Chapitre 1 : Probabilités

N. Biheng

1 Généralités et premières notions

Rappelons qu'une expérience aléatoire [1] est une expérience dont on ne peut prévoir l'issue à l'avance mais dont on connaît toutes les issues possibles. Le mot aléa vient du latin *alea* qui est un jeu de dés. Traditionnellement différents types d'incertitude ou de hasard sont distingués. Plus précisément, les situations probabilisables sont distinguées de celle qui ne le sont pas.
Voici quelques exemples d'expériences aléatoires :

- lancer d'une pièce,
- expérience de Bernoulli à deux issues : « Succès » et « Echec » aussi appelée *tirage de Bernoulli*,
- lancer d'un dé,
- choix d'une personne dans une population.

L'ensemble des issues sera appelé l'*univers* et noté Ω. Par définition, un *événement* sera un sous-ensemble de Ω. Nous dirons qu'une issue *a réalise* l'événement A lorsque $a \in A$.
Avant d'aborder les variables aléatoires, revoyons le conditionnement.
Soient A et B deux événements tels que $P(A) \neq 0$ et $P(B) \neq 0$.
La probabilité conditionnelle de l'événement B sachant que l'événement A est réalisé est définie par :

$$P_A(B) := \frac{P(A \cap B)}{P(A)}$$

Nous en déduirons la formule de Bayes [2] :

$$\boxed{P_B(A) = \frac{P_A(B) \times P(A)}{P_A(B) \times P(A) + P_{\bar{A}}(B) \times P(\bar{A})}}$$

1. Pour des développements non techniques sur les probabilités et leurs applications, le lecteur pourra consulter [1].
2. sa démonstration qui procède de la définition des probabilités conditionnelles est laissée en exercice.

Nous considérerons, durant ce cours, des *variables aléatoires* c'est à dire des fonctions définies sur l'univers et à valeurs dans $\mathbb{R}$ et des *vecteurs aléatoires* qui sont des fonctions définies sur l'univers Ω et à valeurs dans $\mathbb{R}^d$ pour un certain entier naturel $d > 1$.
Lorsque l'ensemble des valeurs prises par une variable aléatoire X est discret, nous dirons qu'elle est discrète et appellerons cet ensemble son *support*[3].
La loi d'une variable aléatoire discrète est donnée par la donnée des réels $\mathbb{P}(X = x_i)$ pour $x_i \in \mathcal{S}$ où $\mathcal{S}$ désigne son support. Il s'agit d'une loi définie sur $\mathcal{S}$ qui est un sous-ensemble de $\mathbb{R}$.
Dans ce cas, pour tout événement $A \subset \mathcal{S}$, $\mathbb{P}(A) = \sum_{x \in A} p_x$.
L'*espérance* est définie par $E(X) := \sum_{x \in \mathcal{S}} x p_x$ et correspond à la « valeur moyenne ». Elle est parfois notée $\bar{x}$.
La *variance* est définie par $V(X) := \sum_{x \in \mathcal{S}} p_x (x - \bar{x})^2$ et est une moyenne pondérée des écarts quadratiques à la moyenne. Elle est parfois notée σ^2.
Nous définirons également l'*écart-type* par $\sigma(X) := \sqrt{V(X)}$.
La variance et l'écart-type sont des paramètres de dispersion.

Proposition 1

1. $V(X) = E(X^2) - E(X)^2$ (formule de Koenig-Huyghens),

2. Si X et Y sont deux variables aléatoires

$$V(X + Y) = V(X) + 2\mathrm{Cov}(X, Y) + V(Y)$$

   où $\mathrm{Cov}(X, Y) := E((X - E(X))(Y - E(Y)) = E(XY) - E(X)E(Y)$.

Cette proposition est vraie dans le cas discret comme dans le cas continu.

Preuve

1. Par définition,

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

   D'où, en développant :

$$V(X) = E(X^2 - 2XE(X) + E(X)^2)$$

3. La définition du support fait intervenir des considérations topologiques et issues de la théorie de la mesure dans le cas continu donc nous ne parlerons pas de support dans ce cas pour éviter toute confusion.

Notons que $E(X)$ et $E(X)^2$ sont des constantes, nous obtenons :

$$V(X) = E(X^2) - 2E(X)^2 + E(X)^2 = E(X^2) - E(X)^2$$

d'où le résultat.

2. Par définition,

$$V(X+Y) = E((X+Y-E(X+Y))^2)$$

d'où

$$V(X+Y) = E((X+Y)^2) - E(X+Y)^2 = E((X+Y)^2) - (E(X)+E(Y))^2$$

d'après le point précédent.
Par conséquent, en développant :

$$V(X+Y) = E(X^2)+2E(XY)+E(Y^2)-(E(X)^2+2E(X)E(Y)+E(Y)^2)$$

et

$$V(X+Y) = E(X^2)-E(X)^2+2(E(XY)-E(X)E(Y))+E(Y^2)-E(Y)^2$$

Ainsi, nous obtenons :

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

Avant de présenter les différentes lois discrètes, définissons le moment d'ordre k et le moment centré d'ordre k pour $k \in \mathbb{N}^\star$.
Le moment d'ordre k est défini par :

$$m_k = E(X^k)$$

Le moment centré d'ordre k est défini par :

$$m_k = E\left((X - E(X))^k\right)$$

Loi uniforme discrète sur $\{1, \ldots, n\}$ pour $n \geq 1$

Elle correspond à l'équiprobabilité sur $\{1, \ldots, n\}$.
Pour tout $k \in \{1, \ldots, n\}$,

$$\boxed{\mathbb{P}(X = k) = \frac{1}{n}}$$

Proposition 2
On a :
$E(X) = \dfrac{n+1}{2}$ et $V(X) = \dfrac{n^2-1}{12}$

Preuve
Calculons tout d'abord $E(X)$.
Par définition, $E(X) = \displaystyle\sum_{k=1}^{n} k\mathbb{P}(X=k) = \frac{1}{n}\sum_{k=1}^{n} k = \frac{n+1}{2}$
en se rappelant que $\displaystyle\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$.
Calculons maintenant $E(X^2)$.
$E(X^2) = \displaystyle\sum_{k=1}^{n} k^2\mathbb{P}(X=k) = \frac{1}{n}\sum_{k=1}^{n} k^2 = \frac{(n+1)(2n+1)}{6}$
en se rappelant que $\displaystyle\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$.
par conséquent,

$$V(X) = E(X^2) - E(X)^2 = \frac{n(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}.$$

Loi de Bernoulli de paramètre p

Cette loi est notée $\mathcal{B}(p)$ et son support est $\{0;1\}$. La probabilité pour une variable X suivant cette loi de prendre la valeur 1 est notée p donc $P(X=0) = 1-p$.

Proposition 3
On a :
$E(X) = p$ et $V(X) = p(1-p)$.

