Polycopié PBS2

N. Biheng

Table des matières

Chapitre 1

Probabilités

1.1 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 $P(B) \neq 0$.

1. Pour des développements non techniques sur les probabilités et leurs applications, le lecteur pourra consulter [3].

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] :

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

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 $A \subset \mathcal{S}$, $\mathbb{P}(A) = \displaystyle\sum_{x \in A} p_x$.
L'*espérance* est définie par $E(X) := \displaystyle\sum_{x \in \mathcal{S}} x p_x$ et correspond à la « valeur moyenne ». Elle est parfois notée $\bar{x}$.
La *variance* est définie par $V(X) := \displaystyle\sum_{x \in \mathcal{S}} p_x (x - \bar{x})^2$ 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. sa démonstration qui procède de la définition des probabilités conditionnelles est laissée en exercice.

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.

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 :

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

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 :

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

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 $V(X)=np(1-p)$.

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

Preuve
Calculons $E(X)$ et $V(X)$.
$E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k(1-p)^{n-k} = p \sum_{k=1}^{n} \binom{n}{k} kp^{k-1}(1-p)^{n-k}$.
Donc, en remarquant que, $k\binom{n}{k} = n\frac{(n-1)!}{(n-1-(k-1))!(k-1)!} = n\binom{n-1}{k-1}$,
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, $E(X(X-1)) = p^2 \sum_{k=2}^{n} \binom{n}{k} k(k-1)p^{k-2}(1-p)^{n-k}$.
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, $\boxed{V(X) = np(1-p)}$.$\square$

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 :
$E(X) = \displaystyle\sum_{k\geq 1} kpq^{k-1} = p\sum_{k\geq 1} kq^{k-1}$.

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

Or $\dfrac{1}{1-q} = \displaystyle\sum_{k\geq 0} q^k$ donc $\dfrac{1}{(1-q)^2} = \displaystyle\sum_{k\geq 1} kq^{k-1}$.
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 $q \in]0;1[$, $\left(\displaystyle\sum_{k\geq 0} q^k\right)'' = \displaystyle\sum_{k\geq 2} k(k-1)q^{k-2}$.
Donc, comme précédemment,

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

Donc

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

5. Le lecteur pourra consulter [10] pour des rappels sur les séries entières.

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 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 $k \geq 0$,
$\boxed{\mathbb{P}(X = k) = e^{-\lambda}\dfrac{\lambda^k}{k!}}$

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

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

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

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

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 [4].

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 $k \in \{1, \ldots, n\}$, $\phi^k(0) = i^k E(X^k)$,
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 $P(X \in [c;d])$.
$\mathbb{P}(X \in [c;d]) = \int_c^d f(x)dx = \int_c^d \dfrac{1}{b-a}dx.$
Par conséquent, $\mathbb{P}(X \in [c;d]) = \left[\dfrac{1}{b-a}x\right]_c^d = \dfrac{d}{b-a} - \dfrac{c}{b-a}.$
En conclusion,
$\boxed{\mathbb{P}(X \in [c;d]) = \dfrac{d-c}{b-a}}.$

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

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

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

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

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

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

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 $E(X)$.

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

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

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

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

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,

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

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

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

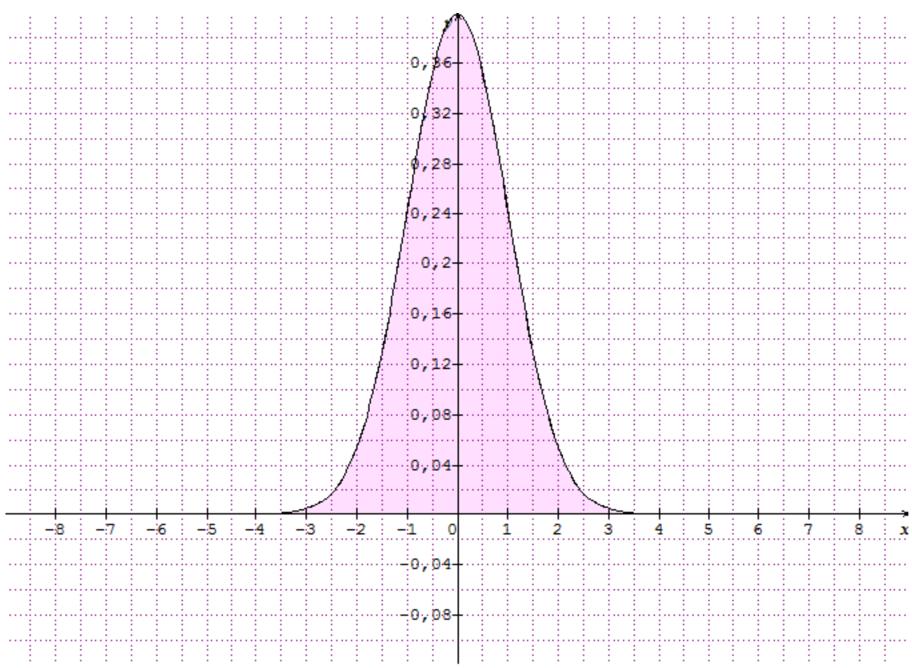
Donc
$\boxed{V(X) = \frac{1}{\lambda^2}}$.

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 [4] 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]



Proposition 9

1. $E(X) = 0$ et $V(X) = 1$,

9. Le lecteur pourra se reporter à [8], [10] ou à [12].

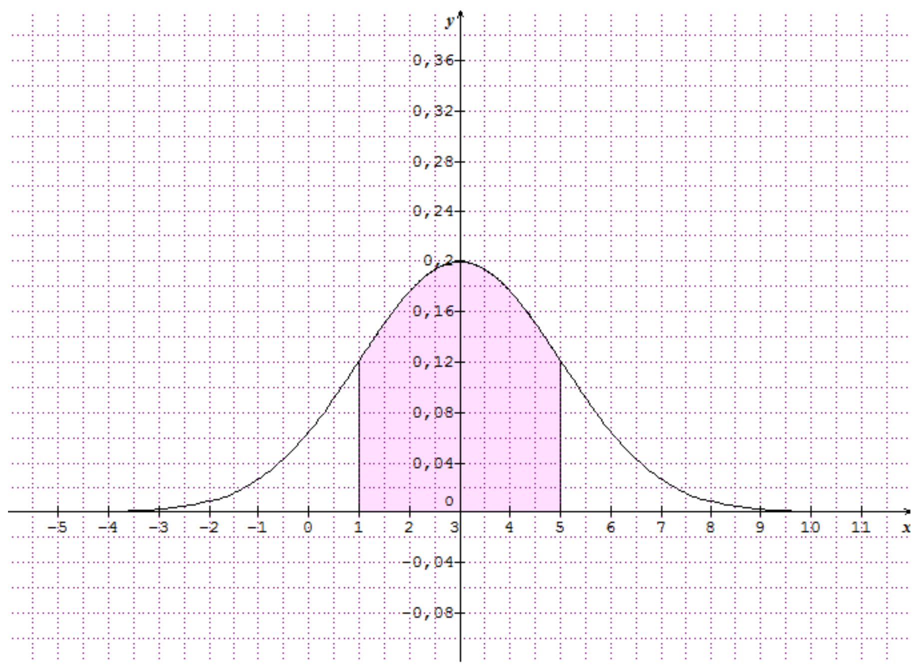
2. $\mathbb{P}(X \leq 0) = \mathbb{P}(X \geq 0) = 0,5$,

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.

1.2 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 $y \in \mathcal{S}_y$,

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

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

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. $\phi_{(X_1,X_2)}(t_1,t_2) = \phi_{X_1}(t_1)\phi_{X_2}(t_2)$

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 $X = \displaystyle\sum_{k=1}^{n} X_k$ 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

11. Démontrer ce résultat.

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 :

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

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 :

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

1.3 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 [**?**] ou à [9] 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 $X_1, \ldots, X_d$ des variables aléatoires indépendantes.
Le vecteur $(X_1, \ldots, X_d)^T$ est un vecteur gaussien.

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

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

car les variables aléatoires $X_1, \ldots, X_d$ 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

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

12. En effet, il suffit de remplacer X_i par $Y_i = X_i - m_i$ puis de faire un changement de variables affine.

Proposition 13
Soit $X = (X_1, \ldots, X_d)$ un vecteur gaussien.
Les variables aléatoires $X_1, \ldots, X_d$ 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 $X = (X_1, \ldots, X_d)$ un vecteur gaussien et A une matrice de taille $d \times p$.
Le vecteur AX suit une loi $\mathcal{N}(Am, A\Sigma A^T)$.

1.4 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.

1.4.1 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.

1.4.2 Convergence en probabilité

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

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

13. cf exercice 14 de la feuille 2

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.

1.4.3 Convergence en loi

Définition 4
La suite (X_i) *converge en loi* vers la variable aléatoire X si
$\lim_{n \longrightarrow +\infty} F_{X_n}(x) = F_X(x)$ 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 [9] 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 [**?**] et dans [9].

1.4.4 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,

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

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

1.4.5 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 $\lim\limits_{n \longrightarrow +\infty} E(|X_n - X|^2) = 0$.

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,

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

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

1.5 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,

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

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 $\lim\limits_{n\longrightarrow+\infty}\phi_{Z_n}(t) = e^{-\frac{t^2}{2}} = \phi_Z(t)$ 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.□

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

- une vitesse de convergence $\frac{1}{\sqrt{n}}$,

14. Le lecteur pourra consulter [8] pour une étude des développements limités.
15. cf. [14] pour une démonstration de ce lemme. Pour de plus amples développements sur l'analyse complexe, le lecteur est invité à se reporter à [11].

- 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. $\sqrt{n}\left(\dfrac{1}{n}\sum_{i=1}^{n} X_i - m\right)$ 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.

1.6 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.

Compléments

1.7 Fonctions caractéristiques

Certaines fonctions caractéristiques devront être connues par cœur.

Loi de Bernoulli de paramètre p

$\phi(t) = 1 - p + pe^{it}$ pour $t \in \mathbb{R}$

Loi binomiale de paramètres n et p

$\phi(t) = (1 - p + pe^{it})^n$ pour $t \in \mathbb{R}$

Loi de Poisson de paramètre λ

$\phi(t) = \phi(t) = \exp(\lambda(e^{it} - 1))$ pour $t \in \mathbb{R}$

Ce résultat n'ayant pas été prouvé en classe, démontrons-le.
Soit X une variable aléatoire suivant une loi de Poisson.
On a : $\phi(t) = E(e^{itX})$.
$\phi(t) = \sum_{k\geq 0} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k\geq 0} \frac{(\lambda e^{it})^k}{k!} = \exp(-\lambda)\exp(\lambda e^{it})$ donc

$$\boxed{\phi(t) = \exp(\lambda(e^{it} - 1))}$$

Loi géométrique de paramètre p Notons $q = 1 - p$.

$$\boxed{\phi(t) = \frac{pe^{it}}{1 - qe^{it}}}$$

Le lecteur est fortement incité à vérifier ce résultat.

Loi exponentielle de paramètre λ

$\phi(t) = \dfrac{\lambda}{\lambda - it}$ pour $t \in \mathbb{R}$

Loi normale centrée réduite

$\phi(t) = e^{-\frac{t^2}{2}}$ pour $t \in \mathbb{R}$.

Nous admettrons ce résultat classique en analyse de Fourier.

Loi normale de paramètres m et σ^2

$\phi(t) = e^{-\frac{\sigma^2 t^2}{2}+itm}$ pour $t \in \mathbb{R}$.

Montrons ce résultat.Si X suit une telle loi, alors $Y = \dfrac{X-m}{\sigma}$ suit une loi normale centrée réduite. Donc $\phi_Y(t) = e^{-\frac{t^2}{2}}$.
Or $\phi_Y(t) = E(e^{itY}) = E(e^{it\frac{X}{\sigma}}e^{-it\frac{m}{\sigma}}) = e^{-itm}\phi_X(\frac{t}{\sigma})$
D'où $\phi_X(t) = \phi_Y(\sigma t)e^{itm} = \boxed{e^{-\frac{\sigma^2 t^2}{2}+itm}}$.

Exemple 4
Soient deux variables aléatoires indépendantes suivant respectivement des lois $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$.
Montrer que la variable aléatoire $X_1 + X_2$ suit une loi normale $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Déterminons la fonction caractéristique de la variable aléatoire $X_1 + X_2$.
Les variables aléatoires sont indépendantes donc, d'après le cours,

$$\phi_{X_1+X_2}(t) = \phi_{X_1}(t)\phi_{X_2}(t) = e^{-\frac{\sigma_1^2 t^2}{2}+itm_1}e^{-\frac{\sigma_2^2 t^2}{2}+itm_2}$$

d'où

$$\phi_{X_1+X_2}(t) = e^{-\frac{(\sigma_1^2+\sigma_2^2)t^2}{2}+it(m_1+m_2)}$$

La proposition suivante sera particulièrement utile :

Proposition 19
Soit X une variable aléatoire telle que $E(|X^2|) < +\infty$.
On a : $\boxed{\phi'(0) = iE(X)}$ et $\boxed{\phi''(0) = i^2 E(X^2) = -E(X^2)}$.

