Le Mans Université

L2 Économie-Gestion

Statistiques et Probabilités

24/11/2025

F. Karamé

Année universitaire 2025-2026

Motivations

- les données sont (collectées) partout
- idée : utiliser les données pour
  - o améliorer les connaissances
  - o répondre à des questions concrètes
  - o l'aide à la décision
- utile dans tous les domaines
- Ce cours est une première approche de la question
- fait le lien entre stat descriptive et probabilités . . .
- . . . pour l'appliquer à l'estimation
- montre à quoi ça sert => plein de nouvelles connaissances. . .
- . . . qui seront complétées au second semestre avec l'économétrie.

On suppose une population : elle est très large et on ne peut pas l'explorer totalement $\Rightarrow$ on travaille sur un échantillon

- **La statistique descriptive pour décrire et comprendre un phénomène**

  Résumer l'information pour en extraire l'essentiel
  Applications dans tous les domaines
  Constructions de tableaux, graphiques, indicateurs synthétiques

- **La statistique mathématique pour l'aide à la décision**

  Modélisation et estimation (théorie/pratique)
  Inférence statistique (tests) (théorie/pratique)
  Déduction sur la population
  Applications dans tous les domaines...

- Pour cela on a besoin de la **théorie des probabilités**

  pour construire des échantillons
  pour avoir des outils d'estimation et connaître leurs propriétés

Vous jouez à pile ou face et vous comptez le nombre de pile. Vous disposez d'un échantillon aléatoire de N tirages $(X_1, X_2, \ldots, X_N)$ pour lesquels vous observez la réalisation de pile (succès codé 1) $(x_1, x_2, \ldots, x_N)$.

- la loi de cette expérience est la loi de **Bernoulli** : donne la probabilité de succès ($p \in [0, 1]$) ou d'échec ($1 - p$) de l'expérience.
- On suppose que chaque tirage ou variable aléatoire de l'échantillon suit la même loi de Bernoulli de paramètre $p \in [0, 1]$ et quelles sont toutes indépendantes.
- Loi de probabilités de Bernoulli de paramètre $p \in [0, 1]$:

$$\mathbb{P}(X_i = k) = p^k(1 - p)^{1-k}$$

avec $k = 0, 1$ (resp. échec, succès).

Vous êtes assureur et vous êtes intéressés par **le nombre des sinistres** sur une certaine période. Vous disposez d'un échantillon aléatoire de N clients $(X_1, X_2, \ldots, X_N)$ pour lesquels vous observez le nombre de sinistres réalisés $(x_1, x_2, \ldots, x_N)$.

- On suppose que la loi sous-jacente est la loi de **Poisson** (loi des événements rares comme les accidents).
- On suppose que chaque variable aléatoire de l'échantillon suit la même loi de Poisson et qu'elles sont toutes indépendantes.
- On suppose donc ici que **tout le monde a la même probabilité de sinistres**...!??
- Sa loi de probabilités :

$$\mathbb{P}(X_i = x_i) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

avec $x_i \in \mathbb{N}$.

- **Echantillon** = 1 groupe d'observations tirées "au hasard" dans la population
- **iid** = **i**dentiquement et **i**ndépendamment **d**istribués
- une observation de l'échantillon x_i est une **réalisation** particulière d'une Variable Aléatoire (**VA**) X_i

i	1	2	...	N
X	x_1	x_2	...	x_N

- Une VA mesure **quantitativement** le résultat d'une expérience aléatoire grâce à sa **loi de probabilités** caractérisée par un ou plusieurs **paramètres**.
- Ces paramètres sont **inconnus**, on va donc chercher à les **estimer** avec des outils **fiables**.

Partie 0 : Rappels des pré-requis :

- o sur l'utilisation des sommes
- o sur les fonctions usuelles (puissance, exp, ln)
- o sur les dérivées
- o sur les intégrales
- o sur les probabilités
- o sur les variables aléatoires
- o sur leurs moments théoriques

Pré-requis :

- o connaître ces notions de bases
- o savoir reconnaître les lois de proba classiques
- o savoir faire les calculs

Partie 1 : Les couples de VA : pour étudier les relations entre deux VA.

- couples de VA discrètes
- couples de VA continues

Partie 2 : L'inférence statistique

- Construire des estimateurs
- Étudier leurs propriétés
- Calculer des intervalles de confiance
- Faire des tests

Partie 0 : Rappels

I- Rappels sur les sommes

Commençons par une définition.

$$x_1 + x_2 + x_3 + \cdots + x_{N-1} + x_N = \sum_{i=1}^{N} x_i$$

$$\sum_{i=1}^{N} x_i = \underbrace{x_1}_{i=1} + \underbrace{x_2}_{i=2} + \underbrace{x_3}_{i=3} + \ldots + \underbrace{x_{N-1}}_{i=N-1} + \underbrace{x_N}_{i=N}$$

On peut "couper" une somme comme on veut :

$$\begin{aligned}
\sum_{i=1}^{N} x_i &= \sum_{i=1}^{10} x_i + \sum_{i=11}^{N} x_i \\
&= \sum_{i=1}^{55} x_i + \sum_{i=56}^{N} x_i \\
&= \sum_{i=1}^{22} x_i + \sum_{i=23}^{27} x_i + \sum_{i=28}^{N} x_i
\end{aligned}$$

si bien sûr $N \geqslant 11$, $N \geqslant 56$ ou $N \geqslant 28$.

