset \mathbb{R}$, la réunion des A_n est l'intervalle semi-ouvert $[0; 2[$. On peut remarquer que si $m < n$ alors $A_m \subset A_n$.

Définition 23 Soient $A_{n \in \mathbb{N}}$ des parties de $\mathcal{A}$; on dit que A est limite monotone croissante des A_n si

- $m < n$ entraîne $A_m \subset A_n$,
- $A = \bigcup_{n \in \mathbb{N}} A_n$, i.e A est l'ensemble des éléments a qui appartiennent à au moins un des A_n .

Théorème 7 Soit $\mathcal{A}$ une algèbre de Boole sur Ω ; $\mathcal{A}$ est une σ -algèbre si et seulement si $\mathcal{A}$ est stable par limite monotone croissante.

DÉMONSTRATION : Admise. □

Proposition B.1.1 Pour toute algèbre de Boole $\mathcal{B}$ sur Ω , il existe une plus petite (au sens de l'inclusion) σ -algèbre $\mathcal{A}$ contenant $\mathcal{B}$ qui est appelée σ -algèbre engendrée par $\mathcal{B}$.

DÉMONSTRATION : Admise. Il suffit de compléter $\mathcal{B}$ par limite monotone. □

Définition 24 Soit $\Omega \subset \mathbb{R}$ (ou plus généralement un espace topologique). On appelle boréliens de Ω les éléments de la σ -algèbre engendrée par les (intervalles) ouverts (ou fermés) de Ω .

On démontre que la tribu des boréliens de $\mathbb{R}$ est distincte de l'ensemble des parties de $\mathbb{R}$: il existe des parties de $\mathbb{R}$ qui ne sont pas des boréliens.

B.2 Mesure

Définition 25 Soit $\mathcal{A}$ une σ -algèbre de Ω . Une mesure positive sur $(\Omega, \mathcal{A})$ est une fonction μ définie sur $\mathcal{A}$ à valeurs dans $[0; +\infty]$ qui vérifie la propriété de σ -additivité :

Si les $A_n \in \mathcal{A}$ sont disjoints avec $A = \bigcup_{n \in \mathbb{N}} A_n$ alors $\mu(A) = \sum_n \mu(A_n)$.

Remarque : dans la définition précédente, la série $\sum_n \mu(A_n)$ peut être divergente (la somme vaut alors $+\infty$), d'où l'intérêt que μ puisse prendre la valeur $+\infty$.

- S'il existe une partie $A \in \mathcal{A}$ telle que $\mu(A) < +\infty$, alors $\mu(\emptyset) = 0$.
- Si $A \subset B$, alors $\mu(A) \leq \mu(B)$.
- Si A est limite monotone des $A_n \in \mathcal{A}$, alors $\mu(A) = \lim_n \mu(A_n)$.

Définition 26 Une probabilité sur $(\Omega, \mathcal{A})$ est une mesure positive p telle que $p(\Omega) = 1$.

Un peu de terminologie

- Un espace *mesurable* est le couple formé d'un ensemble Ω et d'une σ -algèbre de Ω .
- Un espace *mesuré* est le triplet formé d'un ensemble Ω , d'une σ -algèbre $\mathcal{A}$ de Ω et d'une mesure μ sur $\mathcal{A}$.
- Un espace *probabilisé* est le triplet formé d'un ensemble Ω , d'une σ -algèbre $\mathcal{A}$ de Ω et d'une probabilité p sur $(\Omega, \mathcal{A})$.

Théorème 8 Si p est une probabilité sur $(\Omega, \mathcal{B})$ où $\mathcal{B}$ est une algèbre de Boole sur Ω et $\mathcal{A}$ la σ -algèbre engendrée par $\mathcal{B}$, alors il existe une probabilité unique sur $\mathcal{A}$ prolongeant p .

DÉMONSTRATION : Admise. □

Ce théorème est important, car il montre qu'on peut définir une probabilité sur $\mathcal{A}$ en la définissant sur un ensemble plus simple $\mathcal{B}$. Par exemple, pour définir une probabilité sur les boréliens de $\Omega = [0; 1] \subset \mathbb{R}$, il suffit de la définir sur les intervalles ouverts de Ω .