Nous renvoyons le lecteur à [4] pour une démonstration.

Exemple 5
Déduisons de cette proposition l'espérance et la variance d'une variable aléatoire X suivant une loi de Poisson de paramètre λ.

Notons que la variable aléatoire X est de carré intégrable.
D'après ce qui précède, pour tout réel t

$$\phi(t) = \exp(\lambda(e^{it} - 1))$$

On a :

$$\phi'(t) = i\lambda e^{it} \exp(\lambda(e^{it} - 1))$$

par la dérivée d'une composée.
D'où

$$\phi'(0) = i\lambda e^{i\times 0} \exp(\lambda(e^{i\times 0} - 1)) = i\lambda$$

donc

$$E(x) = \frac{\phi'(0)}{i} = \boxed{\lambda}$$

de même

$$\phi''(t) = i^2\lambda e^{it} \exp(\lambda(e^{it} - 1)) + i^2\lambda^2 e^{2it} \exp(\lambda(e^{it} - 1))$$

et

$$\phi''(0) = -\lambda - \lambda^2$$

D'où

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

et

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

1.8 Résultats complémentaires sur les convergence

Théorème 9 (Théorème de Slutsky)
Soient $g : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction continue, (X_n) une suite de variables aléatoires et X une variable aléatoire.
Si la suite (X_n) converge en probabilité vers la variable aléatoire X en probabilité alors la suite $(g(X_n))$ converge en probabilité vers la variable aléatoire $g(X)$.

De même pour la convergence en loi,

Théorème 10
Soient $g : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction continue, (X_n) une suite de variables aléatoires et X une variable aléatoire.
Si la suite (X_n) converge en loi vers la variable aléatoire X alors la suite $(g(X_n))$ converge en loi vers la variable aléatoire $g(X)$.

Ces propriétés s'étendent aisément aux vecteurs aléatoires avec des fonctions $g : \mathbb{R}^d \longrightarrow \mathbb{R}^p$.

1.9 Rappels d'analyse

Le lecteur est invité à se référer à [8] pour de plus amples détails ou à consulter l'excellent [**?**] qui lui permettra d'envisager les différentes notions sous un angle historique.
Naturellement, ces compléments ont vocation à constituer des rappels mais l'étudiant n'est pas tenu de connaître les démonstrations des différents énoncés.
Une fonction f définie sur un intervalle $]-a;a[$ est développable en série entière s'il existe un réel $b \in]0;a]$ tel que $f(x) = \sum_{k\geq 0} a_k x^k$ pour tout réel $x \in]-b;b[$. La plus grande valeur de b possible sera appelé *rayon de convergence.*
Dans ces cas, le développement en série entière est **unique** et la fonction f est de classe $\mathcal{C}^\infty$.

Proposition 20
Soit f une fonction développable en série entière telle que
pour tout $x \in]-R;R[$,
$f(x) = \sum_{k\geq 0} a_k x^k$.
Alors

1. la fonction f de classe $\mathcal{C}^\infty$.
2. Sa dérivée est $f'(x) = \sum_{k\geq 1} k a_k x^{k-1}$.

De plus, nous disposerons du résultat suivant

Proposition 21
Soit f une fonction développable en série entière telle que
pour tout $x \in]-R;R[$, $f(x) = \sum_{k\geq 0} a_k x^k$.

Alors, pour tout réel $x \in]-R;R[$.

$$f(x) = \sum_{k \geq 0} \frac{f^k(x_0)}{k!} (x - x_0)^k$$

Ainsi, la série est somme de sa *série de Taylor*.

Exemple 6
Nous avons utilisé ce résultat dans le chapitre 1 lors de la détermination de l'espérance d'une loi géométrique.
Nous avions $f(q) = \sum_{k \geq 0} q^k = \dfrac{1}{1-q}$ pour $q \in]0;1[$.
Par ailleurs, $f(q) = \dfrac{1}{1-q}$.
La première proposition assure que, pour tout réel $q \in]-1;1[$,

$$f'(q) = \sum_{k \geq 1} kq^{k-1}$$

donc

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

Le lecteur est renvoyé au chapitre 1 pour revoir le lien avec la loi géométrique.

Exemple 7
Nous aurions pu utiliser ce résultat pour obtenir l'espérance et la variance d'une variable aléatoire suivant une loi $\mathcal{P}(\lambda)$.
En effet, on sait que la fonction exponentielle est développable en série entière sur $\mathbb{R}$ et que :

$$f(\lambda) = e^{\lambda} = \sum_{k \geq 0} \frac{\lambda^k}{k!}$$

pour tout réel λ.

Ce qu'il faut retenir

- Lien entre les lois et leur interprétation : binomiale, géométrique...

- Espérance et variance des lois usuelles.
- Savoir refaire les calculs pour les déterminer si cela est demandé.
- Propriétés de la loi normale dont les suivantes :
  1. $\mathbb{P}(X \leq 0) = \mathbb{P}(X \geq 0) = 0,5$,
  2. $\mathbb{P}(X \leq -a) = \mathbb{P}(X \geq a)$,
  3. $\mathbb{P}(-1,96 \leq X \leq 1,96) \approx 0,95$ et $\mathbb{P}(-2,58 \leq X \leq 2,58) \approx 0,99$
  4. D'où $z_{0,975} \approx 1,96$ et $z_{0,995} \approx 2,58$ en particulier [16].
- Définition d'un vecteur gaussien.
- Enoncé du TCL
- Calcul de la fonction caractéristique : en particulier, il faudra l'utiliser lorsqu'on veut déterminer la loi de la somme de variables aléatoires indépendantes.
- Fonctions caractéristiques usuelles.
- Modes de convergence (plus théorique et moins utile pour l'examen).

16. Ces valeurs doivent être connues par cœur.

Chapitre 2

Estimation

2.1 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 $P(B) \neq 0$.

1. Pour des développements non techniques sur les probabilités et leurs applications, le lecteur pourra consulter [3].

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] :

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

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 22

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

---

2. sa démonstration qui procède de la définition des probabilités conditionnelles est laissée en exercice.

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.

2. Si X et Y sont deux variables aléatoires

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

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

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 :

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

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 23
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 24
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 $\{0,\dots,n\}$.
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 25
On a :
$E(X)=np$ et $V(X)=np(1-p)$.

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

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

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, $E(X(X-1)) = p^2 \sum_{k=2}^{n} \binom{n}{k} k(k-1) p^{k-2} (1-p)^{n-k}$.

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, $\boxed{V(X) = np(1-p)}$. $\square$

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 26
On a :
$E(X) = \dfrac{1}{p}$ et $V(X) = \dfrac{q}{p^2}$,

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

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

Or $\dfrac{1}{1-q} = \displaystyle\sum_{k\geq 0} q^k$ donc $\dfrac{1}{(1-q)^2} = \displaystyle\sum_{k\geq 1} kq^{k-1}$.
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 $q \in]0;1[$, $\left(\displaystyle\sum_{k\geq 0} q^k\right)'' = \displaystyle\sum_{k\geq 2} k(k-1)q^{k-2}$.

Donc, comme précédemment,

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

Donc

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

5. Le lecteur pourra consulter [10] pour des rappels sur les séries entières.

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 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 $k \geq 0$,
$\boxed{\mathbb{P}(X = k) = e^{-\lambda}\dfrac{\lambda^k}{k!}}$

Proposition 27
On a :
$E(X) = \lambda$ et $V(X) = \lambda$.

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

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

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

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 8
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 [4].

Théorème 11
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 $\lim\limits_{u \longrightarrow 0} \varepsilon(|u^n|) = 0$.

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 28

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

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,
$E(X) = \dfrac{b^2}{2(b-a)} - \dfrac{a^2}{2(b-a)} = \dfrac{b^2-a^2}{2(b-a)} = \dfrac{(b-a)(b+a)}{2(b-a)}$.
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,
$V(X) = \frac{b^2-2ab-a^2}{12} = \boxed{\frac{(b-a)^2}{12}}$. □

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 29

.

1. $E(X) = \frac{1}{\lambda}$ et $V(X) = \frac{1}{\lambda^2}$,
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 $E(X)$.

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

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

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

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

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,

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

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

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

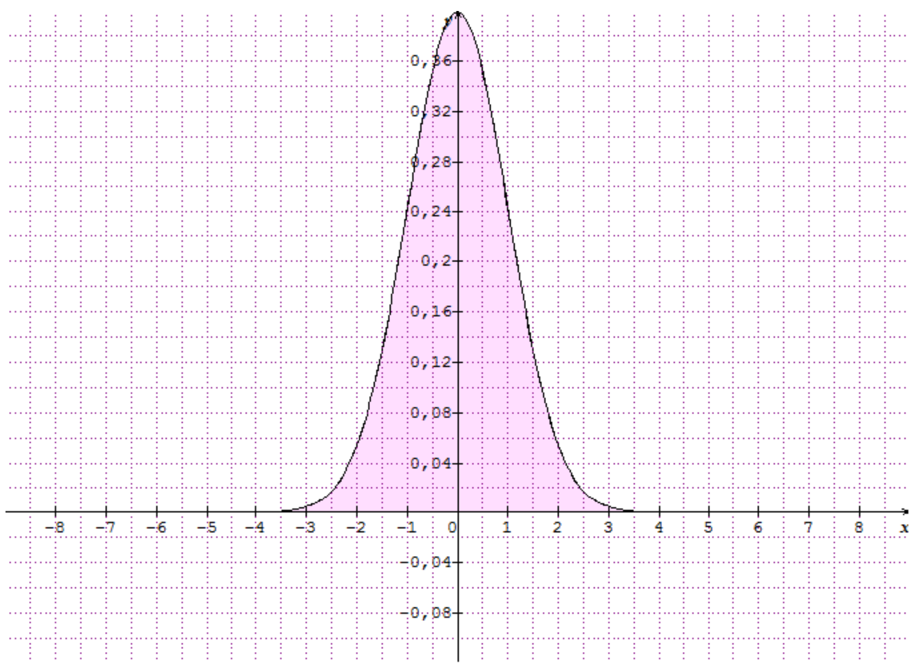
Donc
$\boxed{V(X) = \frac{1}{\lambda^2}}$.

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 [4] 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]



Proposition 30

1. $E(X) = 0$ et $V(X) = 1$,

9. Le lecteur pourra se reporter à [8], [10] ou à [12].

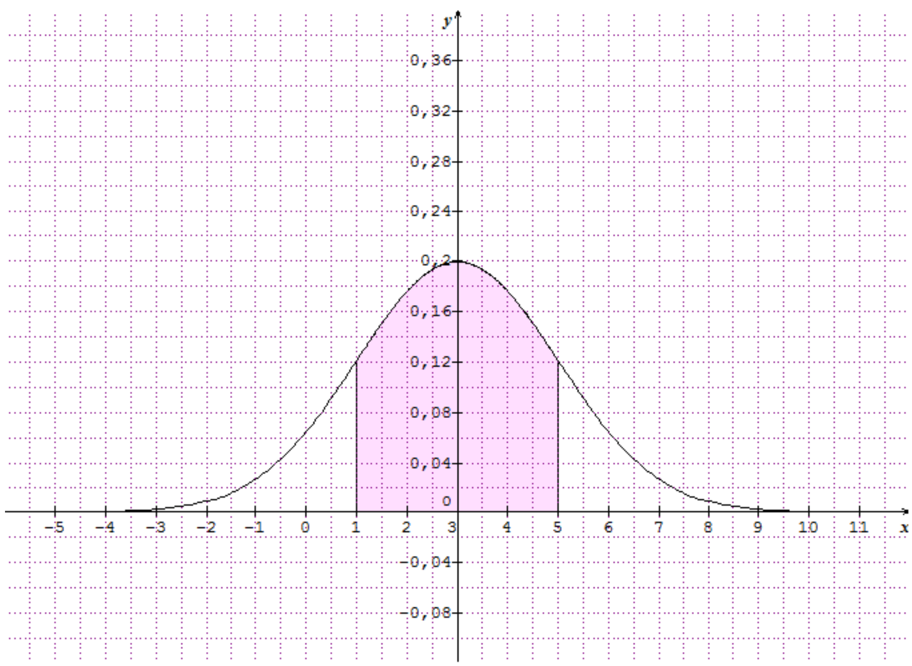
2. $\mathbb{P}(X \leq 0) = \mathbb{P}(X \geq 0) = 0,5$,

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 31

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.