Une formule utile :

$$\frac{1}{N}\sum_{i=1}^{N} x_i = \frac{x_1 + x_2 + x_3 + \cdots + x_{N-1} + x_N}{N} \equiv \bar{x}$$

On reconnaît la moyenne arithmétique. De cette égalité on déduit :

$$\sum_{i=1}^{N} x_i = N\bar{x}$$

(peut être utile parfois pour se "débarrasser" d'une somme dans une expression)

Une écriture plus compacte :

$$
\begin{aligned}
\sum_{k=0}^{4} a_k x^k &= a_0 x^0 + a_1 x^1 + a_2 x^2 + a_3 x^3 + a_4 x^4 \\
&= a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4
\end{aligned}
$$

Même chose pour la multiplication :

$$\prod_{k=1}^{n} k = 1 \times 2 \times 3 \times \ldots \times (n-1) \times n$$
$$\equiv n!$$

On reconnaît ici la définition de "factoriel n" pour $n \geqslant 1$.

Notons que $0! = 1$.

Remarque :

$$n! = n \times \underbrace{(n-1) \times (n-2) \times (n-3) \times \cdots \times 2 \times 1}_{=(n-1)!}$$
$$= n \times (n-1)!$$

Passons aux règles de calcul usuelles sur les sommes.

Ici on les démontre, ce qui nous fait manipuler les sommes, mais il faudra connaître ces formules pour pouvoir les utiliser directement ensuite.

La multiplication par une constante a qui ne dépend pas de l'indice de sommation :

$$\forall a \in \mathbb{R}, \quad \sum_{i=1}^{N} ax_i = a \sum_{i=1}^{N} x_i$$

$$\begin{aligned}
\sum_{i=1}^{N} ax_i &= ax_1 + ax_2 + ax_3 + \cdots + ax_{N-1} + ax_N \\
&= a \underbrace{\left(x_1 + x_2 + x_3 + \cdots + x_{N-1} + x_N\right)}_{\text{on utilise l'écriture en somme}} \\
&= a \sum_{i=1}^{N} x_i
\end{aligned}$$

L'addition d'une constante b qui ne dépend pas de l'indice de sommation :

$$\forall b \in \mathbb{R}, \quad \sum_{i=1}^{N} (b + x_i) = bN + \sum_{i=1}^{N} x_i$$

$$\begin{aligned}
\sum_{i=1}^{N} (b + x_i) &= (b + x_1) + (b + x_2) + (b + x_3) + \cdots + (b + x_{N-1}) + (b + x_N) \\
&= \underbrace{b + b + \cdots + b}_{N \text{ termes}} + \underbrace{x_1 + x_2 + x_3 + \cdots + x_{N-1} + x_N}_{\text{on utilise l'écriture en somme}} \\
&= Nb + \sum_{i=1}^{N} x_i
\end{aligned}$$

Même question mais avec un terme en plus dans la somme ($j = 0$) : il faut donc apprendre à compter les termes d'une somme !

$$\forall b \in \mathbb{R}, \quad \sum_{j=0}^{N} (b + x_j) = (N+1)b + \sum_{j=0}^{N} x_j$$

$$\begin{aligned}
\sum_{j=0}^{N} (b + x_j) &= (b + x_0) + (b + x_1) + (b + x_2) + \cdots + (b + x_{N-1}) + (b + x_N) \\
&= \underbrace{b + b + \cdots + b}_{N+1 \text{ termes !}} + \underbrace{x_0 + x_1 + x_2 + x_3 + \cdots + x_{N-1} + x_N}_{\text{on utilise l'écriture en somme}} \\
&= (N+1)b + \sum_{j=0}^{N} x_j
\end{aligned}$$

Règle de la distribution de la somme

$$\sum_{l=1}^{M}(x_l + y_l) = \sum_{l=1}^{M} x_l + \sum_{l=1}^{M} y_l$$

$$\begin{aligned}
\sum_{l=1}^{M}(x_l + y_l) &= (x_1 + y_1) + (x_2 + y_2) + \cdots + (x_M + y_M) \\
&= (x_1 + x_2 + \cdots + x_M) + (y_1 + y_2 + \cdots + y_M) \\
&= \sum_{l=1}^{M} x_l + \sum_{l=1}^{M} y_l
\end{aligned}$$

Se généralise à plusieurs termes

Même chose avec une soustraction

$$\sum_{i=1}^{N}(x_i - y_i) = \sum_{i=1}^{N} x_i - \sum_{i=1}^{N} y_i$$

$$\begin{aligned}
\sum_{i=1}^{N}(x_i - y_i) &= (x_1 - y_1) + (x_2 - y_2) + \cdots + (x_N - y_N) \\
&= x_1 + x_2 + \cdots + x_N - y_1 - y_2 - \cdots - y_N \\
&= (x_1 + x_2 + \cdots + x_N) - (y_1 + y_2 + \cdots + y_N) \\
&= \sum_{i=1}^{N} x_i - \sum_{i=1}^{N} y_i
\end{aligned}$$

$$\sum_{k=1}^{n-1} C_n^k \, x^k \, y^{n-k}$$

avec

$$C_n^k = \frac{n!}{k!(n-k)!}$$

Cette formule ressemble au binôme de Newton :

$$\sum_{k=0}^{n} C_n^k \, x^k \, y^{n-k} = (x+y)^n$$

avec deux termes en moins ($k = 0$ et $k = n$)

Entraînez-vous à manipuler la formule en calculant $(x+y)^2$, $(x+y)^3$, ...