Définition 27 Il existe une mesure μ , dite mesure de Lebesgue, définie sur la tribu borélienne de $\mathbb{R}$ telle que

$$\mu([a; b]) = b - a.$$

Cette mesure est aussi notée dx car c'est la mesure qui intervient lorsqu'on parle d'intégration des fonctions réelles.

Définition 28 Soit $(\Omega, \mathcal{A}, p)$ un espace probabilisé et $A \in \mathcal{A}$ tel que $p(A) \neq 0$. L'application $\mathcal{A} \rightarrow [0; 1]$ qui à $B \in \mathcal{A}$ associe $\frac{p(A \cap B)}{p(A)}$ est une probabilité, appelée probabilité conditionnelle. Traditionnellement la probabilité conditionnelle est notée $p(B/A)$.

B.3 Mesurabilité

Définition 29 Soient $(\Omega, \mathcal{A})$ et $(\Omega', \mathcal{A}')$ deux espaces mesurables. Une application $f : \Omega \rightarrow \Omega'$ est dite mesurable si

$$\forall A' \in \mathcal{A}', f^{-1}(A') \in \mathcal{A},$$

où $f^{-1}(A')$ désigne l'image réciproque de A' , c'est-à-dire $\{a \in \Omega \mid f(a) \in A'\}$.

Exercice 25. Soit $(\Omega, \mathcal{A})$ un espace mesurable et f une fonction mesurable à valeurs dans $(\mathbb{R}, \mathcal{B})$.

1. Que dire de f si $\mathcal{A}$ est la tribu grossière $\{\emptyset, \Omega\}$?
2. Que peut-on dire si $\mathcal{A} = \mathcal{P}(\Omega)$?
3. Soit $(A_n)_{n \in I}$ une partition de Ω . Quelle est la tribu engendrée par les A_i lorsque $I = 1, 2$? lorsque $I = 1, 2, 3$? Dans chaque cas préciser quelles sont les fonctions mesurables.

Proposition B.3.1 Soit $f : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ où $\mathcal{A}'$ est la σ -algèbre engendrée par $\mathcal{B}'$; f est mesurable si et seulement si

$$\forall B' \in \mathcal{B}', f^{-1}(B') \in \mathcal{A}.$$

DÉMONSTRATION : Admise □

Par exemple si $(\Omega', \mathcal{A}')$ est $(\mathbb{R}, \mathcal{B})$, pour montrer que f est mesurable, il suffit de vérifier que l'image réciproque d'un intervalle appartient à $\mathcal{A}$.

Proposition B.3.2 Le composé de deux applications mesurables est mesurable.

DÉMONSTRATION : Soit $f : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ et $g : (\Omega', \mathcal{A}') \rightarrow (\Omega'', \mathcal{A}'')$; considérons $A'' \subset \mathcal{A}''$.

$$(g \circ f)^{-1}(A'') = f^{-1} \circ g^{-1}(A'') = f^{-1}(g^{-1}(A'')).$$

Comme g est mesurable, $g^{-1}(A'') = A'$ est un élément de $\mathcal{A}'$ et de même $f^{-1}(A')$ appartient à $\mathcal{A}$. □

Proposition B.3.3 Soit $f : (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$ une fonction.

- Si f est monotone, elle est mesurable.
- Si f est continue, elle est mesurable.
- Si f est dérivable sur $\mathbb{R}$, la dérivée f' est mesurable.

DÉMONSTRATION : Admise. □

En fait, il est assez difficile de construire une fonction de $\mathbb{R}$ dans $\mathbb{R}$ qui ne soit pas mesurable par rapport à la tribu des boréliens.

B.4 Mesure produit

Les définitions suivantes sont données dans le cas le plus général d'un produit cartésien de deux ensembles Ω_1 et Ω_2 , mais en pratique on aura toujours $\Omega_1 = \Omega_2$; elles se généralisent de façon évidente au produit cartésien de n ensembles.