2.2 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 $y \in \mathcal{S}_y$,

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

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

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. $\phi_{(X_1,X_2)}(t_1,t_2) = \phi_{X_1}(t_1)\phi_{X_2}(t_2)$

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 9
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 $X = \displaystyle\sum_{k=1}^{n} X_k$ 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

11. Démontrer ce résultat.

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.□

Exemple 10
Soient X et Y deux variables aléatoires indépendantes. Montrer que :

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

Complétons cette section par la formule du changement de variables :

Théorème 12
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 :

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

2.3 Vecteurs gaussiens

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

Définition 7
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 [**?**] ou à [9] par exemple pour accéder aux démonstrations.

Proposition 32
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 33
Soient $X_1, \ldots, X_d$ des variables aléatoires indépendantes.
Le vecteur $(X_1, \ldots, X_d)^T$ est un vecteur gaussien.

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

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

car les variables aléatoires $X_1, \ldots, X_d$ 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

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

12. En effet, il suffit de remplacer X_i par $Y_i = X_i - m_i$ puis de faire un changement de variables affine.

Proposition 34
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 35
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 $\mathcal{N}(Am, A\Sigma A^T)$.

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.

2.4.1 Convergence presque sûre

Définition 8
La suite (X_i) *converge presque sûrement* vers la variable aléatoire X si $\lim_{n \longrightarrow +\infty} X_n(\omega) = X(\omega)$ 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.

2.4.2 Convergence en probabilité

Définition 9
La suite (X_i) *converge en probabilité* vers la variable aléatoire X si

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

13. cf exercice 14 de la feuille 2

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

Proposition 36
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.

2.4.3 Convergence en loi

Définition 10
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 13 (Théorème de Paul Lévy)

1. Si la suite de v.a. (X_n) converge en loi vers une v.a. X alors $\lim\limits_{n \longrightarrow +\infty} \phi_{X_n}(t) = \phi_X(t)$ 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 [9] par exemple.

Proposition 37
La convergence en probabilité implique la convergence en loi.

Preuve
Rappelons un lemme classique.

Lemme 3
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, $\lim\limits_{n \longrightarrow +\infty} \mathbb{P}(\mid X_n - X \mid \geq \varepsilon) = 0$ 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 14 (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 [**?**] et dans [9].

2.4.4 Convergence en norme L^1

Définition 11
La suite (X_i) *converge en norme* L^1 vers la variable aléatoire X si
$\lim_{n \longrightarrow +\infty} E(|X_n - X|) = 0$.

Proposition 38
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,

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

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

2.4.5 Convergence en norme L^2

Définition 12
La suite (X_i) *converge en norme* L^2 (ou en *moyenne quadratique*) vers la variable aléatoire X si $\lim\limits_{n \longrightarrow +\infty} E(|X_n - X|^2) = 0$.

Proposition 39
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,

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

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

2.5 Théorème Central Limite

Théorème 15 (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 16 (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,

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

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

Lemme 4
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 $\lim\limits_{n \longrightarrow +\infty} \phi_{Z_n}(t) = e^{-\frac{t^2}{2}} = \phi_Z(t)$ 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 $\dfrac{1}{\sqrt{n}}$,

14. Le lecteur pourra consulter [8] pour une étude des développements limités.
15. cf. [14] pour une démonstration de ce lemme. Pour de plus amples développements sur l'analyse complexe, le lecteur est invité à se reporter à [11].

- un moyen d'obtenir des intervalles de confiance.

Théorème 17 (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. $\sqrt{n}\left(\dfrac{1}{n}\sum_{i=1}^n X_i - m\right)$ 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.

2.6 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 18 (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.

2.7 Cadre général

> « L'objectif de la statistique mathématique est, principalement, d'aider à établir un jugement sur un échantillon. »
> (Philippe Tassi in [13])

Notre point de départ sera une variable aléatoire X dont l'ensemble des valeurs est noté $\mathcal{H}$. Dans ce chapitre, nous supposerons que la variable aléatoire suit une loi $\mathbb{P}_\theta$ où $\theta \in \Theta$. Θ sera appelé l'espace des paramètres.

Exemple 11
Nous savons que le poids des pommes d'un verger peut être modélisé par une loi normale. Si nous connaissons sa variance et ignorons m, nous aurons :
$\Theta =]0; +\infty[\subset \mathbb{R}$.
Notons que m est strictement positif car il s'agit du poids moyen des pommes du verger.
Nous pourrions également ignorer les paramètres m et σ^2.
Dans ce cas, $\Theta =]0; +\infty[\times]0; +\infty[\subset \mathbb{R}^2$.

Définition 13
Un *modèle statistique* sera la donnée de l'ensemble des valeurs que peut prendre la variable aléatoire X noté $\mathcal{H}$, d'un ensemble $\Theta \subset \mathbb{R}^d$ appelé l'ensemble des paramètres et d'une famille $\mathcal{P} := \{\mathbb{P}_\theta | \theta \in \Theta\}$ de lois indexée par Θ.

Définition 14
Un échantillon de taille n est la donnée de n variables aléatoires indépendantes et identiquement distribuées (i.i.d.) qui suivent la même loi que la variable aléatoire X.

Définition 15
Un *estimateur* est une fonction (mesurable) de $\hat{\theta} : \mathcal{H}^n \longrightarrow \mathbb{R}^d$.
Plus généralement, une *statistique* est une fonction (mesurable) définie sur $\mathcal{H}^n$.

Nous voudrons estimer $g(\theta)$ pour une certaine fonction g.
Par exemple, nous pourrons estimer $E(X) = \dfrac{1}{\lambda}$ pour la loi exponentielle.
Fondamentalement, un estimateur sera conçu comme une fonction des observations à valeurs dans $\mathbb{R}^d$.

Définition 16
Le biais d'un estimateur $\hat{\theta}$ de $g(\theta)$ est défini par : $b(\hat{\theta}) := E(\hat{\theta}) - g(\theta)$.

Lorsque $g(\theta) = \theta$, le biais est simplement $b(\hat{\theta}) = E(\hat{\theta}) - \theta$.

Définition 17
L'estimateur $\hat{\theta}$ dit sans biais lorsque $b(\hat{\theta}) = 0$.

Exemple 12
La moyenne empirique $\bar{X} := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ (parfois notée $\bar{X}_n$) est un estimateur sans biais de l'espérance de X.

Soit $(X_1, \dots, X_n)$ un échantillon de taille n. Les variables aléatoires X_i suivent la même loi que la variable aléatoire X donc $E(X_i) = E(X)$.
En effet, par linéarité de l'espérance,

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} \times nE(X) = E(X)$$

car $E(X_i) = E(X)$ pour tout i.
Ainsi, pour la loi normale $\bar{X}$ est un estimateur sans biais du paramètre m et pour la loi géométrique $\bar{X}$ est un estimateur sans biais de $\dfrac{1}{p}$.

Exemple 13
Par contre, la variance empirique $S'^2 := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$ est biaisée en tant qu'estimateur de la variance.
En effet, $E(S'^2) = \dfrac{n-1}{n}\sigma^2$ où σ^2 désigne la variance de la variable aléatoire X [16].
Lorsque, nous voudrons considérer un estimateur sans biais de la variance, nous aurons recours à la variance empirique modifiée (corrigée) $S^2 := \dfrac{n}{n-1}S'^2$ qui est clairement sans biais car
$E(S^2) = E\left(\dfrac{n}{n-1}S'^2\right) = \dfrac{n}{n-1} \times \dfrac{n-1}{n}\sigma^2 = \sigma^2.$

En dehors des contextes où nous étudierons la convergence d'un estimateur, nous noterons simplement $\hat{\theta}$ comme précédemment mais lorsque nous aborderons la convergence, il sera pertinent d'indiquer l'indice n comme dans la définition suivante.

Définition 18
L'estimateur $\hat{\theta}_n$ est dit *asymptotiquement sans biais* si $\lim\limits_{n \longrightarrow +\infty} b(\hat{\theta}_n) = 0$.

Nous voyons aisément que $S_n'^2$ est asymptotiquement sans biais.
L'espérance d'un estimateur sans biais de $g(\theta)$ est $g(\theta)$ donc sa variance est l'écart en moyenne quadratique (ou la distance L^2) entre l'estimateur et la valeur estimée d'où la définition suivante :

16. Nous renvoyons le lecteur vers [?] ou [13] pour la démonstration de ce résultat.

Définition 19
Soient $\hat{\theta}^1$ et $\hat{\theta}^2$ deux estimateurs sans biais de $g(\theta)$.
L'estimateur $\hat{\theta}^1$ est *plus efficace* que l'estimateur $\hat{\theta}^2$ si $V(\hat{\theta}^1) < V(\hat{\theta}^2)$ pour tout $\theta \in \Theta$.

Nous définirons l'efficacité d'un estimateur en fin de chapitre.

Définition 20

1. Un estimateur $\hat{\theta}_n$ est dit *convergent* si $\hat{\theta}_n$ *converge en probabilité* vers θ.

2. Un estimateur $\hat{\theta}_n$ est dit *fortement convergent* si $\hat{\theta}_n$ *converge presque sûrement* vers θ.

La convergence presque sûre est beaucoup plus forte que la convergence en probabilité. La convergence presque sûre, dans ce contexte, signifie que $\hat{\theta}(X(\omega))$ converge vers θ (sauf éventuellement sur un sous-ensemble de mesure nulle). En dehors de cet éventuel sous-ensemble, il y a convergence de $\hat{\theta}(X(\omega))$ vers θ pour tout ω.
Par contre, la convergence en probabilité signifie simplement, pour tout $\varepsilon > 0$, la probabilité de ne pas appartenir à l'intervalle $[\theta-\varepsilon;\theta+\varepsilon]$ peut être rendue aussi petite que nous le souhaitons pour n assez grand.

Théorème 19
Soit X une variable aléatoire de carré intégrable i.e. telle que $E(X^2) < +\infty$.

1. $\bar{X}$ est un estimateur fortement convergent de l'espérance.

2. S^2 et S'^2 sont des estimateurs fortement convergents de la variance.

Ce théorème est une conséquence directe de la loi forte des grands nombres.

2.8 Estimation paramétrique

2.8.1 Méthode des moments

Elle peut être exprimée à l'aide d'équations mais nous nous contenterons d'une présentation plus intuitive. Nous exprimerons le paramètre que nous voudrons estimer en fonction des moments $m_k := E(X^k)$ ou des moments centrés $m'_k := E((X - E(X))^k)$ notamment $m'_2 = V(X)$. Ensuite, nous remplacerons ces moments par les moments empiriques correspondants.

Exemple 14
Pour la loi géométrique, $E(X) = \frac{1}{p}$ donc $p = \frac{1}{E(X)}$.
A l'ordre 1, la méthode des moments nous donne l'estimateur $\hat{p} := \frac{1}{\bar{X}}$.

Exemple 15
Pour la loi Gamma[17] de paramètres α et β, il faudra résoudre un système :

$$\begin{cases} E(X) = \frac{\alpha}{\beta} \\ V(X) = \frac{\alpha}{\beta^2} \end{cases} \tag{2.1}$$

donc

$$\begin{cases} E(X) = \frac{\alpha}{\beta} \\ \frac{E(X)}{V(X)} = \frac{\frac{\alpha}{\beta}}{\frac{\alpha}{\beta^2}} = \beta \end{cases} \tag{2.2}$$

d'où

$$\begin{cases} \alpha = \frac{E(X)^2}{V(X)} \\ \beta = \frac{E(X)}{V(X)} \end{cases} \tag{2.3}$$

Nous en déduisons les estimateurs donnés par la méthode des moments : $\hat{\alpha} = \frac{\bar{X}^2}{S^2}$ et $\beta = \frac{\bar{X}}{S^2}$

2.8.2 Méthode du maximum de vraisemblance

> « C'est une vérité très certaine que si nous ne pouvons pas déterminer ce qui est vrai, nous devons suivre ce qui est le plus probable. »
> (Descartes)

La loi dépend du paramètre θ donc la densité associée aussi, nous la noterons $f(x, \theta)$.

17. cf. Complément sur la loi Gamma.

Nous supposerons $f(x, \theta) > 0$ pour tout x et tout θ.
Nous définirons sur $\mathcal{H}^n \times \Theta$ la *fonction de vraisemblance* par :

$$L(x_1, \ldots, x_n, \theta) := \prod_{i=1}^{n} f(x_i, \theta)$$

Pour obtenir le paramètre le plus probable à partir des observations $(x_1, \ldots, x_n)$, nous maximiserons sur Θ la fonction $L(x_1, \ldots, x_n, .)$ i.e. nous résoudrons le problème de maximisation suivant :

$$\left\{ \begin{array}{l} \max L(x_1, \ldots, x_n, \theta) \\ \theta \in \Theta \end{array} \right. \tag{2.4}$$

Supposons que la densité soit de classe $\mathcal{C}^2$ par rapport à θ (i.e. deux fois dérivable par rapport à θ et dont la dérivée seconde par rapport à θ est continue).
Le problème est un problème d'optimisation différentiable. Nous pourrons ainsi considérer le système suivant :

$$\left\{ \begin{array}{l} \dfrac{\partial L}{\partial \theta} = 0 \\ \dfrac{\partial^2 L}{\partial \theta^2} < 0 \end{array} \right. \tag{2.5}$$

La première condition est une condition nécessaire et la seconde est une condition suffisante. En pratique, la première condition nous permettra d'identifier des valeurs de θ susceptibles d'être des solutions.