Partons de là et faisons apparaître les deux termes manquants :

$$\underbrace{\sum_{k=0}^{n} C_n^k\, x^k\, y^{n-k}}_{\text{le binôme}} = \underbrace{C_n^0\, x^0\, y^{n-0}}_{k=0} + \underbrace{\sum_{k=1}^{n-1} C_n^k\, x^k\, y^{n-k}}_{\text{le terme qui nous intéresse}} + \underbrace{C_n^n\, x^n\, y^{n-n}}_{k=n}$$

$$= \underbrace{C_n^0}_{=1}\, \underbrace{x^0}_{=1}\, y^{n-0} + \sum_{k=1}^{n-1} C_n^k\, x^k\, y^{n-k} + \underbrace{C_n^n}_{=1}\, x^n\, \underbrace{y^0}_{=1}$$

$$= y^n + \sum_{k=1}^{n-1} C_n^k\, x^k\, y^{n-k} + x^n$$

Donc :

$$(x+y)^n = y^n + \sum_{k=1}^{n-1} C_n^k\, x^k\, y^{n-k} + x^n$$

$$\Leftrightarrow \sum_{k=1}^{n-1} C_n^k\, x^k\, y^{n-k} = (x+y)^n - y^n - x^n$$

$$\Leftrightarrow \sum_{k=1}^{n-1} C_n^k\, x^k\, y^{n-k} = (x+y)^n - (y^n + x^n)$$

$$\sum_{i=1}^{N}(x_i - \bar{x}) = 0$$

$$\begin{aligned}\sum_{i=1}^{N}(x_i - \bar{x}) &= \sum_{i=1}^{N} x_i - N\bar{x} \\ &= N\bar{x} - N\bar{x} = 0\end{aligned}$$

On est en train de sommer les termes d'une variable **centrée, c'est-à-dire dont les termes sont calculés en écart à leur moyenne arithmétique. La somme des termes d'une variable centrée est 0.**

$$\begin{aligned}
\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x}) &= \sum_{i=1}^{N}(y_ix_i - x_i\bar{y} - y_i\bar{x} + \bar{x}\bar{y}) \\
&= \sum_{i=1}^{N} y_ix_i - \sum_{i=1}^{N} x_i\bar{y} - \sum_{i=1}^{N} y_i\bar{x} + \sum_{i=1}^{N} \bar{x}\bar{y} \\
&= \sum_{i=1}^{N} y_ix_i - \bar{y}\sum_{i=1}^{N} x_i - \bar{x}\sum_{i=1}^{N} y_i + \bar{x}\bar{y} \underbrace{\sum_{i=1}^{N} 1}_{=1+\cdots+1=N} \\
&= \sum_{i=1}^{N} y_ix_i - \bar{y}\sum_{i=1}^{N} x_i - \bar{x}\sum_{i=1}^{N} y_i + \bar{x}\bar{y}N \\
&= \sum_{i=1}^{N} y_ix_i - \bar{y}N\bar{x} - \bar{x}N\bar{y} + N\bar{x}\bar{y} \\
&= \sum_{i=1}^{N} y_ix_i - N\bar{x}\bar{y}
\end{aligned}$$

C'est N fois la **covariance empirique** : $Cov(x,y) = \frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})$.

$$\sum_{i=1}^{N}(x_i - \bar{x})^2$$

Soit on recommence tout le calcul, soit on adapte le précédent en posant : $y_i = x_i$. Donc $\bar{x} = \bar{y}$, donc $y_i - \bar{y} = x_i - \bar{x}$ et $\sum_{i=1}^{N}(y_i - \bar{y})(x_i - \bar{x}) = \sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x}) = \sum_{i=1}^{N}(x_i - \bar{x})^2$. Il vient :

$$\begin{aligned}\sum_{i=1}^{N}(x_i - \bar{x})^2 &= \sum_{i=1}^{N} x_i x_i - N\bar{x}\bar{x} \\ &= \sum_{i=1}^{N} x_i^2 - N\bar{x}^2\end{aligned}$$

C'est N fois la **variance empirique de la variable X.**

$$\sum_{i=1}^{N} \frac{x_i}{\sum_{j=1}^{N} x_j}$$

Pas de difficulté ici si on réalise que le terme $\sum_{j=1}^{N} x_j$ est une constante.

L'expression est donc de la forme $\sum_{i=1}^{N} ax_i$ avec $a = \frac{1}{\sum_{j=1}^{N} x_j} = \frac{1}{N\bar{x}}$.

$$\begin{aligned} \sum_{i=1}^{N} \frac{x_i}{N\bar{x}} &= \frac{\sum_{i=1}^{N} x_i}{N\bar{x}} \\ &= \frac{N\bar{x}}{N\bar{x}} = 1 \end{aligned}$$

On a normalisé chaque x par la somme des x pour que la somme étudiée soit ramenée à 1. A ne pas confondre avec la **réduction** d'une variable où il faut diviser par son écart-type pour que la variance de la variable réduite soit égale à 1.

Attention à la priorité : pas de simplification !!!!!

$$\frac{\sum_{i=1}^{N}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \neq \frac{\sum_{i=1}^{N}(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})}$$

Passons à présent aux sommes infinies de termes positifs ou nuls.

Ici, l'intuition est que quand on additionne une infinité de termes positifs, le résultat devrait tendre vers l'infini.

Sauf si les derniers termes qu'on additionne se mettent à tendre vers 0 à un moment dans la somme. Additionner des 0 ne change plus la somme.

Pour que le résultat de la somme tende vers une limite finie, il faut donc que le terme générique converge vers 0 suffisamment rapidement (c'est-à-dire avant que la somme ait divergé).

Il faut donc d'abord vérifier une condition nécessaire mais non suffisante qui est que **le terme générique tend vers 0 quand l'indice de sommation tend vers l'infini** (en fin de somme donc).

- Si c'est le cas, la somme va peut-être converger vers une limite finie mais ce n'est pas certain.
- Si ce n'est pas le cas, on est sûr que la série diverge.

Petit conseil : révisez les résultats sur les sommes rencontrées l'année dernière en cours : somme des termes d'une suite arithmétique, des termes au carré, d'une suite géométrique, ... Cela sera souvent utile.

$$\sum_{n=1}^{\infty} \frac{1}{2^{n-1}}$$

Vérifions la condition de convergence :

$$\lim_{n\to\infty} \frac{1}{2^{n-1}} = \lim_{n\to\infty} \left(\frac{1}{2}\right)^{n-1} = 0$$

puisque $\frac{1}{2} < 1$. La condition est remplie, on va peut-être converger vers une valeur finie.