Preuve
Les calculs de l'espérance et de la variance sont particulièrement simples dans ce cas. En effet,

$$E(X) = (1-p)\times 0 + p\times 1 = \boxed{p}$$

et

$$V(X) = (1-p)\times(0-p)^2 + p\times(1-p)^2 = (1-p)p^2 + p(1-p)^2 = p(1-p)(p+1-p)$$

D'où

$$\boxed{V(X) = p(1-p)}$$

Loi binomiale de paramètres n et p

Initialement, introduite par Jacques Bernoulli en 1713 dans son célèbre ouvrage *Ars Conjectandi*, elle suscitera de nombreux travaux tant sur le plan théorique que sur le plan des applications.
Elle est notée $\mathcal{B}(n, p)$ et correspond au nombre de succès après n répétitions d'une expérience de Bernoulli.
Par conséquent, son support est $\{0, \ldots, n\}$.
Nous verrons qu'une variable suivant une telle loi est correspond à la somme de n variables indépendantes suivant une loi $\mathcal{B}(p)$.
Par définition,
$\boxed{\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}}$ pour $k \in \{0, \cdots, n\}$ [4].

Proposition 4
On a :
$E(X) = np$ et $V(X) = np(1-p)$.

Preuve
Calculons $E(X)$ et $V(X)$.
$E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = p \sum_{k=1}^{n} \binom{n}{k} k p^{k-1} (1-p)^{n-k}$.
Donc, en remarquant que, $k \binom{n}{k} = n \dfrac{(n-1)!}{(n-1-(k-1))!(k-1)!} = n \binom{n-1}{k-1}$,
on obtient : $E(X) = np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)}$.
Après un changement de variables, on obtient :
$E(X) = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{(n-1)-j} = (p + (1-p))^{(n-1)} = \boxed{np}$.
Pour calculer la variance, calculons tout d'abord $E(X(X-1))$.
$E(X(X-1)) = \sum_{k=0}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k}$
Ainsi, $E(X(X-1)) = p^2 \sum_{k=2}^{n} \binom{n}{k} k(k-1) p^{k-2} (1-p)^{n-k}$.

4. $\binom{n}{k}$ est appelé un *coefficient binomial*.

Donc, en remarquant que,

$$k(k-1)\binom{n}{k} = n(n-1)\frac{(n-2)!}{((n-2)-(k-2))!(k-2)!} = n(n-1)\binom{n-2}{k-2}$$

Ainsi, on obtient : $E(X(X-1)) = n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2}(1-p)^{(n-2)-(k-2)}$

Après un changement de variables,
$E(X(X-1)) = n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^{k-2}(1-p)^{(n-2)-j}$.
Ainsi, $E(X(X-1)) = n(n-1)p^2(p+(1-p))^{n-2} = \boxed{n(n-1)p^2}$.
Par conséquent, $E(X^2) = E(X(X-1)) + E(X) = \boxed{n(n-1)p^2 + np}$.
En conclusion,
$V(X) = E(X^2) - E(X)^2 = n(n-1)p^2 + np - n^2p^2 = np((n-1)p + 1 - np)$.
En conclusion, $\boxed{V(X) = np(1-p)}$.□

Loi géométrique de paramètre p

Cette loi est notée $\mathcal{G}(p)$ et correspond au nombre d'essais avant le premier succès dans une répétition de tirages indépendants de Bernoulli.
Comme précédemment, p correspond à la probabilité de l'issue « Succès ».
Son support est $\mathbb{N}^\star$. Pour tout entier $k \geq 1$, $\boxed{\mathbb{P}(X = k) = pq^{k-1}}$ où $q = 1-p$.

Proposition 5
On a :
$E(X) = \dfrac{1}{p}$ et $V(X) = \dfrac{q}{p^2}$,

Preuve
Calculons $E(X)$ et $V(X)$.
On a :
$E(X) = \sum_{k\geq 1} kpq^{k-1} = p \sum_{k\geq 1} kq^{k-1}$.

En tant que fonction [5] de $q \in]0;1[$, $\left(\sum_{k\geq 0} q^k\right)' = \sum_{k\geq 1} kq^{k-1}$.

5. Le lecteur pourra consulter [6] pour des rappels sur les séries entières.

Or $\dfrac{1}{1-q} = \sum_{k\geq 0} q^k$ donc $\dfrac{1}{(1-q)^2} = \sum_{k\geq 1} kq^{k-1}$.
Par conséquent,

$$E(X) = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \boxed{\frac{1}{p}}$$

Pour calculer $V(X)$, calculons tout d'abord $E(X(X-1))$.

$$E(X(X-1)) = \sum_{k\geq 2} k(k-1)pq^{k-1} = pq\sum_{k\geq 2} k(k-1)q^{k-2}$$

en remarquant que cette variable aléatoire est nulle en 1.
En tant que fonction de $q \in]0;1[$, $\left(\sum_{k\geq 0} q^k\right)'' = \sum_{k\geq 2} k(k-1)q^{k-2}$.
Donc, comme précédemment,

$$\frac{2}{(1-q)^3} = \sum_{k\geq 2} k(k-1)q^{k-2}$$

Donc

$$E(X(X-1)) = \frac{2pq}{(1-q)^3} = \frac{2pq}{p^3} = \boxed{\frac{2q}{p^2}}$$

Par conséquent,
$E(X^2) = E(X(X-1)) + E(X) = \dfrac{2q}{p^2} + \dfrac{1}{p} = \dfrac{2q+p}{p^2} = \dfrac{2(1-p)+p}{p^2} = \boxed{\dfrac{2-p}{p^2}}$.
En conclusion,

$$V(X) = E(X^2) - E(X)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

Ainsi $V(X) = \boxed{\dfrac{q}{p^2}}$. $\square$

Loi de Poisson de paramètre $\lambda > 0$

La loi de Poisson a été introduite par Siméon-Denis Poisson en 1838 dans son ouvrage *Recherche sur les probabilités des jugements en matière criminelle et en matière civile*[6]. Elle correspond aux événements rares et a

6. Cet ouvrage peut être lu sur *Gallica*, le site de la BNF.

donc été utilisée pour modéliser l'occurence de suicides, d'accidents. Depuis l'étude des processus de Poisson, son champ d'application s'étend notamment aux télécommunications, à la biologie et à la finance.
Elle est notée $\mathcal{P}(\lambda)$ et son support est $\mathbb{N}$.
Pour tout entier $k \geq 0$,
$\boxed{\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}}$

Proposition 6
On a :
$E(X) = \lambda$ et $V(X) = \lambda$.