Pour simplifier les calculs, nous remplacerons souvent la vraisemblance par la log-vraisemblance car la fonction logarithme est de classe $\mathcal{C}^2$ et strictement croissante sur $]0; +\infty[$.
Le système précédent est équivalent au système :

$$\left\{ \begin{array}{l} \dfrac{\partial \log L}{\partial \theta} = 0 \\ \dfrac{\partial^2 \log L}{\partial \theta^2} < 0 \end{array} \right. \tag{2.6}$$

L'équation $\dfrac{\partial \log L}{\partial \theta} = 0$ s'appelle l'*équation de vraisemblance.*
Rappelons quelques propriétés du logarithme qui nous seront utiles. Soient a et b des réels strictement positifs et x un réel, on a :

1. $\log(ab) = \log a + \log b$
2. $\log\left(\frac{a}{b}\right) = \log a - \log b$
3. $\log a^x = x \log a$

Exemple 16
Déterminons l'estimateur du maximum de vraisemblance pour la loi exponentielle de paramètre λ.
Par définition, $\mathcal{H} = [0; +\infty[$ et $\Theta =]0; +\infty[$.
On sait que : $f(x, \lambda) = \lambda e^{-\lambda x}$.
Remarquons que$f(x, \lambda) > 0$ pour tout x et tout λ.
Soit n un entier naturel supérieur ou égal à 1.
Par définition, $L(x, \lambda) = \prod_{k=1}^{n} \lambda e^{-\lambda x_k} = \lambda^n e^{-\lambda \sum_{k=1}^{n} x_k}$.
Passons au logarithme :

$$\log L(x_1, \ldots, x_n, \lambda) = n \log \lambda - \lambda \sum_{k=1}^{n} x_k$$

Ecrivons les conditions du premier et second ordre :

$$\left\{ \begin{array}{l} \dfrac{\partial \log L}{\partial \lambda} = 0 \\ \dfrac{\partial^2 \log L}{\partial \lambda^2} < 0 \end{array} \right. \tag{2.7}$$

On a : $\dfrac{\partial L}{\partial \lambda}(x, \lambda) = \dfrac{n}{\lambda} - \sum_{k=1}^{n} x_k$.

La condition nécessaire est que la solution $\hat{\lambda}(x)$ doit vérifier :

$$\frac{n}{\lambda} - \sum_{k=1}^{n} x_k = 0$$

D'où : $\hat{\lambda}(x) = \dfrac{n}{\sum_{k=1}^{n} x_k}$.

Pour conclure qu'il s'agit bien d'un maximum, vérifions la condition suffisante : $\dfrac{\partial^2 \log L}{\partial \lambda^2}(x_1, \ldots, x_n, \lambda) = -\dfrac{n}{\lambda^2} < 0$ pour tout λ donc a fortiori pour $\hat{\lambda}(x)$.

La condition suffisante est satisfaite donc $\hat{\lambda}(x)$ est bien le maximum de vraisemblance.

Nous en déduisons l'estimateur du maximum de vraisemblance :
$\hat{\lambda}_n = \dfrac{n}{\sum_{k=1}^{n} X_k} = \dfrac{1}{\bar{X}_n}$. Notons qu'il est *fortement convergent* d'après la loi forte des grands nombres en notant que la fonction inverse est continue.

Etudions un autre exemple qui concernera une loi discrète.
Naturellement, la fonction de densité sera remplacée par $P(X = x)$.

Exemple 17
Déterminons l'estimateur du maximum de vraisemblance pour la loi de Poisson de paramètre λ. Ici, on a : $\mathcal{H} = \mathbb{N}$ et $\Theta =]0; +\infty[$.
On sait que : $P(X = x) = e^{-\lambda}\dfrac{\lambda^x}{x!} > 0$ pour $x \in \mathbb{N}$ et $\lambda \in]0; +\infty[$.
Par définition, $L(x_1, \ldots, x_n, \lambda) = \displaystyle\prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$.
Passons au logarithme :
$\log L(x_1, \ldots, x_n, \lambda) = -n\lambda + \displaystyle\sum_{i=1}^{n} x_i \log(\lambda) - \log \prod_{i=1}^{n} x_i!$.
Dérivons cette fonction par rapport à λ, on a :

$$\frac{\partial \log L}{\partial \lambda}(x_1, \ldots, x_n, \lambda) = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda}$$

La condition nécessaire est que $\hat{\lambda}(x)$ doit être solution de :

$$-n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$

ce qui nous donne : $\hat{\lambda}(x) = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} x_k$.
Pour conclure, nous devons vérifier que la **condition suffisante** est satisfaite :

$$\frac{\partial^2 \log L}{\partial \lambda^2}(x_1, \ldots, x_n, \lambda) = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} < 0$$

Ceci est vrai pour tout λ donc a fortiori pour $\hat{\lambda}(x)$.
La condition suffisante est satisfaite donc $\hat{\lambda}(x)$ est bien le maximum de vraisemblance.

$\hat{\lambda}_n = \frac{1}{n}\sum_{k=1}^{n} X_k = \bar{X}_n$ est l'EMV et nous savons qu'il est *fortement convergent* d'après la loi forte des grands nombres.

Exemple 18
Déterminons l'estimateur du maximum de vraisemblance du couple (m, σ^2) pour la loi normale $\mathcal{N}(m, \sigma^2)$.
La densité associée à cette loi est :

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

donc

$$L(x_1, \ldots, x_n, m, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-m)^2}{2\sigma^2}\right) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(\frac{-\sum_{i=1}^{n}(x_i-m)^2}{2\sigma^2}\right)$$

Passons au logarithme

$$\log L(x_1, \ldots, x_n, m, \sigma^2) = -n\log\sigma - n\log\sqrt{2\pi} - \sum_{i=1}^{n}\frac{(x_i-m)^2}{2\sigma^2}$$

ou encore

$$\log L(x_1, \ldots, x_n, m, \sigma^2) = -\frac{n}{2}\log\sigma^2 - n\log\sqrt{2\pi} - \sum_{i=1}^{n}\frac{(x_i-m)^2}{2\sigma^2}$$

car $\log(a^x) = x\log a$ donc $\log\sigma = \log((\sigma^2)^{\frac{1}{2}}) = \frac{1}{2}\log\sigma$.
Comme nous l'avons déjà vu (cf. correction des exercices), en dérivant par rapport à m, on obtient :

$$\hat{m} := \bar{X}$$

Dérivons par rapport à σ^2 (d'où la modification précédente de la log-vraisemblance).

$$\frac{\partial \log L}{\partial\sigma^2}(x_1, \ldots, x_n, m, \sigma^2) = -\frac{n}{2\sigma^2} + \sum_{i=1}^{n}\frac{(x_i-m)^2}{2\sigma^4}$$

Donc $\frac{\partial \log L}{\partial\sigma^2}(x_1, \ldots, x_n, m, \sigma^2) = 0$ donne

$$\frac{n}{2\sigma^2} = \sum_{i=1}^{n}\frac{(x_i-m)^2}{2\sigma^4}$$

d'où

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{m})^2$$

Il faut naturellement remplacer m par la solution trouvée en dérivant par rapport à m.
Vérifions la condition du second ordre.

$$\frac{\partial^2 \log L}{\partial(\sigma^2)^2} = \frac{n}{2\sigma^4} - \sum_{i=1}^{n}\frac{(x_i - m)^2}{\sigma^6}$$

En $(\hat{m}, \hat{\sigma^2})$, on veut savoir si la condition suivante est vraie :

$$\frac{n}{2\hat{\sigma}^4} - \sum_{i=1}^{n}\frac{(x_i - \hat{m})^2}{\hat{\sigma}^6} < 0$$

qui se ramène à

$$\frac{\hat{\sigma}^6}{2\hat{\sigma}^4} < \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{m})^2$$

Or $\hat{\sigma}^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \hat{m})^2$ donc la condition se ramène à

$$\frac{1}{2}\hat{\sigma}^2 < \hat{\sigma}^2$$

qui est vraie car $\hat{\sigma}^2 > 0$.
En conclusion, $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S'^2$ est l'estimateur du maximum de vraisemblance de la variance.
Nous avons vu précédemment l'égalité $E(S'^2) = \dfrac{n-1}{n}\sigma^2$.
L'estimateur $\hat{\sigma}^2 = S'^2$ est biaisé tandis que l'estimateur $\hat{m} = \bar{X}$ est sans biais. *Lorsque nous estimerons une distribution gaussienne par la méthode du maximum de vraisemblance, l'espérance sera correctement estimée tandis que la variance sera systématiquement sous-estimée.* Naturellement, à mesure que la valeur de n augmentera, le biais deviendra moins important.

Proposons un autre exemple classique.

Exemple 19
Déterminer l'estimateur du maximum de vraisemblance pour la loi géométrique de paramètre p.
La fonction de vraisemblance est :

$$L(x_1, \ldots, x_n, p) = \prod_{i=1}^{n} p(1-p)^{x_i - 1}$$

i.e.

$$L(x_1, \ldots, x_n, p) = p^n (1-p)^{\left(\sum_{i=1}^{n} x_i\right) - n}$$

En passant au logarithme (ce qui est licite car les quantités considérées sont strictement positives), nous obtenons :

$$\log L(x_1, \ldots, x_n, p) = n \log(p) + \left[\left(\sum_{i=1}^{n} x_i\right) - n\right] \log(1-p)$$

puis en dérivant :

$$\frac{\partial \log L}{\partial p}(x_1, \ldots, x_n, p) = \frac{n}{p} - \left[\left(\sum_{i=1}^{n} x_i\right) - n\right] \frac{1}{1-p}$$

Ecrivons l'équation de log-vraisemblance.

$$\frac{\partial \log L}{\partial p}(x_1, \ldots, x_n, p) = 0 \Leftrightarrow \frac{n}{p} - \left[\left(\sum_{i=1}^{n} x_i\right) - n\right] \frac{1}{1-p} = 0$$

En multipliant par $p(1-p)$ puis, en simplifiant, nous obtenons :

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} x_i}$$

Vérifions que la condition suffisante est satisfaite.

$$\frac{\partial^2 \log L}{\partial p^2}(x_1, \ldots, x_n, p) = -\frac{n}{p^2} - \left[(\sum_{i=1}^{n} x_i) - n\right] \frac{1}{(1-p)^2}$$

En remarquant que[18], pour tout $i \in \{1, \ldots, n\}$, $x_i \geq 1$, nous déduisons : $\sum_{i=1}^{n} x_i - n \geq 0$ et $\dfrac{\partial^2 \log L}{\partial p^2}(x_1, \ldots, x_n, p) < 0$.

18. La loi géométrique ne charge que les valeurs entières strictement positives i.e. $\mathcal{H} = \mathbb{N} \setminus \{0\}$.

Par conséquent, $\hat{p} = \dfrac{n}{\sum_{i=1}^{n} X_i} = \dfrac{1}{\bar{X}}$ est l'estimateur du maximum de vraisemblance.
Remarquons qu'il s'agit de l'estimateur obtenu à l'aide de la méthode des moments d'ordre 1.

Nous conclurons par un théorème qui assure la convergence de l'estimateur du maximum de vraisemblance vers le paramètre considéré.

Théorème 20
Sous les conditions suivantes :

1. Θ ouvert de $\mathbb{R}$
2. $\mathbb{P}_\theta \neq \mathbb{P}_{\theta'}$ si $\theta \neq \theta'$
3. $f(x, \theta) > 0$ pour tout réel $x \in \mathcal{H}$ et tout réel $\theta \in \theta$
4. $\dfrac{\partial \log L}{\partial \theta}$ existe pour tout réel $x \in \mathcal{H}$ et tout réel $\theta \in \theta$

Alors il existe une suite de solutions $\hat{\theta}_n$ de l'équation de vraisemblance qui converge fortement vers la vraie valeur du paramètre θ_0.

Ce théorème est démontré dans [13]. Nous y renvoyons le lecteur.