On reconnaît ici la somme des termes d'une suite géométrique de raison $\frac{1}{2} < 1$ et dont le premier terme égale 1 (quand $n = 1$).

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{2^{n-1}} &= \lim_{N\to\infty} \sum_{n=1}^{N} \left(\frac{1}{2}\right)^{n-1} \\
&= \lim_{N\to\infty} \frac{1 - \left(\frac{1}{2}\right)^{N}}{1 - \frac{1}{2}} \quad \text{(on peut utiliser la formule générale)} \\
&= \frac{1 - \lim\limits_{N\to\infty} \left(\frac{1}{2}\right)^{N}}{\frac{1}{2}} \\
&= \frac{1}{\frac{1}{2}} = 2
\end{aligned}$$

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$$

Vérifions la condition de convergence :

$$\lim_{n \to \infty} \frac{1}{n(n+1)} = 0$$

La condition est remplie, on va peut-être converger vers une valeur finie.

L'astuce consiste ici à utiliser le fait que :

$$\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}$$

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n(n+1)} &= \lim_{N\to\infty} \sum_{n=1}^{N} \frac{1}{n(n+1)} \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \left(\frac{1}{n} - \frac{1}{n+1}\right) \\
&= \lim_{N\to\infty} [\left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots + \\
&\qquad \left(\frac{1}{N-1} - \frac{1}{N}\right) + \left(\frac{1}{N} - \frac{1}{N+1}\right)] \\
&= \lim_{N\to\infty} \left(1 - \frac{1}{N+1}\right) \\
&= 1 - \lim_{N\to\infty} \frac{1}{N+1} \\
&= 1
\end{aligned}$$

$$\sum_{n=1}^{\infty} \frac{1}{5^{n+2}}$$

Vérifions la condition de convergence :

$$\lim_{n\to\infty} \frac{1}{5^{n+2}} = \lim_{n\to\infty} \left(\frac{1}{5}\right)^{n+2} = 0$$

puisque $\frac{1}{5} < 1$. La condition est remplie, on va peut-être converger vers une valeur finie.

On reconnaît ici la somme des termes d'une suite géométrique de raison $\frac{1}{5}$ dont le premier terme est $\left(\frac{1}{5}\right)^3 \neq 1$ (quand $n = 1$, c'est le piège !).

Pour utiliser la formule habituelle, il faut que le premier terme de la somme soit 1.

On va donc factoriser par $\left(\frac{1}{5}\right)^3$.

$$
\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{5^{n+2}} &= \lim_{N\to\infty} \sum_{n=1}^{N} \frac{1}{5^{n+2}} \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \left(\frac{1}{5}\right)^{n+2} \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \left(\frac{1}{5}\right)^{3} \left(\frac{1}{5}\right)^{n-1} \\
&= \lim_{N\to\infty} \left(\frac{1}{5}\right)^{3} \sum_{n=1}^{N} \left(\frac{1}{5}\right)^{n-1} \\
&= \lim_{N\to\infty} \left(\frac{1}{5}\right)^{3} \frac{1-\left(\frac{1}{5}\right)^{N}}{1-\frac{1}{5}} \quad \text{(maintenant on peut)} \\
&= \left(\frac{1}{5}\right)^{3} \frac{1-\lim\limits_{N\to\infty}\left(\frac{1}{5}\right)^{N}}{\frac{4}{5}} \\
&= \left(\frac{1}{5}\right)^{3} \frac{1}{\frac{4}{5}} = \frac{1}{5^3}\,\frac{5}{4} = \frac{1}{5^2}\,\frac{1}{4} = \frac{1}{100}
\end{aligned}
$$

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

Vérifions la condition de convergence :

$$\lim_{n \to \infty} \frac{1}{n} = 0$$

La condition nécessaire est remplie, cependant on va montrer que la limite des termes à 0 n'est pas atteinte suffisamment rapidement pour que la somme converge. Cela illustre bien son caractère non suffisant.

Raisonnons par l'absurde. Posons une hypothèse de départ pour construire un raisonnement : soit

$$S_N = \sum_{n=1}^{N} \frac{1}{n}$$

et supposons que la somme converge vers une limite ℓ :

$$\lim_{N\to\infty} S_N = \ell$$

C'est donc aussi le cas si on calcule S_{2N} :

$$\lim_{N\to\infty} S_{2N} = \lim_{N\to\infty} \sum_{n=1}^{2N} \frac{1}{n} = \ell$$

Dès lors, cela implique :

$$\begin{aligned}\lim_{N\to\infty} (S_{2N} - S_N) &= \lim_{N\to\infty} S_{2N} - \lim_{N\to\infty} S_N \\ &= \ell - \ell = 0\end{aligned}$$

Donc si on arrive à une conclusion différente, c'est que l'hypothèse de départ est fausse.

$$
\begin{aligned}
S_{2N} - S_N &= \sum_{n=1}^{2N} \frac{1}{n} - \sum_{n=1}^{N} \frac{1}{n} \\
&= \sum_{n=N+1}^{2N} \frac{1}{n} \\
&= \frac{1}{N+1} + \frac{1}{N+2} + \cdots + \frac{1}{2N}
\end{aligned}
$$

Comme

$$
\begin{aligned}
& N+1 < 2N \Leftrightarrow \frac{1}{N+1} > \frac{1}{2N} \\
& N+2 < 2N \Leftrightarrow \frac{1}{N+2} > \frac{1}{2N} \\
& \ldots \\
& 2N \leqslant 2N \Leftrightarrow \frac{1}{2N} \geqslant \frac{1}{2N}
\end{aligned}
$$

$$
S_{2N} - S_N > \frac{1}{2N} + \frac{1}{2N} + \cdots + \frac{1}{2N} = \frac{N}{2N} = \frac{1}{2}
$$

Donc $\lim\limits_{N\to\infty} (S_{2N} - S_N) \neq 0$ puisque plus grand que $\frac{1}{2}$. Donc l'hypothèse de départ $\lim\limits_{N\to\infty} S_N = \ell$ est fausse et donc la série diverge.

$$\sum_{k=1}^{\infty} \frac{1}{k^2}$$

Vérifions la condition de convergence :

$$\lim_{k \to \infty} \frac{1}{k^2} = 0$$

La condition nécessaire est remplie, on va peut-être converger.