Preuve
Calculons $E(X)$ et $V(X)$.
$E(X) = \sum_{k\geq 0} ke^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k\geq 1} ke^{-\lambda}\frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda}\sum_{k\geq 1}\frac{\lambda^{k-1}}{(k-1)!}$.
Après un changement de variables, on obtient :
$E(X) = \lambda e^{-\lambda}\sum_{j\geq 0}\frac{\lambda^j}{j!}$.
En conclusion,

$$E(X) = \lambda e^{-\lambda}e^{\lambda} = \boxed{\lambda}$$

Afin de calculer $V(X)$, nous calculerons tout d'abord $E(X(X-1))$.

$$E(X(X-1)) = \sum_{k\geq 0} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k\geq 2} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!}$$

D'où

$$E(X(X-1) = \lambda^2 e^{-\lambda}\sum_{k\geq 2}\frac{\lambda^{k-2}}{(k-2)!}$$

Après un changement de variables, on obtient :
$E(X(X-1)) = \lambda^2 e^{-\lambda}\sum_{j\geq 0}\frac{\lambda^j}{j!}$.
En conclusion,

$$E(X(X-1)) = \lambda^2 e^{-\lambda}e^{\lambda} = \boxed{\lambda^2}$$

Par conséquent,

$$E(X^2) = E(X(X-1)) + E(X) = \boxed{\lambda^2 + \lambda}$$

D'où

$$V(X) = E(X^2) - E(X)^2 = \lambda^2 + \lambda - \lambda^2$$

Ainsi $V(X) = \boxed{\lambda}$.$\square$

Pour les lois continues, le principe est similaire.
Nous supposerons l'existence d'une fonction f définie sur $\mathbb{R}$ telle que, pour tout intervalle $[a;b]$, $\mathbb{P}(X \in [a;b]) = \int_a^b f(x)dx$
Une telle fonction est appelée une **fonction de densité** et doit vérifier deux conditions :

1. $f(x) \geq 0$, pour tout réel x,
2. $\int_{\mathbb{R}} f(x)dx = 1$.

Une loi continue est entièrement caractérisée par sa fonction de répartition F définie sur $\mathbb{R}$ par : $F_X(x) := \mathbb{P}(X \leq x) = \displaystyle\int_{-\infty}^{x} f(t)dt$.

Elle est également caractérisée par sa fonction de survie :
$R_X(x) := \mathbb{P}(X > x) = 1 - F_X(x)$.

Elle est enfin caractérisée par sa fonction caractéristique qui est définie par : $\phi_X(t) = E(e^{itX})$ pour $t \in \mathbb{R}$.
Notons que ces trois fonctions sont définies et caractérisent également les lois dans le cas discret.
Exemple 1
Déterminer la fonction caractéristique de la loi $\mathcal{B}(n,p)$.
Soit X une variable aléatoire suivant une loi binomiale de paramètres n et p.
On a :

$$E(e^{itX}) = \sum_{k=0}^{n} e^{itk} \binom{n}{k} p^k(1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pe^{it})^k(1-p)^{n-k} = (pe^{it}+1-p)^n$$

en reconnaissant le binôme de Newton

Contrairement à l'espérance ou à la variance, la fonction caractéristique est toujours définie.
En effet, $\int_{\mathbb{R}} |e^{itX}|f(x)dx = \int_{\mathbb{R}} f(x)dx = 1 < +\infty$.
Son lien avec les moments est donné par le théorème suivant dont la preuve pourra être trouvée dans [2].

Théorème 1
Soit X une variable aléatoire de fonction caractéristique ϕ. Supposons qu'il existe un entier $n \geq 1$ tel que $E(|X|^n) < +\infty$.
Alors

1. la fonction ϕ est de classe $\mathcal{C}^n$,
2. pour tout $k \in \{1, \dots, n\}$, $\phi^k(0) = i^k E(X^k)$,
3. la fonction ϕ admet le développement de Taylor d'ordre n au voisinage de 0 :
$$\phi(u) = 1 + \sum_{k=0}^{n} \frac{(iu)^k}{k!} E(X^k) + u^n \varepsilon(|u^n|)$$
avec $\lim_{u \longrightarrow 0} \varepsilon(|u^n|) = 0$.

Espérance et variance, quand elles existent, sont respectivement données par :

- $E(X) = \int_{\mathbb{R}} x f(x) dx$
- $V(X) = \int_{\mathbb{R}} (x - E(X))^2 f(x) dx$.

Notons que l'espérance et la variance existent toutes deux dès lors que la condition $E(X^2) < +\infty$ est satisfaite.

Nous avons vu, dans la feuille d'exercices 1, que l'espérance n'existe pas pour $\alpha \in]0; 1]$ dans le cas de la *loi de Pareto*.

Loi uniforme sur l'intervalle $[a; b]$ ($a < b$)

Cette loi est notée $\mathcal{U}([a; b])$. Sa densité f est définie par :
$f(x) = \dfrac{1}{b-a}$ pour $x \in [a; b]$ et $f(x) = 0$ pour $x \notin [a; b]$.

Proposition 7
1. $\mathbb{P}(X \in [c; d]) = \dfrac{d-c}{b-a}$ si $a \leq c \leq d \leq b$ et $a < b$.
2. $E(X) = \dfrac{a+b}{2}$ et $V(X) = \dfrac{(b-a)^2}{12}$,

Preuve
Calculons tout d'abord $P(X \in [c;d])$.
$\mathbb{P}(X \in [c;d]) = \int_c^d f(x)dx = \int_c^d \frac{1}{b-a}dx$.
Par conséquent, $\mathbb{P}(X \in [c;d]) = \left[\frac{1}{b-a}x\right]_c^d = \frac{d}{b-a} - \frac{c}{b-a}$.
En conclusion,
$\boxed{\mathbb{P}(X \in [c;d]) = \frac{d-c}{b-a}}$.

Calculons maintenant $E(X)$.

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_a^b xf(x)dx = \int_a^b x \times \frac{1}{b-a}dx = \left[\frac{1}{b-a} \times \frac{x^2}{2}\right]_a^b.$$

Par conséquent,
$E(X) = \frac{b^2}{2(b-a)} - \frac{a^2}{2(b-a)} = \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)}$.
En conclusion,

$$\boxed{E(X) = \frac{a+b}{2}}$$

Pour obtenir $V(X)$, nous calculerons tout d'abord $E(X^2)$.

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_a^b x^2 f(x)dx = \int_a^b x^2 \times \frac{1}{b-a}dx = \left[\frac{1}{b-a} \times \frac{x^3}{3}\right]_a^b.$$

Par conséquent,
$E(X^2) = \frac{b^3}{3(b-a)} - \frac{a^3}{3(b-a)} = \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)(a^2+ab+b^2)}{3(b-a)}$.

$$\boxed{E(X^2) = \frac{a^2+ab+b^2}{3}}$$

Ainsi,

$$V(X) = E(X^2) - E(X)^2 = \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2$$

D'où

$$V(X) = \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4}$$

Par conséquent,

$$V(X) = \frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2}{12}$$

En conclusion,
$V(X) = \dfrac{b^2 - 2ab - a^2}{12} = \boxed{\dfrac{(b-a)^2}{12}}$. $\square$

Loi exponentielle de paramètre $\lambda > 0$

Cette loi est notée $\mathcal{E}(\lambda)$ et correspond à la durée de vie d'un phénomène sans mémoire. Elle modélise aussi l'arrivée des clients dans une file d'attente. Sa densité est donnée par
$f(x) = \lambda e^{-\lambda x}$ pour $x \geq 0$ et $f(x) = 0$ pour $x < 0$

Proposition 8
.