2.9 Information de Fisher

Nous souhaitons déterminer la quantité d'information contenue dans une observation ou dans un échantillon. En d'autres termes, nous souhaitons savoir en quoi la connaissance d'une ou de plusieurs observations nous renseignent sur la valeur d'un paramètre d'une loi. Sous les hypothèses suivantes

1. $\forall x \in \mathcal{H}$ et $\forall \theta \in \Theta \subset \mathbb{R}$, $f(x, \theta) > 0$,
2. f est dérivable par rapport à θ.

nous pourrons définir l'*information de Fisher*[19] :

$$I(\theta) := E\left(\frac{f'(X, \theta)}{f(X, \theta)}\right)$$

19. Ce concept est dû comme le maximum de vraisemblance au biologiste et statisticien Ronald Aylmer Fisher.

ou encore la définition équivalente

$$I(\theta) := E\left[\left(\frac{\partial \log f}{\partial \theta}(X,\theta)\right)^2\right]$$

Définition 21
Le *score* est défini par :

$$S(x,\theta) := \frac{\partial \log f}{\partial \theta}(x,\theta)$$

Pour utiliser cette formule, les calculs reposeront en général sur la linéarité de l'espérance, la connaissance des moments des lois et sur les identités remarquables suivantes :

- $(a+b)^2 = a^2 + 2ab + b^2$,
- $(a-b)^2 = a^2 - 2ab + b^2$,

Exemple 20
Pour la loi exponentielle, comme nous l'avons vu, $f(x,\lambda) = \lambda e^{-\lambda x} > 0$.
Donc

$$\log f(x,\lambda) = \log \lambda - \lambda x$$

et

$$\frac{\partial \log f}{\partial \lambda}(x,\lambda) = \frac{1}{\lambda} - x$$

Par conséquent,

$$I(\lambda) = E\left[\left(\frac{\partial \log f}{\partial \lambda}(X,\lambda)\right)^2\right] = E\left[\left(\frac{1}{\lambda} - X\right)^2\right]$$

D'où

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

et

$$I(\lambda) = \frac{1}{\lambda^2} - \frac{2}{\lambda^2} + \frac{2}{\lambda^2} = \boxed{\frac{1}{\lambda^2}}$$

Exemple 21
De même, nous calculerons $I(p)$ pour la loi binomiale de paramètres n et p.
On sait que $f(x,p) = \binom{n}{x} p^x(1-p)^{n-x} > 0$. Donc

$$\log f(x,p) = \log \binom{n}{x} + x\log p + (n-x)\log(1-p)$$

et

$$\frac{\partial \log f}{\partial p}(x,p) = \frac{x}{p} - \frac{n-x}{1-p} = \frac{x-np}{p(1-p)}$$

Par conséquent,

$$I(p) = E\left[\left(\frac{\partial \log f}{\partial p}(X,p)\right)^2\right] = \left[E\left(\frac{X-np}{p(1-p)}\right)^2\right]$$

donc

$$I(p) = \frac{1}{p^2(1-p)^2}\left[E(X^2) - 2npE(X) + n^2p^2\right]$$

Or $E(X) = np$ et $E(x^2) = n(n-1)p^2 + np$ d'après le chapitre 1.
Donc

$$I(p) = \frac{1}{p^2(1-p)^2}(n(n-1)p^2 + np - 2n^2p^2 + n^2p^2)$$

et

$$I(p) = \frac{np(1-p)}{p^2(1-p)^2} = \boxed{\frac{n}{p(1-p)}}$$

Nous privilégierons une autre méthode de calcul.

Théorème 21
Sous réserve que f soit deux fois dérivable en θ, $I(\theta) = -E\left(\frac{\partial^2 \log f}{\partial \theta^2}(X,\theta)\right)$.

Exemple 22
Revenons à la loi binomiale de paramètres n et p.
On a : $\frac{\partial \log f}{\partial p}(x,p) = \frac{x}{p} - \frac{n-x}{1-p} = \frac{x-np}{p(1-p)}$.
Par la dérivée d'un quotient [20] :

$$\frac{\partial^2 \log f}{\partial p^2}(x,p) = -\frac{-np(1-p) - (x-np)(1-2p)}{p^2(1-p)^2}$$

20. $\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$

Après simplification, nous obtenons :

$$\frac{\partial^2 \log f}{\partial p^2}(x,p) = \frac{(2p-1)x - 2np^2}{p^2(1-p)^2}$$

Par conséquent

$$I(p) = -E\left(\frac{\partial^2 \log f}{\partial p^2}(x,p)\right) = -\frac{1}{p^2(1-p)^2}\left[(2p-1)E(X) - np^2\right]$$

Donc

$$I(p) = -\frac{1}{p^2(1-p)^2}(2p-1)np - np^2)$$

et

$$I(p) = \frac{np(1-p)}{p^2(1-p)^2} = \boxed{\frac{n}{p(1-p)}}$$

Exemple 23
Pour la loi normale de paramètre m et σ^2 avec σ^2 fixé et connu.
On a : $f(x,m) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\dfrac{(x-m)^2}{2\sigma^2}\right)$.
Donc $\log f(x,m) = -\dfrac{1}{2}\log(2\pi\sigma^2) - \dfrac{(x-m)^2}{2\sigma^2}$
et $\dfrac{\partial \log f}{\partial m}(x,m) = -\dfrac{-2x+2m}{2\sigma^2} = \dfrac{-m+x}{\sigma^2}$.
D'où

$$\frac{\partial^2 \log f}{\partial m^2}(x,m) = \frac{-1}{\sigma^2}$$

En conclusion,

$$I(m) = -E\left(\frac{\partial^2 \log f}{\partial m^2}(X,m)\right) = \frac{1}{\sigma^2}$$

Exemple 24
Pour la loi exponentielle de paramètre λ, nous pouvons la calculer de nouveau.
On a : $f(x,\lambda) = \lambda e^{-\lambda x}$.
Donc $\log f(x,\lambda) = \log\lambda - \lambda x$.
D'où $\dfrac{\partial \log f}{\partial \lambda}(x,\lambda) = \dfrac{1}{\lambda} - x$.
Par conséquent,

$$\frac{\partial^2 \log f}{\partial \lambda^2}(x,\lambda) = -\frac{1}{\lambda^2}$$

En conclusion,

$$I(\lambda) = -E\left(\frac{\partial^2 \log f}{\partial \lambda^2}(X, \lambda)\right) = \frac{1}{\lambda^2}$$

Voici les informations usuelles :

- loi de Poisson $I(\lambda) = \dfrac{1}{\lambda}$
- loi binomiale $I(p) = \dfrac{n}{p(1-p)}$
- loi normale $I(m) = \dfrac{1}{\sigma^2}$
- loi normale $I(\sigma^2) = \dfrac{1}{2\sigma^4}$
- loi exponentielle $I(\lambda) = \dfrac{1}{\lambda^2}$.

Lorsque des variables aléatoires X_1,...,X_n sont **indépendantes**, l'information de Fisher associée au n-uplet $(X_1, \ldots, X_n)$ est :

$$\boxed{I_n(\theta) := I_{(X_1,\ldots,X_n)}(\theta) = nI(\theta)}$$

où $I(\theta) := I_{X_1}(\theta)$.

Exemple 25
Par exemple, pour la loi normale,

$$I_n(m) = \frac{n}{\sigma^2}$$

2.10 Inégalité FDCR

Nous étudierons dans cette partie l'inégalité de Fréchet-Darmois-Cramer-Rao.
Comme nous l'avons vu, pour éviter la sous-estimation ou la sur-estimation des paramètres, nous privilégierons des estimateurs sans biais.
De plus, nous souhaitons déterminer parmi eux ceux dont la variance sera la plus faible car la variance correspondra à la distance (dans l'espace fonctionnel des variables aléatoires de carré intégrable) entre l'estimateur et θ (vu

comme une fonction constante de carré intégrable).
L'inégalité FDCR fournira une borne inférieure à la variance d'un estimateur sans biais.
Nous notons, à titre culturel, les hypothèses de Cramer-Rao. Elles seront toujours vérifiées dans les exercices que nous traiterons.
Hypothèses de Cramer-Rao [21]

- H_1 : Θ est un ouvert de $\mathbb{R}$ et $f(x,\theta) > 0$ pour tout x et tout θ.
- H_2 : $\dfrac{\partial f}{\partial \theta}(x,\theta)$ et $\dfrac{\partial^2 f}{\partial \theta^2}(x,\theta)$ existent et sont finies presque sûrement et pour tout θ.
- H_3 : Pour tout A (mesurable), $\int_A f(x,\theta)dx$ est dérivable au moins deux fois sous le signe $\int$ et

$$\frac{\partial}{\partial \theta}\int_A f(x,\theta)dx = \int_A \frac{\partial f}{\partial \theta}(x,\theta)dx$$

- H_4 : pour tout θ, $0 < I(\theta) < +\infty$.

Les hypothèses sont d'ordre technique et assurent notamment la licéité de la dérivation sous le signe $\int$.

Théorème 22
Sous les hypothèses de Cramer-Rao. Soit T un estimateur sans biais de $g(\theta)$ de variance finie.
Sous l'hypothèse supplémentaire :
$H5$: Pour tout A (mesurable), $\int_A T(x)f(x,\theta)dx$ est dérivable par rapport à θ et $\int_A |T(x)|\dfrac{\partial f}{\partial \theta}(x,\theta)dx$.
On a :

- $g(\theta)$ est dérivable.
- pour tout θ,

$$\boxed{V(T) \geq \frac{g'^2(\theta)}{I_n(\theta)}}$$

21. Nous renvoyons le lecteur intéressé à [13].

La borne $\dfrac{g'^2(\theta)}{I_n(\theta)}$ est appelée *borne de Fréchet* ou borne de *Cramer-Rao*.
Lorsque nous étudierons des estimateurs sans biais du paramètre θ, $g(\theta) = \theta$ et donc la borne de Fréchet sera $\dfrac{1}{I_n(\theta)}$.

Définition 22
Un estimateur sans biais T de $g(\theta)$ est dit *efficace* lorsque la borne de Fréchet est atteinte i.e.

$$V(T) = \frac{g'^2(\theta)}{I_n(\theta)}$$

Exemple 26
Avant d'étudier la moyenne empirique, rappelons deux propriétés importantes sur la variance.
soient X et Y deux variables aléatoires indépendantes et a un réel.
On a :

- $V(aX) = a^2V(X)$
- $V(X + Y) = V(X) + V(Y)$

La preuve de ces propriétés est laissée en exercice. La première propriété procède de la définition de la variance et la seconde procède du fait que $E(XY) = E(X)E(Y)$ pour des variables aléatoires X et Y indépendantes.
Nous savons que la moyenne empirique $\bar{X}$ est un estimateur sans biais de l'espérance pour toutes les lois (qui admettent une espérance).
Pour la loi normale, $V(\bar{X}) = V\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i\right) = \dfrac{1}{n^2}\displaystyle\sum_{i=1}^{n} V(X_i)$ car les variables aléatoires X_i sont indépendantes.
$V(\bar{X}) = \dfrac{n}{n^2}V(X_1) = \dfrac{\sigma^2}{n}$ car les variables aléatoires X_i sont de même loi.
D'après ce qui précède, $I_n(m) = \dfrac{n}{\sigma^2}$.
Donc

$$V(\bar{X}) = \frac{1}{I_n(m)}$$

La moyenne empirique est donc un estimateur efficace du paramètre m.

Nous verrons dans les compléments et en exercices comment traiter les estimateurs biaisés et nous étudierons d'autres méthodes d'estimation.

Ce qu'il faut retenir

- Notion d'estimateur.
- Estimateur sans biais, estimateur (fortement) convergent.
- Application de la méthode des moments.
- Détermination de l'estimateur du maximum de vraisemblance.
- Calcul de l'information de Fisher.
- Il est conseillé de connaître les informations de Fisher usuelles ou de savoir aisément les retrouver.
- Inégalité FDCR, notion d'estimateur efficace (plus théorique, moins important pour l'examen).

Chapitre 3

Intervalles de confiance

3.1 Cadre général

J'ai encore une fois oublié que le hasard est le résultat d'une immense équation dont nous ne connaissons pas toutes les racines.
(Honoré de Balzac in *Z. Marcas*)

On définit la « qualité » d'une estimation en lui associant un intervalle ayant une forte probabilité de contenir la vraie valeur du paramètre[...].
(André Vessereau in [**?**])

Lors du chapitre précédent, nous avons étudié l'estimation ponctuelle d'un paramètre θ (ou plus généralement de $g(\theta)$). En revanche, nous ne disposons pas d'éléments pour quantifier la qualité d'une estimation.
Rappelons deux définitions du chapitre précédent.

Définition 23
Un *modèle statistique* est la donnée de l'ensemble des valeurs que peut prendre la variable aléatoire X noté $\mathcal{H}$, d'un ensemble $\Theta \subset \mathbb{R}^d$ appelé l'ensemble des paramètres et d'une famille $\mathcal{P} := \{\mathbb{P}_\theta | \theta \in \Theta\}$ de lois indexée par Θ.

Définition 24
Un échantillon de taille n est la donnée de n variables aléatoires indépendantes et identiquement distribuées (i.i.d.) qui suivent la même loi que la variable aléatoire X.