D'abord, notons que $\forall k \in \mathbb{N}$:

$$\begin{aligned}
& k > (k-1) \\
\Leftrightarrow & k^2 > k(k-1) \\
\Leftrightarrow & \frac{1}{k^2} < \frac{1}{k(k-1)} \quad \forall k > 1 \text{ c'est-à-dire } \forall k \geqslant 2 \\
\Leftrightarrow & \sum_{k=2}^{n} \frac{1}{k^2} < \sum_{k=2}^{n} \frac{1}{k(k-1)}
\end{aligned}$$

On évite $k = 1$ et on commence en $k = 2$ pour ne pas diviser par 0.
Utilisons l'astuce de la cascade en constatant que :

$$\frac{1}{k(k-1)} = \frac{1}{k-1} - \frac{1}{k}$$

$$\begin{aligned}\sum_{k=2}^{n} \frac{1}{k^2} &< \sum_{k=2}^{n} \left(\frac{1}{k-1} - \frac{1}{k} \right) \\ &< \left(\frac{1}{1} - \frac{1}{2} \right) + \left(\frac{1}{2} - \frac{1}{3} \right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n} \right) \\ &< 1 - \frac{1}{n}\end{aligned}$$

On y est presque.
A gauche, il manque le premier terme de la somme, quand $k = 1$. Dans ce cas, le terme générique vaut 1. Additionner 1 de chaque coté de l'inégalité ne change rien.

$$\begin{aligned}1 + \sum_{k=2}^{n} \frac{1}{k^2} &< 1 + 1 - \frac{1}{n} \\ \Leftrightarrow \sum_{k=1}^{n} \frac{1}{k^2} &< 2 - \frac{1}{n}\end{aligned}$$

Il ne reste plus qu'à prendre la limite :

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k^2} < \lim_{n\to\infty} \left(2 - \frac{1}{n}\right)$$
$$< 2 - \lim_{n\to\infty} \frac{1}{n}$$
$$< 2$$

La série a donc bien convergé même si ici on n'a pas donné explicitement sa limite.

$$\left(\sum_{i=1}^{N} a_i\right)^2 \neq \sum_{i=1}^{N} a_i^2$$

Ici la question est juste de vérifier qu'il n'y a pas de confusion entre $\sum_{i=1}^{N} a_i^2$ et $\left(\sum_{i=1}^{N} a_i\right)^2$. En effet, le dernier contient le précédent du fait des termes croisés . . . !

$$\begin{aligned}(a+b)^2 &= a^2 + ab + ba + b^2 \\ &= a^2 + b^2 + 2ab\end{aligned}$$

$$
\begin{aligned}
\left(\sum_{i=1}^{N} a_i\right)^2 &= \sum_{i=1}^{N} a_i^2 + \sum_{i=1}^{N} \sum_{j \neq i}^{N} a_i a_j \\
&= \sum_{i=1}^{N} a_i^2 + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} a_i a_j \quad (\text{on passe de } j \neq i \text{ à } j > i) \\
&= \sum_{i=1}^{N} a_i^2 + 2 \sum_{i=1, j>i}^{N} a_i a_j \quad (\text{notation compacte})
\end{aligned}
$$

Appliquez pour $N = 3 \ldots.$

II- Rappels sur les fonctions usuelles

Rappelez les règles de calculs des fonctions usuelles

Bien qu'on ait déjà bien étudié ces questions en première année, il y a encore beaucoup d'approximations dans l'utilisation des fonctions avec des puissances, causant des erreurs ou des simplifications de résultats non abouties. Il est très important de bien maîtriser ces règles.

$$x^a \, x^b$$

On se souvient que $x^a = \underbrace{x \times \cdots \times x}_{a \text{ fois}}$.

Donc en revenant à cette définition si on a un doute :

$$\begin{aligned} x^a \, x^b &= \underbrace{\underbrace{x \times \cdots \times x}_{a \text{ fois}} \times \underbrace{x \times \cdots \times x}_{b \text{ fois}}}_{a+b \text{ fois}} \\ &= x^{a+b} \end{aligned}$$

$$(x^a)^b$$

$$\begin{aligned}(x^a)^b &= \underbrace{x^a \times \ldots \times x^a}_{b \text{ fois}} \\ &= x^{ab}\end{aligned}$$

donc $(x^{\frac{1}{a}})^b = x^{\frac{b}{a}}$.

$$x^a\ y^a$$

$$\begin{aligned} x^a\ y^a &= \underbrace{x \times \cdots \times x}_{a \text{ fois}} \times \underbrace{y \times \cdots \times y}_{a \text{ fois}} \\ &= \underbrace{xy \times \cdots \times xy}_{a \text{ fois}} \\ &= (xy)^a \end{aligned}$$

$$x^{\frac{1}{a}}$$

c'est la racine a-ième de x. Préférez cette écriture plutôt que celle avec la racine.

$$\frac{1}{x^a}$$

Ici il est important de comprendre que $\frac{1}{x^a} = x^{-a}$.