1. $E(X) = \dfrac{1}{\lambda}$ et $V(X) = \dfrac{1}{\lambda^2}$,
2. $F(x) = 1 - e^{-\lambda x}$ pour $x \geq 0$ et $F(x) = 0$ sinon,
3. $R(x) = e^{-\lambda x}$ pour $x \geq 0$ et $R(x) = 1$ sinon.

La loi exponentielle est *sans mémoire* car [7]

$$\forall s \geq 0, \forall t \geq 0, \mathbb{P}_{\{T>t\}}(T > s + t) = \mathbb{P}(T > s)$$

Preuve
Calculons tout d'abord $E(X)$.

$$E(X) = \int_0^{+\infty} x e^{-\lambda x} dx$$

par définition.
Il s'agit d'une intégrale impropre [8].
Notons que toutes les fonctions qui interviennent dans cette preuve sont de

7. Cette propriété est simple. Sa preuve est laissée en exercice.
8. Le lecteur pourra se reporter à [4] pour un développement sur les intégrales impropres.

classe $\mathcal{C}^\infty$ donc les intégrales de Riemann sont bien définies.
Soit $A > 0$.

$$\int_{-\infty}^{A} xf(x)dx = \int_0^A x(\lambda e^{-\lambda x})dx = \left[x(-e^{-\lambda x})\right]_0^A - \int_0^A 1 \times (-e^{-\lambda x})dx$$

en intégrant par parties.

$$\int_0^A xe^{-\lambda x}dx = Ae^{-\lambda A} + \int_0^A e^{-\lambda x}dx$$

D'où

$$\int_0^A xe^{-\lambda x}dx = A\lambda e^{-\lambda A} + \left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^A$$

Ainsi,

$$\int_0^A xe^{-\lambda x}dx = Ae^{-\lambda A} - \frac{1}{\lambda}e^{-\lambda A} + \frac{1}{\lambda}e^{-\lambda \times 0} = \frac{\lambda A - 1}{\lambda}e^{-\lambda A} + \frac{1}{\lambda}$$

Or, on sait que : $\lim\limits_{x \longrightarrow +\infty} e^{-x} = 0$ et, par croissance comparée, $\lim\limits_{x \longrightarrow +\infty} xe^{-x} = 0$.
Donc,

$$\lim_{A \longrightarrow +\infty} \int_0^A xe^{-\lambda x}dx = \frac{1}{\lambda}$$

Par conséquent, l'intégrale $\int_0^{+\infty} xe^{-\lambda x}dx$ existe et vaut $\frac{1}{\lambda}$.
En conclusion,

$$\boxed{E(X) = \frac{1}{\lambda}}$$

Calculons maintenant $E(X^2)$.
Comme précédemment, il faut traiter une intégrale impropre.
Soit $A > 0$.

$$\int_0^A x^2(\lambda e^{-\lambda x})dx = \left[x^2(-e^{-\lambda x})\right]_0^A - \int_0^A 2x(-e^{-\lambda x})dx$$

en intégrant par parties.
D'où, d'après ce qui précède,

$$\int_0^A x^2(\lambda e^{-\lambda x})dx = (-\lambda A^2 - 2\lambda A)e^{-\lambda A} + \left[\frac{2}{\lambda^2}(-e^{-\lambda x})\right]_0^A$$

Enfin,

$$\int_0^A x^2(\lambda e^{-\lambda x})dx = (-\lambda A^2 - 2\lambda A)e^{-\lambda A} + \frac{2}{\lambda^2}(-e^{-\lambda A}) + \frac{2}{\lambda^2}$$

Or, $\lim\limits_{x \longrightarrow +\infty} xe^{-x} = 0$ et $\lim\limits_{x \longrightarrow +\infty} x^2e^{-x} = 0$.
Donc,

$$\lim_{A \longrightarrow +\infty} \int_0^A x^2 e^{-\lambda x}dx = \frac{2}{\lambda^2}$$

Par conséquent, l'intégrale $\int_0^{+\infty} xe^{-\lambda x}dx$ existe et vaut $\frac{2}{\lambda^2}$.
En conclusion,

$$\boxed{E(X^2) = \frac{2}{\lambda^2}}$$

D'où

$$V(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2$$

Donc
$\boxed{V(X) = \frac{1}{\lambda^2}}$.

La fonction de répartition est :
$F(x) = \int_0^x \lambda e^{-\lambda t}dt = 1 - e^{-\lambda x}$ pour $x \geq 0$ et $F(x) = 0$ sinon.
Par conséquent, la fonction de survie est :
$R(x) = 1 - F(x) = e^{-\lambda x}$ pour $x \geq 0$ et $R(x) = 1$ sinon.$\square$

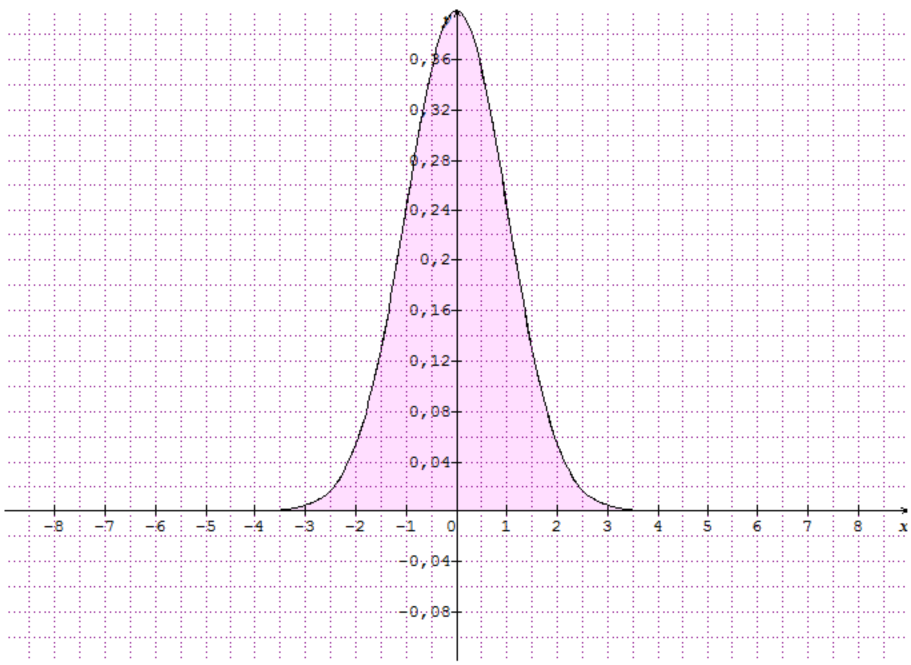
Nous conclurons cette section par des rappels sur la loi normale.

Loi normale centrée réduite

Le lecteur pourra se reporter aux développements dans [2] par exemple. Cette loi est notée $\mathcal{N}(0;1)$.
Sa densité est donnée par
$\boxed{f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}}$ pour $x \in \mathbb{R}$.