Soient $(\Omega_1, \mathcal{A}_1)$ et $(\Omega_2, \mathcal{A}_2)$ deux espaces mesurables, la σ -algèbre engendrée par les produits $A_1 \times A_2$ pour $A_1 \in \mathcal{A}_1$ et $A_2 \in \mathcal{A}_2$ est appelée σ -algèbre produit de $\mathcal{A}_1$ et $\mathcal{A}_2$; elle est définie sur $\Omega = \Omega_1 \times \Omega_2$.

Proposition B.4.1 La σ -algèbre produit est la plus petite tribu rendant mesurable les projections $p_1 : \Omega \rightarrow \Omega_1$ et $p_2 : \Omega \rightarrow \Omega_2$

DÉMONSTRATION : Admise. □

Corollaire B.4.2 *La tribu borélienne de $\mathbb{R}^2$ est la σ -algèbre engendrée par les pavés de la forme $]a; b[\times]c; d[$ ou plus généralement par les produits d'intervalles.*

Proposition B.4.3 *Soit $\Omega_1 \times \Omega_2$ muni de la tribu produit de $\mathcal{A}_1$ et $\mathcal{A}_2$; $f : \Omega_1 \times \Omega_2 \rightarrow (\Omega, \mathcal{A})$ est mesurable (pour la tribu produit) si et seulement si toutes les applications partielles f_{ω_1} et f_{ω_2} sont mesurables.*

DÉMONSTRATION : Admise. □

Proposition B.4.4 *Soit $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^2, \mathcal{B})$; f est mesurable si et seulement si chacune des fonctions coordonnées est une fonction mesurable de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B})$.*

DÉMONSTRATION : Supposons f_1 et f_2 mesurables et soit $P =]a; b[\times]c; d[$. On a

$$\begin{aligned} f^{-1}(P) &= \{a \mid f_1(a) \in]a; b[\text{ et } f_2(a) \in]c; d[\} = \{a \mid f_1(a) \in]a; b[\} \cap \{a \mid f_2(a) \in]c; d[\} \\ &= f_1^{-1}(]a; b[) \cap f_2^{-1}(]c; d[). \end{aligned}$$

Par hypothèse $f_1^{-1}(]a; b[)$ et $f_2^{-1}(]c; d[)$ appartiennent à $\mathcal{A}$, donc leur intersection aussi : f est donc mesurable. Réciproque admise. □

A titre d'exemple,

Proposition B.4.5 *Soient f et g deux fonctions mesurables $(\Omega, \mathcal{A}) \rightarrow \mathbb{R}$, k un réel fixé. Alors les fonctions kf , $f + g$, fg sont mesurables.*

DÉMONSTRATION : Par exemple, la fonction $x \mapsto (f(x), g(x))$ de $(\Omega, \mathcal{A})$ dans $\mathbb{R}^2$ est mesurable. Les fonctions de $\mathbb{R}^2$ dans $\mathbb{R} : (y, z) \mapsto y + z$ (resp. $(y, z) \mapsto yz$) est continue, donc mesurable. Le composé de ces fonctions est donc mesurable. □

Théorème 9 *Avec les notations précédentes, supposons que p_1 (resp. p_2) soit une probabilité sur Ω_1 (resp. Ω_2); il existe une unique mesure de probabilité p , notée $p_1 \otimes p_2$, définie sur la tribu produit et telle que*

$$p(A_1 \times A_2) = p(A_1) \times p(A_2) \quad \forall A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2.$$

DÉMONSTRATION : Admise : la démonstration de ce résultat nécessite de connaître la notion d'intégrale d'une fonction par rapport à une mesure, que nous n'aborderons pas ici. □

Corollaire B.4.6 *Il existe une mesure sur les boréliens de $\mathbb{R}^2$ telle que la mesure d'un pavé $]a; b[\times]c; d[$ soit égale à $(b - a)(d - c)$; cette mesure est appelée mesure de Lebesgue, généralement notée $dx dy$.*

B.5 Variable aléatoire

Définition 30 *Soit $(\Omega, \mathcal{A}, p)$ un espace probabilisé et $(\Omega', \mathcal{A}')$ un espace mesurable; une variable aléatoire X sur $(\Omega, \mathcal{A}, p)$ à valeurs dans Ω' est une application mesurable $X : \Omega \rightarrow \Omega'$.*