Pour s'en convaincre, calculons :

$$\begin{aligned} x^a \times x^{-a} &= x^{a-a} \\ &= x^0 \\ &= 1 \end{aligned}$$

grâce au point 1. De l'égalité, il vient :

$$x^a \times x^{-a} = 1 \Leftrightarrow x^{-a} = \frac{1}{x^a}$$

$$\frac{1}{x^{-a}}$$

Grâce au point précédent et aux règles de calcul des fractions :

$$\frac{1}{x^{-a}} = \frac{1}{\frac{1}{x^a}} = 1 \times \frac{x^a}{1} = x^a$$

$\lim_{x \to 0^+} \ln(x) = -\infty$

$\ln(1) = 0.$

$\lim_{x \to +\infty} \ln(x) = +\infty.$

$\ln(a \times b) = \ln(a) + \ln(b).$

$\ln(a^n) = n \times \ln(a).$

$\ln(\frac{a}{b}) = \ln(a) - \ln(b).$

$\ln(\frac{1}{x}) = -\ln(x).$

$\lim_{x \to -\infty} e^x = 0.$

$e^0 = 1.$

$e^1 = e$, la constante d'Euler.

$\lim_{x \to +\infty} e^x = +\infty.$

$e^{a+b} = e^a \times e^b$ (propriété 3).

$e^{-a} = \frac{1}{e^a}$ (propriété 6).

$e^{a-b} = \frac{e^a}{e^b}$ (propriétés 1 et 6).

$e^{q \times a} = (e^q)^a = (e^a)^q$ (propriété 2).

$e^{\ln(u(x))} = u(x).$

$\ln(e^{u(x)}) = u(x).$

III- Rappels sur les dérivées

Rappelez les définitions :

- de la dérivée d'une fonction en un point. Comment peut-on l'interpréter ?
- de la fonction dérivée. A quoi sert-elle ?
- de la fonction dérivée seconde. A quoi sert-elle ?
- de la fonction dérivée partielle. A quoi sert-elle ?

Soient a, b et α des réels. Rappelez les définitions de la dérivée $f'(x)$ de la fonction $f(x)$ selon qu'elle est définie comme :

- $f(x) = a \Rightarrow f'(x) = 0$
- $f(x) = ax + b \Rightarrow f'(x) = a$
- $f(x) = x^{\alpha} \Rightarrow f'(x) = \alpha x^{\alpha-1}$
- $f(x) = \sum_{i=0}^{N} a_i \, x^i \Rightarrow f'(x) = \sum_{i=1}^{N} i \times a_i \, x^{i-1}$
- $f(x) = u(x) + v(x) \Rightarrow f'(x) = u'(x) + v'(x)$
- $f(x) = \sum_{i=1}^{N} u_i(x) \Rightarrow f'(x) = \sum_{i=1}^{N} u_i'(x)$
- $f(x) = u(x) \times v(x) \Rightarrow f'(x) = u'(x) \times v(x) + u(x) \times v'(x)$
- $f(x) = \frac{u(x)}{v(x)} \Rightarrow f'(x) = \frac{u'(x)v(x) - u(x)v'(x)}{v^2(x)}$

- $f(x) = h(g(x)) \Rightarrow f'(x) =$
- $f(x) = [u(x)]^{\alpha} \Rightarrow f'(x) = \alpha u'(x)[u(x)]^{\alpha-1}$
- $f(x) = \frac{1}{v(x)} \Rightarrow f'(x) = -\frac{v'(x)}{v^2(x)}$
- $f(x) = \sqrt{x} \Rightarrow f'(x) = \frac{1}{2\sqrt{x}}$
- $f(x) = \ln(x) \Rightarrow f'(x) = \frac{1}{x}$
- $f(x) = \ln[u(x)] \Rightarrow f'(x) = \frac{u'(x)}{u(x)}$
- $f(x) = e^x \Rightarrow f'(x) = e^x$
- $f(x) = e^{u(x)} \Rightarrow f'(x) = u'(x) \times e^{u(x)}$

IV- Rappels sur les intégrales

Qu'est-ce qu'une intégrale ?

$$\int_a^b f(x)\ dx = \lim_{n \to +\infty} \sum_{i=0}^{n-1} \left(\frac{b-a}{n} \right) \times f\left(a + i \times \frac{b-a}{n} \right)$$

Comment peut-on l'interpréter ?

Qu'est-ce qu'une primitive ?

$F(.)$ est une primitive de $f(.)$ (à une constante près) ssi : $F'(x) = f(x)$.

Il vient : $\int_a^b f(x)\ dx = \left[F(x)\right]_a^b = F(b) - F(a)$.

Donc la constante d'intégration ne compte pas dans le calcul d'une intégrale.

Soient a, k_0 et α des réels. Les primitives $F(x)$ (à k_0 la constante d'intégration près) de la fonction $F'(x) = f(x)$ selon qu'elle est définie comme :

- $f(x) = 0 \Rightarrow F(x) = k_0$
- $f(x) = a \Rightarrow F(x) = ax + k_0$
- $f(x) = x^\alpha \Rightarrow F(x) = \frac{1}{\alpha+1} x^{\alpha+1} + k_0$
- $f(x) = u'(x)[u(x)]^\alpha \Rightarrow F(x) = \frac{1}{\alpha+1}[u(x)]^{\alpha+1} + k_0$
- $f(x) = \frac{v'(x)}{v^2(x)} \Rightarrow F(x) = -\frac{1}{v(x)} + k_0$
- $f(x) = \frac{1}{x} \Rightarrow F(x) = \ln(x) + k_0$
- $f(x) = \frac{u'(x)}{u(x)} \Rightarrow F(x) = \ln[u(x)] + k_0$
- $f(x) = e^x \Rightarrow F(x) = e^x + k_0$
- $f(x) = u'(x) \times e^{u(x)} \Rightarrow F(x) = e^{u(x)} + k_0$
- $f(x) = u'(x) + v'(x) \Rightarrow F(x) = u(x) + v(x) + k_0$