Il s'agit bien d'une densité car $\int_0^{+\infty} e^{-\frac{x^2}{2}}dx = \sqrt{2\pi}$.
Ce résultat peut notamment être obtenu à l'aide d'une intégrale double ou d'une transformation de Fourier [9]

9. Le lecteur pourra se reporter à [4], [6] ou à [8].

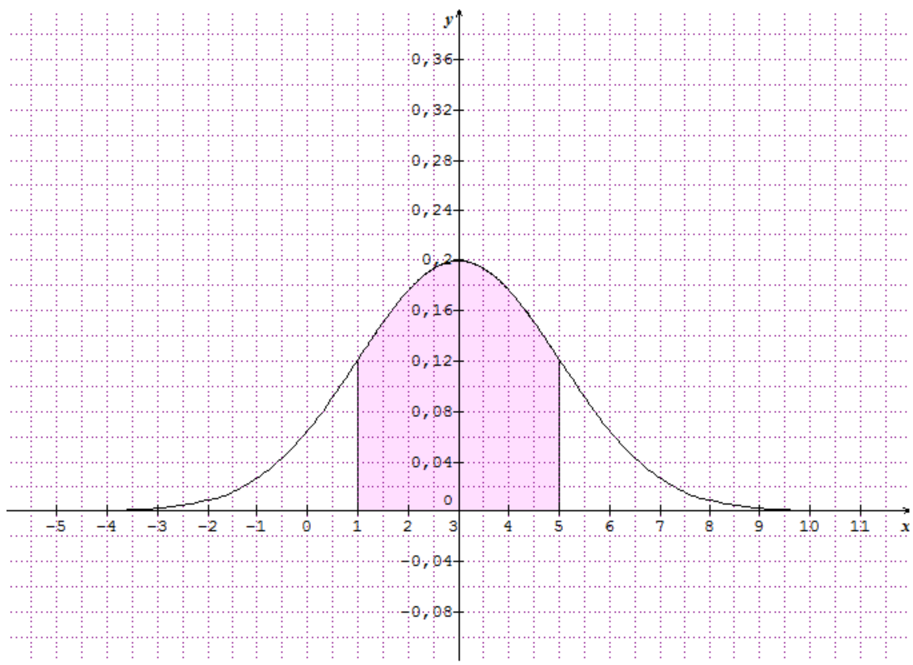


Proposition 9

1. $E(X) = 0$ et $V(X) = 1$,
2. $\mathbb{P}(X \leq 0) = \mathbb{P}(X \geq 0) = 0,5$,
3. $\mathbb{P}(X \leq -a) = \mathbb{P}(X \geq a)$,
4. $\mathbb{P}(-1,96 \leq X \leq 1,96) \approx 0,95$ et $\mathbb{P}(-2,58 \leq X \leq 2,58) \approx 0,99$

Loi normale de paramètres μ et σ

Cette loi est notée $\mathcal{N}(\mu, \sigma^2)$ et une variable aléatoire X suit une loi $\mathcal{N}(\mu, \sigma^2)$ si la variable aléatoire $Y := \frac{X-\mu}{\sigma}$ suit une loi normale centrée réduite.
Sa densité est donnée par
$f(x) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ pour $x \in \mathbb{R}$.



Proposition 10

1. $\mathbb{P}(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.683$,
2. $\mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.954$,
3. $\mathbb{P}(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.997$.

Pour conclure, rappelons quelques éléments sur les vecteurs aléatoires.
En ce qui concerne leurs lois :

- dans le cas discret, la loi du vecteur $(X_1, \ldots, X_d)$ est donnée par : $\mathbb{P}(X_1 = x_1, \ldots, X_d = x_d)$ dans le cas discret
- elle est donnée par la densité $f_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d)$ dans le cas continu.

L'espérance du vecteur aléatoire $(X_1, \ldots, X_d)$ est, par définition, $(E(X_1), \ldots, E(X_d))$. La contrepartie de la variance dans le cas multidimensionnel est la *matrice de variances-covariances*.
Elle est carrée d'ordre d et définie par
$aij = \mathrm{Cov}(X_i, X_j) := E(X_iX_j) - E(X_i)E(X_j)$ pour tout couple $(i, j) \in \{1, \ldots, d\}^2$. par construction, elle est symétrique et d'après ce qui précède, $a_{ii} = E(X_i^2) - E(X_i)^2 = V(X_i)$.

Les lois des variables aléatoires $X_1, \ldots, X_d$ sont appelées *lois marginales du vecteur aléatoire.*
Notons également que, pour un vecteur aléatoire $X = (X_1, \ldots, X_d)$, la fonction caractéristique est définie par : $\phi_X(t) = E(e^{it \bullet X})$ pour $t \in \mathbb{R}^d$ où $\bullet$ désigne le produit scalaire de $\mathbb{R}^d$.
La détermination des lois marginales peut s'avérer difficile et ne sera vue que dans des cas relativement simples dans ce cours.

2 Indépendance

Deux événements sont indépendants [10] si la connaissance de la réalisation de l'un ne change pas la probabilité de réalisation de l'autre. dans le langage bayésien, cela signifie que probabilités a priori et a posteriori coïncident. Formellement, deux événements A et B sont *indépendants* si :

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$$

ou encore $\mathbb{P}(A) = \mathbb{P}_B(A)$.
Pour des variables aléatoires, le principe est le même. Dans le cas discret, deux variables aléatoires X et Y sont dites *indépendantes* si, pour tous réels $x \in \mathcal{S}_X$ et $y \in \mathcal{S}_y$,

$$\mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(\{X = x\}) \times \mathbb{P}(\{Y = y\})$$

où $\mathcal{S}_X$ et $\mathcal{S}_y$ désignent respectivement les supports respectifs des variables aléatoires X et Y.
Dans le cas continu, plusieurs conditions assurent l'indépendance de deux variables aléatoires :

1. $\mathbb{P}(\{X \leq x\} \cap \{Y \leq y\}) = \mathbb{P}(\{X \leq x\}) \times \mathbb{P}(\{Y \leq y)$ i.e. $F_{(X,Y)}(x, y) = F_X(x) F_Y(y)$.
2. $f_{(X,Y)}(x, y) = f_X(x) f_Y(y)$
3. $\phi_{(X_1,X_2)}(t_1, t_2) = \phi_{X_1}(t_1)\phi_{X_2}(t_2)$

10. Ce concept n'est pertinent que pour des événements tous deux de probabilité non nulle.

Lorsque deux variables aléatoires X et Y sont indépendantes,
$E(XY) = E(X)E(Y)$ [11]. Donc $\mathrm{Cov}(X,Y) = 0$.
En d'autres termes deux variables aléatoires indépendantes sont *décorrélées.*

Plus généralement, dès que les expressions considérées ont un sens (en termes d'intégrabilité des fonctions f et g considérées et de leur produit), si X et Y sont deux variables aléatoires sont indépendantes, alors
$E(f(X)g(Y)) = E(f(X))E(g(Y))$.