Théorème 10 Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, p)$ à valeurs dans $(\Omega', \mathcal{A}')$; la fonction p^X définie sur $\mathcal{A}'$ par $p^X(A') = p(X^{-1}(A'))$ est une mesure de probabilité, appelée probabilité image ou distribution de la variable aléatoire. Pour $A' \in \mathcal{A}'$ on a

$$p^X(A) = p(X^{-1}(A)) = p(\{\omega \in \Omega \mid X(\omega) \in A\}).$$

DÉMONSTRATION : Admise. □

Dans toutes les applications, si Ω' n'est pas dénombrable, ce sera un intervalle de $\mathbb{R}$ muni de la tribu borélienne.

Remarque : Dans le cas d'une variable X à valeurs réelles, on notera $p(X = y)$ la quantité $p^X(\{y\}) = p(\{\omega \in \Omega \mid X(\omega) = y\})$ et $p(X < y)$ la quantité $p^X(]-\infty; y[) = p(\{\omega \in \Omega \mid X(\omega) < y\})$.

Remarque : Si X est une v.a.r l'ensemble $\mathcal{U}(X) = \{X^{-1}(B) \mid B \in \mathcal{B}\}$ est une σ -algèbre, appelée σ -algèbre engendrée par X . C'est la plus petite σ -algèbre sur Ω par rapport à laquelle X est mesurable.

Il est important de comprendre qu'en termes de probabilités, la σ -algèbre engendrée par X peut être interprétée comme représentant « toute l'information contenue dans X ». En particulier si Y est une fonction de X , de la forme $\phi(X)$ où ϕ est une fonction numérique, alors Y est mesurable par rapport à $\mathcal{U}(X)$. Réciproquement si Y est mesurable par rapport à $\mathcal{U}(X)$, alors il existe une fonction ϕ telle que $Y = \phi(X)$; bien évidemment ce résultat est théorique et ne fournit (hélas!) aucune construction pratique de la fonction ϕ .

Dans le cas où $\Omega' = \mathbb{R}^2$, une application mesurable est appelée un *vecteur aléatoire*; la distribution est appelée *loi conjointe* de X ; On appelle *loi marginales* de X les probabilités images de P^X par les projections p_1 et p_2 de $\mathbb{R}^2$ dans $\mathbb{R}$.

Table des matières

1	Introduction et rappels	2
1.1	Rappels et définitions	2
1.1.1	Ensemble des parties	2
1.1.2	Fonctions	3
1.1.3	Produit cartésien	3
2	Modèle probabiliste	5
3	Variables discrètes	7
3.1	Espérance	7
3.2	Couple de v.a.r	8
3.2.1	Indépendance	8
3.2.2	Covariance	9
3.3	Lois discrètes usuelles	9
3.4	Introduction aux v.a.r continues.	11
3.4.1	Fonction de répartition	11
4	V.a.r continues	13
4.1	Densité	13
4.1.1	Introduction	13
4.1.2	Définition	14
4.2	Exemples	15
4.2.1	La loi uniforme	15
4.2.2	Loi de Laplace	16
4.2.3	Loi exponentielle	16
4.2.4	Loi normale	16
4.3	Moments	17
4.3.1	Rappels d'intégration	17
4.3.2	Espérance et variance	18
4.3.3	Exemples	18
4.4	Loi dérivées de la loi normale	19
4.5	Compléments	22
5	Couple de v.a.r	24
5.1	Intégrale double	24
5.1.1	Définition	24
5.1.2	Changement de variables	25
5.2	Densité	26
5.3	Exemples	27

6	Convergences	31
6.1	Définitions	31
6.1.1	Convergence presque sûre	31
6.1.2	Convergence en probabilité	31
6.1.3	Convergence en loi	31
6.2	Théorèmes limites	32
6.2.1	Inégalité de Bienaymé-Tchebicheff	32
6.2.2	Loi faible des grands nombres	33
6.2.3	Théorème central limite	34
A	Introduction	36
B	Modèle probabiliste	38
B.1	σ -algèbre	38
B.2	Mesure	39
B.3	Mesurabilité	40
B.4	Mesure produit	40
B.5	Variable aléatoire	41