Ce chapitre reposera sur la notion de *fonction pivotale* (*asymptotiquement pivotale* pour les intervalles de confiance asymptotiques).
Une *fonction pivotale* est une fonction Ψ définie sur $\mathcal{H}^n \times \Theta$ telle que $\Psi(X_1, \ldots, X_n, \theta)$ suit une loi « indépendante » du paramètre θ. Nous pourrons ainsi encadrer $\Psi(X_1, \ldots, X_n, \theta)$ à l'aide de fractiles[1] de cette loi et en déduire un encadrement de θ qui sera fonction de l'échantillon.

Exemple 27
Pour un échantillon $(X_1, \ldots, X_n)$ de loi normale $\mathcal{N}(m, \sigma^2)$ avec σ^2 connu, $\Psi(X_1, \ldots, X_n, m) := \dfrac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$ suit une loi normale centrée réduite donc la fonction Ψ est une fonction pivotale dans ce cas.

Nous exploiterons différentes fonctions pivotales dans ce chapitre sans les expliciter.

3.2 Intervalle de confiance pour la moyenne m

3.2.1 Cas 1 : Variance connue

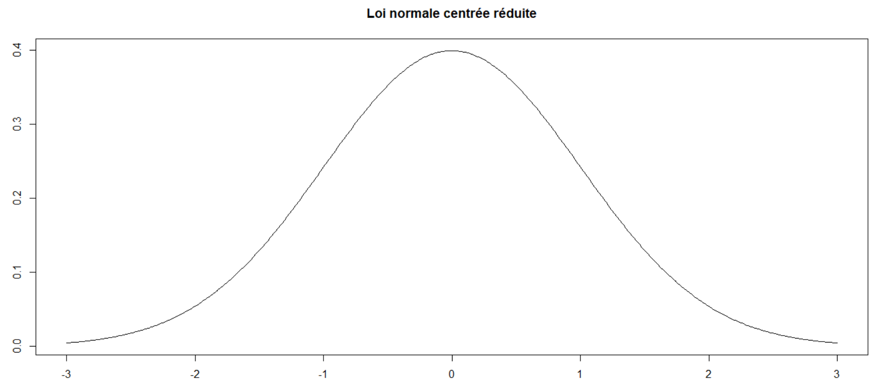
Rappelons que nous nous placerons dans le cas gaussien, i.e. nous chercherons à estimer la moyenne m d'une variable aléatoire X dont nous connaissons la variance σ^2(et donc l'écart-type σ). La proposition suivante, vue dans le chapitre 1, nous sera d'une grande utilité :

Proposition 40
Soit X une variable aléatoire suivant une loi normale centrée réduite ($\mathcal{N}(0, 1)$).

1. $\mathbb{P}(X \leq 0) = \mathbb{P}(X \geq 0) = 0,5$
2. Plus généralement, pour tout réel a, $\mathbb{P}(X \leq -a) = \mathbb{P}(X \geq a)$
3. $\mathbb{P}(-1,96 \leq X \leq 1,96) \approx 0,95$ et $\mathbb{P}(-2,58 \leq X \leq 2,58) \approx 0,99$.

1. Soit X une variable aléatoire. Un *fractile* d'ordre α de la loi de X est un réel u_α tel que $\mathbb{P}(u_\alpha \leq X) = \alpha$.

Rappelons également que la loi normale est symétrique par rapport à sa moyenne m et donc, que la loi normale centrée réduite l'est par rapport à 0 comme l'indique l'illustre la proposition précédente.



Soit $(X_1, \dots, X_n)$ un échantillon de taille $n \geq 1$ de la loi $\mathcal{N}(m, \sigma^2)$ i.e. la donnée de n variables aléatoires indépendantes suivant cette loi.

Les variables aléatoires X_i sont indépendantes donc $\sum_{i=1}^{n} X_i$ suit une loi $\mathcal{N}(nm, n\sigma^2)$ et $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ suit une loi $\mathcal{N}(m, \frac{\sigma^2}{n})$.

Ceci repose sur le résultat suivant bien connu [2] :

Proposition 41

1. Soient X_1 et X_2 deux variables aléatoires indépendantes suivant respectivement des lois $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$ alors la variable aléatoire $X_1 + X_2$ suit une loi $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

2. Soit X une variable aléatoire suivant une loi $\mathcal{N}(m, \sigma^2)$ alors la variable aléatoire $Y = X - m$ suit une loi $\mathcal{N}(0, \sigma^2)$.

3. Soit X une variable aléatoire suivant une loi $\mathcal{N}(m, \sigma^2)$ et a un réel différent de 0, la variable aléatoire aX suit une loi $\mathcal{N}(am, a^2\sigma^2)$.

D'après le deuxième point, $\bar{X}_n - m$ suit une loi $\mathcal{N}(0, \frac{\sigma^2}{n})$.

Donc, en utilisant le troisième point, $\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$ suit une loi normale

2. Nous avons démontré le premier point lors du chapitre 1 à l'aide de la fonction caractéristique. Les autres points peuvent être obtenus de manière similaire.

centrée réduite.
Ex supra,

$$\mathbb{P}\left(-1,96 \leq \frac{\sqrt{n}(\bar{X}_n - m)}{\sigma} \leq 1,96\right) \approx 0,95$$

Notons que ceci ne dépend pas de la valeur de n.
Par conséquent, le réel m que nous cherchons à estimer vérifie :

$$m \in \left[\bar{X}_n - 1,96\frac{\sigma}{\sqrt{n}}; \bar{X}_n + 1,96\frac{\sigma}{\sqrt{n}}\right]$$

au niveau de confiance 0,95, au sens où il appartient à cet intervalle avec une probabilité égale à 0,95. En d'autres termes, il appartient à cet intervalle pour 95% des échantillons.
En considérant les réalisations d'un échantillon $(X_1, \ldots, X_n)$, c'est à dire des observations $(x_1, \ldots, x_n)$ dont nous noterons la moyenne $\bar{x}_n$, nous pourrons définir l'intervalle de confiance au niveau de confiance $0,95$ pour la moyenne m.

Définition 25
L'*intervalle de confiance* au niveau $0,95$ pour la moyenne m est :

$$\left[\bar{x}_n - 1,96\frac{\sigma}{\sqrt{n}}; \bar{x}_n + 1,96\frac{\sigma}{\sqrt{n}}\right].$$

De même, pour le niveau de confiance 0,99, nous utiliserons le dernier point de la proposition 1.

Définition 26
L'*intervalle de confiance* au niveau $0,99$ pour la moyenne m est :

$$\left[\bar{x}_n - 2,58\frac{\sigma}{\sqrt{n}}; \bar{x}_n + 2,58\frac{\sigma}{\sqrt{n}}\right].$$

Plus généralement en notant $z_{1-\frac{\alpha}{2}}$ le fractile d'ordre $1 - \dfrac{\alpha}{2}$ de la loi $\mathcal{N}(0,1)$ [3], nous pouvons définir :

3. Nous laissons $\frac{\alpha}{2}$ à gauche de l'intervalle et $\frac{\alpha}{2}$ à droite de l'intervalle. La forme de l'intervalle est simple car la loi normale centrée réduite est symétrique par rapport à 0.

Définition 27
L'intervalle de confiance au niveau de confiance $1-\alpha$ pour la moyenne m est :
$$\left[\bar{x}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \bar{x}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right].$$

Exemple 28
Proposer un intervalle de confiance au niveau 0,90 pour la moyenne m pour une variable aléatoire gaussienne de variance 2 dont nous connaissons les observations suivantes : 3,1 ; 2,4 ; 5 ; 7 et 2,8.

Calculons tout d'abord $\bar{x}$. On a :
$$\bar{x} = \frac{3,1+2,4+5+7+2,8}{5} = 4,06$$
Nous aurons recours au fractile $z_{0,95} \approx 1,64$.
L'intervalle de confiance au niveau 0,90 pour la moyenne est :
$$\left[4,06 - 1,64\frac{\sqrt{2}}{\sqrt{5}}; 4,06 + 1,64\frac{\sqrt{2}}{\sqrt{5}}\right]$$
i.e. approximativement l'intervalle $[3,02; 5,10]$.

3.2.2 Cas 2 : Variance inconnue

Nous aurons besoin d'introduire la loi du Khi-deux et la loi de Student.

Définition 28
Soient $(X_1, \ldots, X_n)$ la donnée de n variables aléatoires indépendantes normales centrées réduites.
La variable aléatoire $U_n := \sum_{i=1}^{n} X_i^2$ suit une loi du Khi-deux à n degrés de liberté notée $\chi^2(n)$.

Enonçons ses propriétés principales dans la proposition suivante :

Proposition 42
1. Sa densité est : $f_{U_n}(x) = \dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} e^{-\frac{x}{2}} x^{\frac{n}{2}-1}$ pour $x \geq 0$,

2. $E(U_n) = n$,

3. $V(U_n) = 2n$,

4. Sa fonction caractéristique est : $\phi_{U_n}(t) = \dfrac{1}{(1-2it)^{\frac{n}{2}}}$.

Le théorème suivant sera également utile :

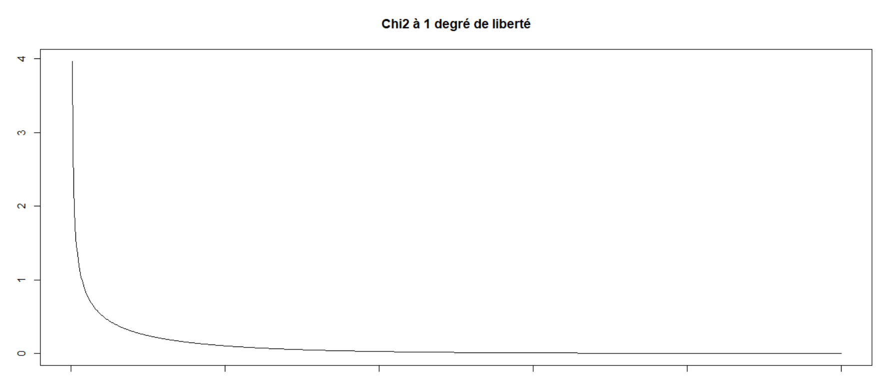
Théorème 23
Soient X et Y deux variables aléatoires indépendantes suivant respectivement des lois $\chi^2(m)$ et $\chi^2(n)$ alors la variable aléatoire $X+Y$ suit une loi $\chi^2(m+n)$.

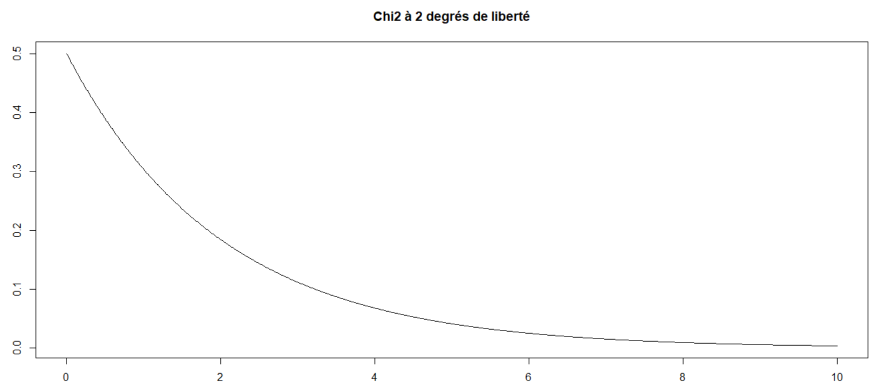
Preuve
Les variables aléatoires X et Y sont indépendantes donc :

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) = \frac{1}{(1-2it)^{\frac{m}{2}}} \times \frac{1}{(1-2it)^{\frac{n}{2}}} = \frac{1}{(1-2it)^{\frac{m+n}{2}}}$$

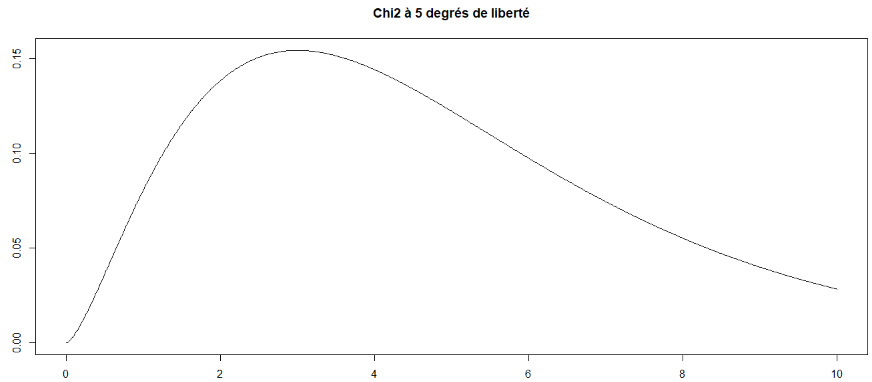
d'où le résultat.