Exemple 2
Montrer que la somme de n variables aléatoires de Bernoulli de paramètre p indépendantes suit une loi binomiale de paramètres n et p.

Considérons la variable aléatoire $X = \sum_{k=1}^{n} X_k$ somme de n v.a. de Bernoulli de paramètre p.
Nous voulons montrer que X une loi binomiale de paramètres n et p.
Calculons $\phi_{X_1}(t)$ pour t réel.

$$\phi_{X_1}(t) = E(e^{itX_1}) = pe^{it\times 1} + (1-p)e^{it\times 0} = 1 - p + pe^{it}$$

Calculons maintenant $\phi_X(t)$ pour t réel.

$$\phi_X(t) = E(e^{itX}) = E(e^{it\sum_{k=1}^{n} X_k}) = E(\prod_{k=1}^{n} e^{itX_k}) = \prod_{k=1}^{n} \phi_{X_k}(t)$$

car les variables aléatoires X_k sont indépendantes.
Elles sont identiquement distribuées donc $\phi_{X_1}(t) = \phi_{X_k}(t)$ pour tout entier k compris entre 1 et n. Ainsi,

$$\phi_X(t) = \phi_{X_1}(t)^n = \left(1 - p + pe^{it}\right)^n$$

Nous reconnaissons la fonction caractéristique de la loi binomiale de paramètres n et p déterminée supra.$\square$

Exemple 3
Soient X et Y deux variables aléatoires indépendantes. Montrer que :

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$$

11. Démontrer ce résultat.

Complétons cette section par la formule du changement de variables :

Théorème 2
Sous les conditions suivantes :

1. $f(x) > 0$ pour tout réel $x \in I$ où I est un intervalle ouvert,
2. g est bijective de I sur $g(I)$,
3. g est dérivable sur l'intervalle I,

la variable aléatoire $Y = g(X)$ admet pour densité :

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

3 Vecteurs gaussiens

Nous étudierons maintenant l'extension de la loi normale en dimension $d > 1$.

Définition 1
Un vecteur aléatoire $(X_1, \ldots, X_d)$ est dit *gaussien* si toute combinaison linéaire des variables aléatoires X_k est gaussienne.

Un vecteur gaussien est entièrement caractérisé par $m = (E(X_1), \ldots, E(X_d))^T$ et sa matrice de variances-covariances Σ. Sa loi sera notée $\mathcal{N}(m, \Sigma)$ et nous parlerons de loi normale multidimensionnelle.
Certaines propositions de cette section ne seront pas démontrées et le lecteur est invité à se reporter à [3] ou à [5] par exemple pour accéder aux démonstrations.

Proposition 11
Soit X un vecteur aléatoire suivant une loi $\mathcal{N}_d(m, \Sigma)$ (i.e. un vecteur gaussien).
Sa densité est donnée par :

$$f(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^d \frac{1}{\sqrt{|\det \Sigma|}} \exp\left(-\frac{1}{2}(x-m)^T \Sigma^{-1}(x-m)\right)$$

Proposition 12
Soient $X_1, \dots, X_d$ des variables aléatoires indépendantes.
Le vecteur $(X_1, \dots, X_d)^T$ est un vecteur gaussien.

Preuve
Sans restreindre la généralité, supposons les variables aléatoires centrées [12].
La densité de $X = (X_1, \dots, X_d)^T$ est donnée par :

$$f(x) = f(x_1, \dots, x_d) = \prod_{i=1}^{d} \left(\frac{1}{\sigma_i \sqrt{2\pi}} \right) \exp\left(-\frac{1}{2\sigma_i^2} x_i^2 \right)$$

car les variables aléatoires $X_1, \dots, X_d$ sont indépendantes.

$$f(x) = \left(\frac{1}{\sqrt{2\pi}} \right)^d \frac{1}{\prod_{i=1}^{d} \sigma_i} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x \right)$$

en remarquant que la matrice de covariance est

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_d^2 \end{bmatrix}$$

Nous pouvons conclure car $\sqrt{|\det \Sigma|} = \prod_{i=1}^{d} \sigma_i$ et

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma_d^2} \end{bmatrix}$$

Proposition 13
Soit $X = (X_1, \dots, X_d)$ un vecteur gaussien.
Les variables aléatoires $X_1, \dots, X_d$ sont indépendantes si et seulement si la matrice Σ est diagonale.

Notons qu'il existe des variables aléatoires gaussiennes décorrélées qui ne sont pas indépendantes [13].

Proposition 14
Soient $X = (X_1, \dots, X_d)$ un vecteur gaussien et A une matrice de taille $d \times p$.
Le vecteur AX suit une loi $\mathcal{N}(Am, A\Sigma A^T)$.

12. En effet, il suffit de remplacer X_i par $Y_i = X_i - m_i$ puis de faire un changement de variables affine.
13. cf exercice 14 de la feuille 2

4 Convergence de variables aléatoires

Soit (X_i) une suite de variables aléatoires définies sur le même espace Ω et X une variable aléatoire également définie Ω. Nous nous intéresserons à différents modes de convergence de la suite (X_i) vers la variable aléatoire X.

4.1 Convergence presque sûre

Définition 2
La suite (X_i) *converge presque sûrement* vers la variable aléatoire X si $\lim\limits_{n \longrightarrow +\infty} X_n(\omega) = X(\omega)$ pour presque tout $\omega \in \Omega$.

Nous n'expliciterons pas l'expression « pour presque tout $\omega \in \Omega$ » car cela nécessiterait d'aborder des éléments de théorie de mesure qui ne figurent pas dans les pré-requis du cours. Il faut simplement comprendre que ce mode de convergence correspond grossièrement à la convergence simple des suites de fonctions, i.e. à une sorte de convergence ponctuelle.

4.2 Convergence en probabilité

Définition 3
La suite (X_i) *converge en probabilité* vers la variable aléatoire X si

$$\forall \varepsilon > 0, \lim_{n \longrightarrow +\infty} \mathbb{P}(\mid X_n - X \mid \geq \varepsilon) = 0$$

Il est important de connaître les liens entre les différents modes de convergence.

Proposition 15
La convergence presque sûre implique la convergence en probabilité.

La preuve de ce résultat reposant sur la théorie de la mesure, ce résultat sera admis.

4.3 Convergence en loi

Définition 4
La suite (X_i) *converge en loi* vers la variable aléatoire X si $\lim\limits_{n \longrightarrow +\infty} F_{X_n}(x) = F_X(x)$ en tout point de continuité de F_X.

Théorème 3 (Théorème de Paul Lévy)

1. Si la suite de v.a. (X_n) converge en loi vers une v.a. X alors $\lim_{n \longrightarrow +\infty} \phi_{X_n}(t) = \phi_X(t)$ pour tout réel t.

2. Réciproquement, si la suite des fonctions caractéristiques (ϕ_{X_n}) converge simplement vers une fonction ϕ quand n tend vers $+\infty$ et si la fonction ϕ est continue en zéro alors ϕ est la fonction caractéristique d'une variable aléatoire X telle que (X_n) converge vers X en loi.