Règles de calcul sur les intégrales

- $\int_a^b f(x)\,dx = \lim\limits_{n\to+\infty} \sum_{i=0}^{n-1} \left(\frac{b-a}{n}\right) \times f\left(a + i \times \frac{b-a}{n}\right)$
- $\int_a^b \{f(x)+g(x)\}dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$
- $\int_a^b c \times f(x)dx = c\int_a^b f(x)dx$
- $\int_b^a f(x)dx = -\int_a^b f(x)dx$
- $\int_a^b f(x)dx + \int_b^a f(x)dx = 0$
- $\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx$
- $\int_a^a f(x)dx = 0$
- $g(x) = f(x) \Rightarrow \int_{x\in I} g(x)dx = \int_{x\in I} f(x)dx$
- $\int_{-u}^u f(x)dx = \begin{cases} 0 & \text{si } f \text{ impair} \\ 2\int_0^u f(x)dx & \text{si } f \text{ pair} \end{cases}$
- $\int_a^b f(x)dx = [F(x)]_a^b = F(b) - F(a)$
- $\int_{-\infty}^b f(x)dx = [F(x)]_{-\infty}^b = F(b) - \lim\limits_{a\to-\infty} F(a)$
- $\int_a^{+\infty} f(x)dx = [F(x)]_a^{+\infty} = \lim\limits_{b\to+\infty} F(b) - F(a)$
- $\int_{-\infty}^{+\infty} f(x)dx = [F(x)]_{-\infty}^{+\infty} = \lim\limits_{b\to+\infty} F(b) - \lim\limits_{a\to-\infty} F(a)$
- $\int_{x=a}^b \int_{y=c}^d f(x,y)dxdy = \int_{y=c}^d \int_{x=a}^b f(x,y)dxdy$

V- Rappels sur les probabilités

Soit E l'ensemble des réalisables :

$$\mathbb{P}(E) = 1$$
$$\mathbb{P}(\varnothing) = 0$$

$\cup$: union d'ensembles ou d'événements : signifie ”ou”

$\cap$: intersection d'ensembles ou d'événements : signifie ”et”

Très liés aux notions d'événements **incompatibles** et **indépendants**

Soient A et B deux événements incompatibles ($A \cap B = \varnothing$) :

$$\mathbb{P}(A \cap B) = 0$$
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

Soient $\left\{A_i\right\}_{i=1}^{n}$ un ensemble de n événements incompatibles :

$$\mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} \mathbb{P}(A_i)$$

Soient A et son événement contraire $\overline{A}$ dans E.

Par définition, ils sont incompatibles ($A \cap \overline{A} = \varnothing$) et forment un système complet d'événements ($A \cup \overline{A} = E$) :

$$\begin{aligned} & \mathbb{P}(A \cup \overline{A}) = \mathbb{P}(E) \\ \Leftrightarrow\ & \mathbb{P}(A) + \mathbb{P}(\overline{A}) = 1 \\ \Leftrightarrow\ & \mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A) \end{aligned}$$

Généralisation si pas incompatibles

$$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) \\ &\quad - \mathbb{P}(A \cap B)\end{aligned}$$

$$\begin{aligned}\mathbb{P}(A \cup B \cup C) &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \\ &\quad - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) \\ &\quad + \mathbb{P}(A \cap B \cap C)\end{aligned}$$

$$\begin{aligned}\mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) &= \sum_{i=1}^{n} \mathbb{P}(A_i) \\ &\quad - \sum_{i=1, i<j}^{n} \mathbb{P}(A_i \cap A_j) \\ &\quad + \sum_{i=1, i<j<k}^{n} \mathbb{P}(A_i \cap A_j \cap A_k) \\ &\quad + \dots \\ &\quad + (-1)^{n+1} \mathbb{P}\Big(\bigcap_{i=1}^{n} A_i\Big)\end{aligned}$$

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

On a aussi

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

On en déduit la **formule des probabilités composées** :

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

et on peut réécrire :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$
$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}$$

Supposons A_1 et A_2 un système complet d'événements (2 événements incompatibles tq $A_1 \cup A_2 = E$). Soit B un événement.

$$\begin{aligned} B &= B \cap E \\ &= B \cap (A_1 \cup A_2) \\ &= (B \cap A_1) \cup (B \cap A_2) \end{aligned}$$

Comme incompatibles :

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}\Big[\big(B \cap A_1\big) \cup \big(B \cap A_2\big)\Big] \\ &= \mathbb{P}\big(B \cap A_1\big) + \mathbb{P}\big(B \cap A_2\big) \\ &= \mathbb{P}\big(B|A_1\big)\mathbb{P}\big(A_1\big) \; + \mathbb{P}\big(B|A_2\big)\mathbb{P}\big(A_2\big) \end{aligned}$$

en utilisant la formule des probabilités composées.

Généralisation

Supposons $\left\{A_i\right\}_{i=1}^{n}$ un système complet d'événements (n événements incompatibles tq $\bigcup_{i=1}^{n} A_i = E$). Soit B un événement.

$$\begin{aligned}
B &= B \cap E \\
&= B \cap \left(\bigcup_{i=1}^{n} A_i\right) \\
&= \bigcup_{i=1}^{n} (B \cap A_i) \\
&= \left(B \cap A_1\right) \cup \left(B \cap A_2\right) \cup \cdots \cup \left(B \cap A_n\right)
\end{aligned}$$

Comme incompatibles :

$$\begin{aligned}
\mathbb{P}(B) &= \mathbb{P}\left[\left(B \cap A_1\right) \cup \left(B \cap A_2\right) \cup \cdots \cup \left(B \cap A_n\right)\right] \\
&= \mathbb{P}\left(B \cap A_1\right) + \mathbb{P}\left(B \cap A_2\right) + \cdots + \mathbb{P}\left(B \cap A_n\right) \\
&= \sum_{i=1}^{n} \mathbb{P}\left(B \cap A_i\right) \\
&= \sum_{i=1}^{n} \mathbb{P}\left(B | A_i\right) \mathbb{P}\left(A_i\right)
\end{aligned}$$

Imaginons qu'on recherche à présent $\mathbb{P}\big(A_j|B\big)$.