Voici les représentations graphiques pour 1 et 2 degrés de liberté.




Chi2 à 2 degrés de liberté
0.0
0.1
0.2
0.3
0.4
0.5
0
2
4
6
8
10

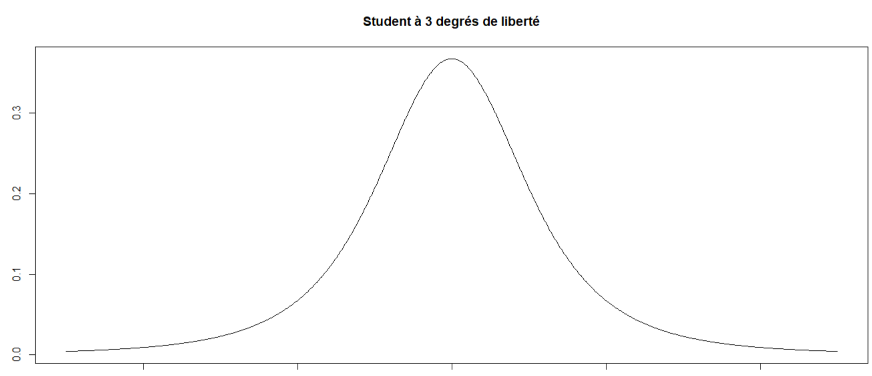
Pour $n > 2$, la forme est la suivante :

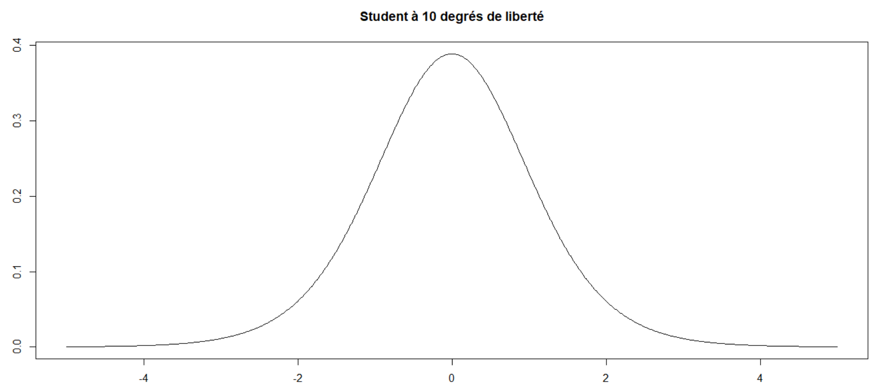


Définition 29
Soient X et Y deux variables aléatoires indépendantes suivant respectivement des lois $\mathcal{N}(0,1)$ et $\chi^2(n)$.
La variable $T_n = \dfrac{X}{\sqrt{\frac{Y}{n}}}$ suit une loi de Student à n degrés de liberté. Elle est notée $\mathcal{T}_n$.

Les propriétés que nous utiliserons sont résumées dans la proposition suivante :
Proposition 43

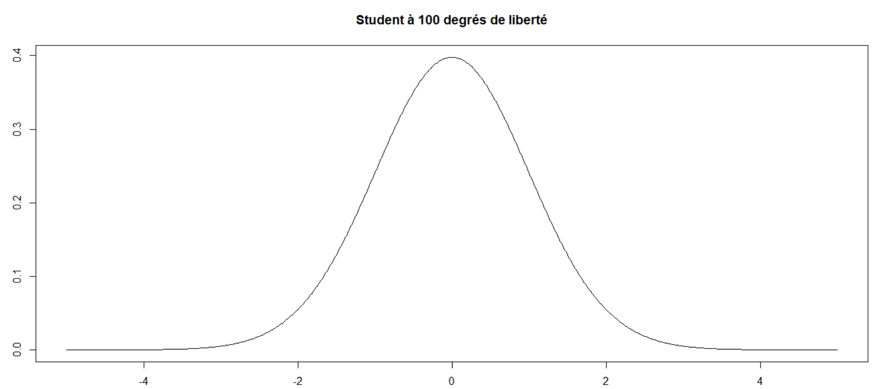
1. $E(T_n) = 0$ (**symétrie** comme la loi normale centrée réduite),
2. $V(T_n) = \dfrac{n}{n-2}$ pour $n > 2$.





Théorème 24
T_n converge en loi vers $\mathcal{N}(0,1)$ lorsque n tend vers $+\infty$.

Pour $n = 100$, la représentation graphique de la densité de la loi de Student à n degrés de libertés est « très proche » de la courbe en cloche.



Rappelons maintenant un résultat du chapitre 1.

Théorème 25
Pour tout échantillon $(X_1, \ldots, X_n)$,
$S_n^2 := \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ est un estimateur sans biais de σ^2.

Nous admettrons le résultat suivant :

Théorème 26
La variable aléatoire $T_n := \dfrac{\sqrt{n}(\bar{X}_n - m)}{\sqrt{S_n^2}}$ suit une loi de Student à $n-1$ degrés de liberté.
Remarque 1
Pourquoi cette perte d'un degré de liberté ? Il est aisément compréhensible que les n variables aléatoires $X_i - \bar{X}_n$ ne sont pas indépendantes car
$\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X}_n) = 0$ d'où la perte d'un degré de liberté.
En suivant la même méthode que dans le cas précédent,

$$\mathbb{P}\left(-t_{1-\frac{\alpha}{2}} \leq T_n \leq t_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

où $t_{1-\frac{\alpha}{2}}$ désigne le fractile d'ordre $1 - \dfrac{\alpha}{2}$ de la loi de Student à $n-1$ degrés de liberté. Nous exploitons ici la symétrie de la loi de Student.
Définition 30
L'intervalle de confiance au niveau $1-\alpha$ pour la moyenne m est :

$$\left[\bar{x}_n - t_{1-\frac{\alpha}{2}}\frac{\sqrt{s_n^2}}{\sqrt{n}}; \bar{x}_n + t_{1-\frac{\alpha}{2}}\frac{\sqrt{s_n^2}}{\sqrt{n}}\right].$$

Exemple 29
Considérons l'échantillon de 6 observations extraites d'une loi normale de moyenne et de variance inconnues : 7, 6, 4, 8, 5 et 9.
Déterminer un intervalle de confiance niveau 0,95 pour la moyenne m.

Calculons tout d'abord $\bar{x}$. On a :

$$\bar{x} = \frac{7+6+4+8+5+9}{6} = 6.5$$

Il faut maintenant calculer s^2.

$$s^2 = \frac{1}{6-1}\left((7-6.5)^2 + (6-6.5)^2 + (4-6.5)^2 + (8-6.5)^2 + (5-6.5)^2 + (9-6.5)^2\right) = 3.5$$

Nous aurons recours au fractile $t_{0{,}975} \approx 2,57$ où $t_{0{,}975}$ désigne le fractile de la loi de Student à $6-1=5$ degrés de liberté d'ordre 0.975.
L'intervalle de confiance pour la moyenne au niveau 0,95 est :

$$\left[6.5 - 2.57\frac{\sqrt{3.5}}{\sqrt{6}}; 6.5 + 2.57\frac{\sqrt{3.5}}{\sqrt{6}}\right]$$

i.e. approximativement l'intervalle $[4.54; 8.46]$.

3.3 Intervalle de confiance pour la variance σ^2

3.3.1 Cas 1 : Moyenne connue

Dans cette partie, nous considérons toujours $(X_1, \ldots, X_n)$ un échantillon de taille $n \geq 1$ de la loi $\mathcal{N}(m, \sigma^2)$. Maintenant, nous cherchons à estimer la variance σ^2.
Nous savons que les variables aléatoires $X_i - m$ sont des variables aléatoires normales indépendantes et centrées donc les variables aléatoires $\dfrac{X_i - m}{\sigma^2}$ sont des variables aléatoires indépendantes normales centrées réduites.
Par conséquent, en considérant $S_n^{2\star} := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - m)^2$, $nS_n^{2\star}$ suit une loi Khi-deux à n degrés de liberté.
Comme précédemment,

$$\mathbb{P}\left(\chi^2_{\frac{\alpha}{2}} \leq \frac{nS_n^{2\star}}{\sigma^2} \leq \chi^2_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

où $\chi^2_{\frac{\alpha}{2}}$ et $\chi^2_{1-\frac{\alpha}{2}}$ désignent respectivement le fractile d'ordre $\frac{\alpha}{2}$ et le fractile d'ordre $1 - \frac{\alpha}{2}$ de la loi $\chi^2(n)$.
Notons qu'ici, nous avons besoin des deux fractiles $\chi^2_{\frac{\alpha}{2}}$ et $\chi^2_{1-\frac{\alpha}{2}}$ car la loi $\chi^2(n)$ n'est pas symétrique.

Définition 31
L'*intervalle de confiance* au niveau $1 - \alpha$ pour la variance σ^2 est :

$$\left[n\frac{s_n^{2\star}}{\chi^2_{1-\frac{\alpha}{2}}}; n\frac{s_n^{2\star}}{\chi^2_{\frac{\alpha}{2}}}\right]$$

Exemple 30
Considérons l'échantillon de 8 observations extraites d'une loi normale de moyenne 2 et de variance inconnue : 3, 1, 5, 8, 5, 10, 12 et 9.
Déterminer un intervalle de confiance niveau 0,90 pour la variance σ^2.

Calculons $s^{2\star}$.
$s^{2\star} = \dfrac{1}{8}\left((3-2)^2 + (1-2)^2 + (5-2)^2 + (8-2)^2 + (5-2)^2 + (10-2)^2 + (12-2)^2 + (9-2)^2\right)$

donc $s^{2\star} \approx 33,63$.

Nous aurons recours aux fractiles de la loi $\chi^2(8)$:
$\chi^2_{0,05} \approx 2,733$ et $\chi^2_{0,95} \approx 15,51$.
Nous en déduisons l'intervalle de confiance niveau 0,90 pour la variance σ^2 :
$\left[8 \times \dfrac{33,63}{15,51} ; 8 \times \dfrac{33,63}{2,733}\right]$ i.e. $[17,35;98,44]$.

3.3.2 Cas 2 : Moyenne inconnue

Ne connaissant pas m, nous allons le remplacer par $\bar{X}_n$ et remplacer $S_n^{2\star}$ par S_n^2. Rappelons que $S_n^2 := \dfrac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. Nous admettrons le théorème suivant :

Théorème 27
Pour tout échantillon $(X_1, \ldots, X_n)$,
$\dfrac{(n-1)S_n^2}{\sigma^2}$ suit une loi $\chi^2(n-1)$.

L'intervalle de confiance sera de forme similaire à celui du cas précédent.

Définition 32
L'*intervalle de confiance* au niveau $1-\alpha$ pour la variance σ^2 est :

$$\left[(n-1)\frac{s_n^2}{\chi^2_{1-\frac{\alpha}{2}}} ; (n-1)\frac{s_n^2}{\chi^2_{\frac{\alpha}{2}}}\right]$$

où $\chi^2_{\frac{\alpha}{2}}$ et $\chi^2_{1-\frac{\alpha}{2}}$ désignent respectivement le fractile d'ordre $\frac{\alpha}{2}$ et le fractile d'ordre $1-\frac{\alpha}{2}$ de la loi $\chi^2(n-1)$.

Exemple 31
Considérons l'échantillon de 8 observations extraites d'une loi normale de moyenne et de variance inconnues : 3, 1, 5, 8, 5, 10, 12 et 9.
Déterminer un intervalle de confiance au niveau 0,95 pour la variance σ^2.

Calculons tout d'abord $\bar{x}$.
On a :

$$\bar{x} = \frac{3+1+5+8+5+10+12+9}{8} = 6,625$$

Il faut maintenant calculer s^2.

$$s^2 = \frac{1}{7}((3-6,625)^2+(1-6,625)^2+(5-6,625)^2+(8-6,625)^2+(5-6,625)^2+ (10-6,625)^2+(12-6,625)^2+(9-6,625)^2)$$

donc $s^2 \approx 13,98$.

Nous aurons recours aux fractiles de la loi $\chi^2(7)$: $\chi^2_{0,025} \approx 1,690$ et $\chi^2_{0,975} \approx 16,01$.

Nous en déduisons l'intervalle de confiance pour la variance σ^2 au niveau 0,95 : $\left[7 \times \dfrac{13,98}{16,01}; 7 \times \dfrac{13,98}{1,690}\right]$ i.e. approximativement $[6,11; 57,91]$.

3.4 Intervalles de confiance asymptotiques

3.4.1 Intervalle de confiance pour la moyenne m

Dans ce paragraphe, la variance est supposée connue. Ainsi, nous nous placerons dans le cas général où nous chercherons à estimer la moyenne m d'une variable aléatoire X de carré intégrable dont nous connaissons la variance σ^2(et donc l'écart-type σ). Rappelons le théorème central limite.