Ce théorème sera admis. Sa preuve peut être trouvée dans [5] par exemple.

Proposition 16
La convergence en probabilité implique la convergence en loi.

Preuve
Rappelons un lemme classique.

Lemme 1
Soient X et Y deux variables aléatoires et c un réel.

$$\mathbb{P}(Y \leq c) \leq \mathbb{P}(X \leq c + \varepsilon) + \mathbb{P}(X - Y > \varepsilon)$$

Le lemme est immédiat car $\{Y \leq c\} \subset \{X \leq c + \varepsilon\} \cup \{X > c + \varepsilon, Y \leq c\}$.

Par hypothèse, $\lim_{n \longrightarrow +\infty} \mathbb{P}(\mid X_n - X \mid \geq \varepsilon) = 0$ pour tout réel $\varepsilon > 0$.
Remarquons que :

1. $\mathbb{P}(X_n \leq a) \leq \mathbb{P}(X \leq a + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)$

2. $\mathbb{P}(X_n \leq a - \varepsilon) \leq \mathbb{P}(X_n \leq a) + \mathbb{P}(|X_n - X| > \varepsilon)$

Donc

$$\mathbb{P}(X_n \leq a-\varepsilon)-\mathbb{P}(|X_n-X| > \varepsilon) \leq \mathbb{P}(X_n \leq a) \leq \mathbb{P}(X \leq a+\varepsilon)+\mathbb{P}(|X_n-X| > \varepsilon)$$

Nous voulons montrer que, en tout point de continuité a de la fonction F_X,

$$\lim_{n \longrightarrow +\infty} F_{X_n}(a) = F_X(a)$$

En d'autres termes, il faut montrer que, pour tout $\delta > 0$, il existe un entier naturel N tel que, pour tout entier $n \geq N$,

$$(\star)|F_{X_n}(a) - F_X(a)| < \delta$$

Fixons un réel $\varepsilon' > 0$. La fonction F_X étant continue en a,
il existe un réel $\varepsilon > 0$ tel que, pour tout $\gamma \in [-\varepsilon; \varepsilon]$, $|F_X(a+\gamma) - F(a)| < \dfrac{\varepsilon'}{2}$.
En particulier, on a :
$|F_X(a-\varepsilon) - F(a)| < \dfrac{\varepsilon'}{2}$ et $|F_X(a+\varepsilon) - F(a)| < \dfrac{\varepsilon'}{2}$.
Par ailleurs, la convergence en probabilité nous assure qu'il existe un entier naturel N tel que pour tout entier $n \geq N$, $\mathbb{P}(|X_n - X| > \varepsilon) < \dfrac{\varepsilon'}{2}$.
L'inégalité $(\star)$ nous permet de conclure que, pour tout entier $n \geq N$,
$|F_{X_n}(a) - F_X(a)| < \dfrac{\varepsilon'}{2} + \dfrac{\varepsilon'}{2} = \varepsilon'$.

Théorème 4 (Théorème de Slutsky)
Soient (X_n) et (Y_n) deux suites de variables aléatoires.
Si la suite (X_n) converge en loi vers une variable aléatoire X et Y_n converge en probabilité vers une constante c alors le couple (X_n, Y_n) converge vers le couple en loi (Y, c).

Ce théorème sera admis. Il est prouvé dans [3] et dans [5].

4.4 Convergence en norme L^1

Définition 5
La suite (X_i) *converge en norme* L^1 vers la variable aléatoire X si
$\lim\limits_{n \longrightarrow +\infty} E(|X_n - X|) = 0$.

Proposition 17
La convergence en moyenne quadratique implique la convergence en probabilité.

Preuve
Rappelons tout d'abord l'*inégalité de Markov* :
Soit Y une variable aléatoire positive et a un réel supérieur à 0,

$$\mathbb{P}(Y \geq a) \leq \frac{E(Y)}{a}$$

Revenons à la démonstration. Soient (X_i) une suite convergeant en norme L^1 vers une variable aléatoire X et un réel $\varepsilon > 0$.
Pour tout entier n,

$$\mathbb{P}(|\ X_n - X\ | \geq \varepsilon) \leq \frac{E(|X_n - X|)}{\varepsilon}$$

grâce à l'inégalité de Markov.
Par hypothèse, $\lim_{n \longrightarrow +\infty} E(|X_n - X|) = 0$.
Donc $\lim_{n \longrightarrow +\infty} \mathbb{P}(| X_n - X | \geq \varepsilon) = 0. \square$

4.5 Convergence en norme L^2

Définition 6
La suite (X_i) *converge en norme* L^2 (ou en *moyenne quadratique*) vers la variable aléatoire X si $\lim_{n \longrightarrow +\infty} E(|X_n - X|^2) = 0$.

Proposition 18
La convergence en moyenne quadratique implique la convergence en probabilité.

Preuve
Soient (X_i) une suite convergeant en moyenne quadratique vers une variable aléatoire X et un réel $\varepsilon > 0$.
Pour tout entier n,

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

grâce à l'inégalité de Markov.
Par hypothèse, $\lim_{n \longrightarrow +\infty} E(|X_n - X|^2) = 0$.
Donc $\lim_{n \longrightarrow +\infty} \mathbb{P}(| X_n - X | \geq \varepsilon) = 0. \square$

5 Théorème Central Limite

Théorème 5 (Loi forte des grands nombres)
Soit (X_i) une suite de variables aléatoires i.i.d. (indépendantes et identiquement distribuées) telle que $E(|X_1|) < +\infty$.
Notons $m := E(X_1)$.

$$\lim_{n \longrightarrow +\infty} \bar{X}_n = m$$

au sens de la convergence p.s. où $\bar{X}_n := \dfrac{X_1 + \cdots + X_n}{n}$

Ce théorème sera fondamental en statistique mathématique. En termes statistiques, il signifie que la *moyenne empirique* est un estimateur consistent de l'espérance.

Théorème 6 (T.C.L. cas unidimensionnel)

1. Soit (X_i) une suite v.a. i.i.d.
2. Notons $m := E(X_i)$ et $\sigma^2 = V(X_i)$.
3. $\dfrac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$ converge en loi vers une loi normale centrée réduite.

Preuve

Remarquons que $\sqrt{n}\dfrac{(\bar{X}_n - m)}{\sigma} = \dfrac{1}{\sqrt{n}}\left(\dfrac{S_n - nm}{\sigma}\right) = \dfrac{1}{\sqrt{n}}\left(\displaystyle\sum_{k=1}^{n}\dfrac{(X_k - m)}{\sigma}\right)$

où $S_n = \displaystyle\sum_{k=1}^{n} X_k$.

D'après l'égalité précédente, quitte à remplacer X_k par $Y_k = X_k - m$, on peut supposer les variables aléatoires X_k centrées.

Nous étudions ainsi $Z_n := \dfrac{1}{\sqrt{n}}\dfrac{S_n}{\sigma}$.

Pour tout $t \in \mathbb{R}$,

$$\phi_{Z_n}(t) = E(e^{itZ_n}) = E(e^{i\frac{t}{\sigma\sqrt{n}}S_n}) = \phi_{S_n}\left(\frac{t}{\sigma\sqrt{n}}\right)$$

Les variables aléatoires X_k sont indépendantes donc

$$\phi_{Z_n}(t) = \phi_{S_n}\left(\frac{t}{\sigma\sqrt{n}}\right) = \left(\phi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n$$

Remarquons que la fonction caractéristique est deux fois dérivable car la variable aléatoire admet un moment d'ordre 2. En utilisant un développement limité [14] d'ordre 2 en 0 :

$$\phi_{X_1}(u) = 1 + u\phi'_{X_1}(0) + \frac{u^2}{2}\phi''_{X_1}(0) + u^2\varepsilon(u)$$

où $\lim\limits_{u\longrightarrow 0}\varepsilon(u) = 0$.

En rappelant que la variable aléatoire X_1 est centrée,
$\phi'_{X_1}(0) = iE(X) = 0$ et $\phi''_{X_1}(0) = -E(X^2) = -V(X) = -\sigma^2$.
Par conséquent,

$$\left(\phi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2\sigma^2 n} + \frac{t^2}{n}\varepsilon(t)\right)^n$$

14. Le lecteur pourra consulter [4] pour une étude des développements limités.

Nous admettrons le lemme suivant d'analyse complexe [15].

Lemme 2
Soit (u_n) une suite de nombres complexes convergeant vers un nombre complexe z.

$$\lim_{n \longrightarrow +\infty} \left(1 + \frac{u_n}{n}\right)^n = e^z$$

En rappelant que $\phi''_{X_1}(0) = -E(X^2) = \sigma^2$ car la variable aléatoire est supposée centrée, nous pouvons ainsi conclure que
$$\lim_{n \longrightarrow +\infty} \phi_{Z_n}(t) = \lim_{n \longrightarrow +\infty} \left(1 - \frac{t^2}{2\sigma^2 n} \times \sigma^2 + \frac{t^2}{n}\varepsilon(t)\right)^n = \lim_{n \longrightarrow +\infty} \left(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon(t)\right)^n$$
donc $\lim_{n \longrightarrow +\infty} \phi_{Z_n}(t) = e^{-\frac{t^2}{2}} = \phi_Z(t)$ où Z désigne la fonction caractéristique de la loi $\mathcal{N}(0,1)$.
La fonction ϕ_Z est continue sur $\mathbb{R}$ donc a fortiori en 0 donc le théorème de Paul Lévy nous permet de conclure.$\square$

Le T.C.L. sera tout aussi fondamental en statistique.
En effet, il nous fournit :

- une vitesse de convergence $\dfrac{1}{\sqrt{n}}$,
- un moyen d'obtenir des intervalles de confiance.

Théorème 7 (T.C.L. cas multidimensionnel)

1. Soit (X_i) une suite vecteurs aléatoires de $\mathbb{R}^d$ i.i.d.
2. Notons $m := E(X_i) \in \mathbb{R}^d$ et Σ la matrice de variances-covariances.
3. $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i - m\right)$ converge en loi vers une loi normale multidimensionnelle $N(0,\Sigma)$.

Preuve
La preuve de ce résultat est laissée en exercice. Il suffira de s'inspirer du cas unidimensionnel.

15. cf. [9] pour une démonstration de ce lemme. Pour de plus amples développements sur l'analyse complexe, le lecteur est invité à se reporter à [7].

6 Approximation de lois

Approximation de la loi binomiale par la loi normale

Commençons cette section par le célèbre théorème de Moivre-Laplace. Moivre l'a initialement prouvé dans le cas symétrique $p = \dfrac{1}{2}$ puis Laplace l'a généralisé pour $p \in]0;1[$ d'où sa dénomination.

Théorème 8 (Théorème de Moivre-Laplace)
Soit X_n une variable aléatoire suivant une loi $\mathcal{B}(n,p)$.
Posons $q := 1 - p$.
Pour tout réel x,

$$\lim_{n \longrightarrow +\infty} \mathbb{P}\left(\frac{X_n - np}{\sqrt{npq}}) \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt = F(x)$$

où F désigne la fonction de répartition de la loi normale centrée réduite.

Preuve
Il s'agit d'un cas particulier du théorème central limite appliqué à des variables aléatoires indépendantes suivant une loi $\mathcal{B}(p)$.

Ce théorème permettra d'approcher la loi binomiale $\mathcal{B}(n,p)$ par la loi normale $\mathcal{N}(n,pq)$ pour n suffisamment grand.
Les conditions suivantes rendent possible cette approximation :

(i) $n \geq 30$,

(ii) $np \geq 5$,

(iii) $nq \geq 5$.

Approximation de la loi binomiale par la loi de Poisson

Soit $\lambda > 0$. Considérons une suite $(X_{n,p})_{n \geq 1}$ de variables aléatoires suivant une loi $\mathcal{B}(n,p)$ et supposons que les valeurs n et p sont reliées par la relation suivante $np = \lambda$.

$$\lim_{n \longrightarrow +\infty, p \longrightarrow 0} \mathbb{P}(X_n = k) = \lim_{n \longrightarrow +\infty, p \longrightarrow 0} \binom{n}{k} p^k (1-p)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}$$

L'expression a toujours un sens pour $n \geq k$.

La loi binomiale $\mathcal{B}(n, p)$ peut ainsi être approchée par la loi de Poisson $\mathcal{P}(np)$ sous certaines conditions.
Voici deux jeux de conditions de validité de l'approximation par la loi de Poisson $\mathcal{P}(np)$:

(i) $n \geq 20$ et $p \leq 0,05$,

(ii) $n \geq 100$ et $np \leq 10$.

Les conditions sur n et p assurent que la valeur de np est relativement stable, ce qui permet d'utiliser le résultat évoqué supra.

Références

[1] Dacunha-Castelle, D. : *Chemins de l'aléatoire*. Flammarion (2000)

[2] Foata, D., Franchi, J. Fuchs, A. : *Calcul de probabilités*. Dunod (2012)

[3] Monfort, A. : *Cours de probabilités*. Economica (1996)

[4] Ramis, E., Deschamp, C., Odoux,J. : *Le cours de mathématiques T3. Topologie et éléments d'analyse*. Dunod (2017)

[5] Revuz, D. : *Probabilités*. Hermann (1997)

[6] Rudin, W. : P*rincipes d'analyse mathématique*. Dunod (2006)

[7] Stein, E., Shakarchi, R. : *Complex Analysis*. Princeton University Press (2003)

[8] Stein, E., Shakarchi, R. : *Fourier Analysis*. Princeton University Press (2003)

[9] Toulouse, P.S. : *Thèmes de probabilités et de statistiques*. Dunod (1999)

[10] Williams, D. : *Probability with Martingales*. Cambrige Mathematical Textbooks (1991)