Théorème 28 (Théorème Central Limite)
Soit (X_i) une suite de v.a. i.i.d. telle que $E(X_1^2) < +\infty$.
Notons $m := E(X_i)$ et $\sigma^2 = V(X_i)$.
$\dfrac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$ converge en loi vers une loi normale centrée réduite.

Par conséquent, pour n assez grand [4],

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\bar{X}_n - m)}{\sigma} \leq z_{1-\frac{\alpha}{2}}\right) \approx 1-\alpha$$

avec les notations précédentes. Ainsi, nous pourrons définir la notion d'intervalle de confiance asymptotique.

4. En pratique, l'approximation est souvent jugée acceptable pour $n \geq 30$.

Définition 33 (Variance connue)
L'*intervalle de confiance asymptotique* au niveau de confiance $1-\alpha$ pour la moyenne m est :

$$\left[\bar{x}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}};\bar{x}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right].$$

Notons qu'ici, nous utilisons la variance σ^2 que nous connaissons.
Dans le cas où elle est inconnue, nous aurons recours au théorème suivant que nous admettrons :

Théorème 29
Soit (X_i) une suite de v.a. i.i.d. telle que $E(X_1^2) < +\infty$.
$\dfrac{\sqrt{n}(\bar{X}_n - m)}{\sqrt{S_n^2}}$ converge en loi vers une loi normale centrée réduite.

Par conséquent, nous obtenons l'intervalle de confiance asymptotique suivant :

Définition 34 (Variance inconnue)
L'*intervalle de confiance asymptotique* au niveau de confiance $1-\alpha$ pour la moyenne m est :

$$\left[\bar{x}_n - z_{1-\frac{\alpha}{2}}\frac{\sqrt{s_n^2}}{\sqrt{n}};\bar{x}_n + z_{1-\frac{\alpha}{2}}\frac{\sqrt{s_n^2}}{\sqrt{n}}\right].$$

Nous pourrons utiliser des intervalles plus simples lorsque la variance est fonction de la moyenne. Nous aborderons deux cas :

- estimation du paramètre p d'une variable aléatoire de Bernoulli,
- estimation du paramètre λ d'une loi de Poisson.

3.4.2 Intervalle de confiance pour la proportion

Dans cette partie, nous proposerons un intervalle de confiance pour la proportion de personnes porteuses d'un caractère dans une population donnée. Chaque personne est porteuse ou non donc une sous-population individus pourra être modélisée [5] par un échantillon de taille n de la loi de Bernoulli

5. Cette modélisation n'est valide que lorsque la population est suffisamment importante.cf. [13]

où $X_i = 1$ si la personne est porteuse du caractère et 0 sinon.
Nous notons $\hat{p}$ la proportion de personnes porteuses du caractère dans l'échantillon, elle correspond à la moyenne empirique des variables aléatoires de Bernoulli associées.
Rappelons le théorème suivant vu dans le chapitre 1 :

Théorème 30 (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.

Ce théorème implique que la loi binomiale peut être approchée par la loi normale pour n suffisamment grand. Rappelons le jeu de conditions d'approximation donné dans le chapitre 1 :

1. $n \geq 30$,
2. $n\hat{p} \geq 5$,
3. $n(1 - \hat{p}) \geq 5$.

Définition 35
L'*intervalle de confiance asymptotique* au niveau $1 - \alpha$ pour la proportion p est :

$$\left[\hat{p} - z_{1-\frac{\alpha}{2}} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} ; \hat{p} + z_{1-\frac{\alpha}{2}} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right]$$

où $z_{1-\frac{\alpha}{2}}$ désigne le fractile d'ordre $1 - \dfrac{\alpha}{2}$ de la loi normale centrée réduite.

Exemple 32
François prélève 300 serpents dans une forêt et constate que 70 d'entre eux sont venimeux.
Déterminer un intervalle de confiance asymptotique pour la proportion de serpents venimeux dans cette forêt au niveau de confiance $0,95$.

Calculons tout d'abord $\hat{p}$.
Par définition, $\hat{p} = \dfrac{70}{300} \approx 0,23$.
Vérifions les conditions d'approximation par la loi normale :

1. $n = 300 \geq 30$,
2. $n\hat{p} = 300 \times 0,23 = 69 \geq 5$,
3. $n(1-\hat{p}) = 231 \geq 5$.

Nous aurons recours au fractile $z_{0,975} \approx 1,96$.
L'intervalle de confiance asymptotique pour la proportion de serpents venimeux dans cette forêt au niveau 0,95 est :

$$\left[0,23 - 1,96\frac{\sqrt{0,23 \times 0,77}}{\sqrt{300}}; 6 + 1,96\frac{\sqrt{0,23 \times 0,77}}{\sqrt{300}}\right]$$

i.e. approximativement l'intervalle $[0,18; 0,28]$.

3.4.3 Intervalle de confiance pour le paramètre λ d'une loi de Poisson

Rappelons que, pour, une variable aléatoire X qui suit une loi $\mathcal{P}(\lambda)$, $E(X) = V(X) = \lambda$.
Par conséquent, le théorème central limite appliqué dans ce cas donne le théorème suivant :
Théorème 31
Soit (X_i) une suite de variables aléatoires i.i.d. de loi $\mathcal{P}(\lambda)$,
$\sqrt{n}\dfrac{\bar{X}_n - \lambda}{\sqrt{\lambda}}$ converge en loi vers une loi normale centrée réduite.
Par conséquent, pour n assez grand

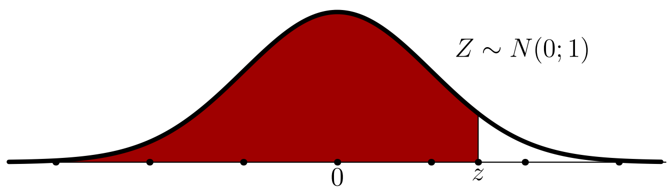
$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{n}\frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \leq z_{1-\frac{\alpha}{2}}\right) \approx 1 - \alpha$$

où $z_{1-\frac{\alpha}{2}}$ désigne le fractile d'ordre $1 - \dfrac{\alpha}{2}$ de la loi normale centrée réduite.
Définition 36
L'*intervalle de confiance asymptotique* au niveau $1 - \alpha$ pour le paramètre λ est :

$$\left[\bar{x}_n - z_{1-\frac{\alpha}{2}}\frac{\sqrt{\bar{x}_n}}{\sqrt{n}}; \bar{x}_n + z_{1-\frac{\alpha}{2}}\frac{\sqrt{\bar{x}_n}}{\sqrt{n}}\right]$$

Tableau N [1]

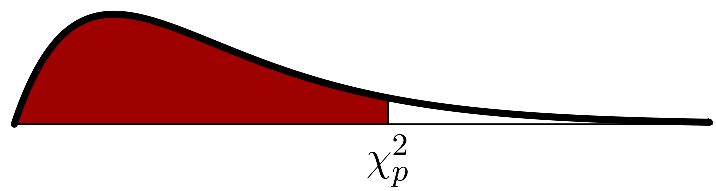
Aire sous la courbe normale à gauche de z, c'est à dire $P[Z \leq z]$, ou $Z \sim N(0;1)$.



	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.00	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.10	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.20	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.30	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.40	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.50	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.60	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.70	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.80	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.90	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09

Tableau C [1/2]

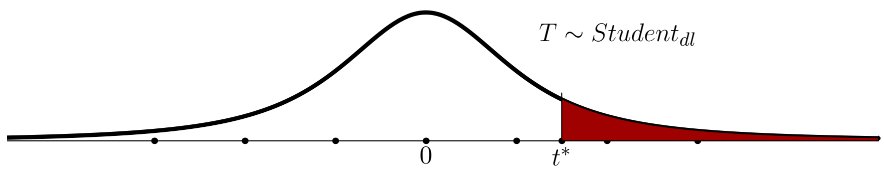
Percentiles de la distribution du χ^2. Valeurs de χ^2_P correspondant à P



dl	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.1}$	$\chi^2_{0.9}$	$\chi^2_{0.95}$	$\chi^2_{0.975}$	$\chi^2_{0.99}$	$\chi^2_{0.995}$
1	.0000	.0002	.0010	.0039	.0158	2.706	3.841	5.024	6.635	7.879
2	.0100	.0201	.0506	.1026	.2107	4.605	5.991	7.378	9.210	10.60
3	.0717	.1148	.2158	.3518	.5844	6.251	7.815	9.348	11.34	12.84
4	.2070	.2971	.4844	.7107	1.064	7.779	9.488	11.14	13.28	14.86
5	.4117	.5543	.8312	1.145	1.610	9.236	11.07	12.83	15.09	16.75
6	.6757	.8721	1.237	1.635	2.204	10.64	12.59	14.45	16.81	18.55
7	.9893	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48	20.28
8	1.344	1.646	2.180	2.733	3.490	13.36	15.51	17.53	20.09	21.95
9	1.735	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67	23.59
10	2.156	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21	25.19
11	2.603	3.053	3.816	4.575	5.578	17.28	19.68	21.92	24.72	26.76
12	3.074	3.571	4.404	5.226	6.304	18.55	21.03	23.34	26.22	28.30
13	3.565	4.107	5.009	5.892	7.042	19.81	22.36	24.74	27.69	29.82
14	4.075	4.660	5.629	6.571	7.790	21.06	23.68	26.12	29.14	31.32
15	4.601	5.229	6.262	7.261	8.547	22.31	25.00	27.49	30.58	32.80
16	5.142	5.812	6.908	7.962	9.312	23.54	26.30	28.85	32.00	34.27
17	5.697	6.408	7.564	8.672	10.09	24.77	27.59	30.19	33.41	35.72
18	6.265	7.015	8.231	9.390	10.86	25.99	28.87	31.53	34.81	37.16
dl	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.1}$	$\chi^2_{0.9}$	$\chi^2_{0.95}$	$\chi^2_{0.975}$	$\chi^2_{0.99}$	$\chi^2_{0.995}$

Tableau T1 [1/2]

Tableau de t^* tel qu'une variable de Student à dl degrés de liberté ait probabilité p d'être supérieure à t^*



	$P[T \geq t^*] = p$											
dl	0.25	0.2	0.15	0.1	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.8165	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.7649	.9785	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.7407	.9410	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.7267	.9195	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.7176	.9057	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.7111	.8960	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.7064	.8889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.7027	.8834	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.6998	.8791	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.6974	.8755	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.6955	.8726	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.6938	.8702	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.6924	.8681	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.6912	.8662	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.6901	.8647	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.6892	.8633	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.6884	.8620	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
	0.25	0.2	0.15	0.1	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005

Ce qu'il faut retenir

- Quatre cas à connaître et à bien distinguer pour les intervalles de confiance dans le cas gaussien :

- pour la moyenne :

  1. (Normale centrée réduite) $\left[\bar{x}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \bar{x}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$ si la variance est connue,
  2. (Student à $n-1$ degrés de liberté) $\left[\bar{x}_n - t_{1-\frac{\alpha}{2}}\frac{\sqrt{s_n^2}}{\sqrt{n}}; \bar{x}_n + t_{1-\frac{\alpha}{2}}\frac{\sqrt{s_n^2}}{\sqrt{n}}\right]$ si la variance est inconnue.
  3. **Lorsque n est grand (i.e. $n > 30$), la loi de Student peut être remplacée par la loi normale centrée réduite.**

- pour la variance :

  1. (Khi-deux à n degrés de libertés) $\left[n\frac{s_n^{2\star}}{\chi^2_{1-\frac{\alpha}{2}}}; n\frac{s_n^{2\star}}{\chi^2_{\frac{\alpha}{2}}}\right]$ si m est connue,
  2. (Khi-deux à $n-1$ degrés de liberté) $\left[(n-1)\frac{s_n^2}{\chi^2_{1-\frac{\alpha}{2}}}; (n-1)\frac{s_n^2}{\chi^2_{\frac{\alpha}{2}}}\right]$ si m est inconnue.

- Intervalles de confiance asymptotiques.

Bibliographie

[1] Baccini, A. : *Présentation des tests du rapport de vraisemblances, de Wald et des multiplicateurs de Lagrange*. Publications du laboratoire de probabilités et statistique, Université Paul Sabatier Toulouse (1989)

[2] Cadre, B., Vial, C. : *Statistique mathématique*. Ellipses (2012)

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

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

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

[6] Monfort, A. : *Cours de statistique mathématique*. Economica (1997)

[7] Neyman, J., Pearson, E.S. : *On the problem of the most efficients tests of statistical hypotheses*. Philosophical Transactions of the royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, Vol. 231, pp 289-337 (1933)

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

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

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

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

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

[13] Tassi, P. : *Méthodes statistiques*. Economica (2004)

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

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

[16] Wilks, S. S. : *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*. Annal of mathematical statistics, Vol. 9 Number 1, 60-62 (1938)