Avec Bayes on a :

$$\begin{aligned}\mathbb{P}(A_j|B) &= \frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\sum_{i=1}^{n}\mathbb{P}(B|A_i)\mathbb{P}(A_i)}\end{aligned}$$

avec la formule des probabilités totales.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$
$$\mathbb{P}(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} \mathbb{P}(A_i)$$

Il vient :

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

VI- Rappels sur les variables aléatoires

Soit X suit une variable aléatoire discrète de loi de probabilités $\mathbb{P}(X = x)$ pour $x \in \mathcal{X}$.

Comment sait-on qu'on est en présence d'une loi de probabilités ?

Une loi de probabilités est une formule qui prédit $\mathbb{P}(X = x)$, la probabilité des résultats numériques discrets $x \in \mathcal{X} \subset \mathbb{Z}$ d'une expérience aléatoire codifiée.

Elle vérifie les conditions suivantes :

- $\forall x \in \mathcal{X}, \quad \mathbb{P}(X = x) \in [0, 1]$ car ce sont des probabilités.
- $\sum_{x \in \mathcal{X}} \mathbb{P}(X = x) = 1$ car $\mathcal{X}$ est l'ensemble des réalisables.

Dans la pratique, on vérifiera simplement la positivité des probabilités et leur somme à 1.

$\forall k \in \mathcal{X} = [1, \dots, n], \;\; \mathbb{P}(X = k) = \frac{1}{n}$

C'est ici la loi uniforme discrète rencontrée à l'exercice 2. Les n résultats possibles de l'expérience sont tous équiprobables (comme le résultat d'un lancer de dé pour $n = 6$).

On constate que $\forall k \in \mathcal{X}, \;\; \mathbb{P}(X = k) = \frac{1}{n} \in [0, 1]$.

Ensuite, il faut montrer que $\sum_{k \in \mathcal{X}} \mathbb{P}(X = k) = 1$:

$$\begin{aligned} \sum_{k \in \mathcal{X}} \mathbb{P}(X = k) &= \sum_{k=1}^{n} \frac{1}{n} \\ &= n \times \frac{1}{n} \\ &= 1 \end{aligned}$$

On est donc bien en présence d'une loi de probabilités.

$\mathbb{P}(X = x) = p^x \ (1-p)^{(1-x)}$ avec $x = 0, 1$.

C'est la loi de Bernoulli, une expérience aléatoire à 2 issues : succès/échec (comme gagner/perdre, pile/face, ...) avec p la probabilité associée au "succès" et $1-p$ celle de l'"échec".

Attention, ici la notion de succès n'est pas toujours positive (comme la probabilité de faire faillite, être malade, mourir, ...).

Comme $\mathbb{P}(X = 1) = p \in [0, 1]$ ou $\mathbb{P}(X = 0) = 1 - p \in [0, 1]$, la première condition est bien vérifiée.

Ensuite, il faut montrer que $\sum_{x \in \mathcal{X}} \mathbb{P}(X = x) = 1$:

$$\begin{aligned}
\sum_{x \in \mathcal{X}} \mathbb{P}(X = x) &= \sum_{x=0}^{1} p^x \ (1-p)^{(1-x)} \\
&= \underbrace{p^0 (1-p)^1}_{\text{si } x=0} + \underbrace{p^1 (1-p)^0}_{\text{si } x=1} \\
&= (1-p) + p \\
&= 1
\end{aligned}$$

On est donc bien en présence d'une loi de probabilités.

$\forall k \in [0, \ldots, n], \;\; \mathbb{P}(X = k) = C_n^k \, p^k \, (1-p)^{(n-k)}$, avec $C_n^k = \frac{n!}{k!(n-k)!}$.

C'est la loi Binomiale. C'est une expérience qui consiste à déterminer la probabilité de k succès dans la répétition de n expériences de Bernoulli indépendantes. Ainsi, l'ensemble des possibles pour k est de 0 succès sur n ou de n succès sur n.

Il est évident que $\mathbb{P}(X = k)$ est positif comme produit des termes positifs.

Ensuite, il faut montrer que $\sum_{k \in \mathcal{X}} \mathbb{P}(X = k) = 1$. En utilisant le binôme de Newton :

$$\begin{aligned}
\sum_{k \in \mathcal{X}} \mathbb{P}(X = k) &= \sum_{k=0}^{n} C_n^k \, p^k \, (1-p)^{(n-k)} \\
&= [p + (1-p)]^n \\
&= 1^n \\
&= 1
\end{aligned}$$

On est donc bien en présence d'une loi de probabilités.

$\forall k \in [1, +\infty[, \quad \mathbb{P}(X = k) = (1-p)^{k-1}\, p.$
C'est la loi géométrique. Elle sert à compter le temps d'attente du premier succès dans la répétition d'une loi de Bernoulli de façon indépendante. On peut avoir un succès dès le premier tirage ou jamais (événement quasi impossible).

De façon évidente, $\forall k,\ \mathbb{P}(X = k) \in [0, 1]$ comme produit de termes positifs et plus petits que 1.

Ensuite, il faut montrer que $\sum_{k\in\mathcal{X}} \mathbb{P}(X = k) = 1$. En utilisant la somme des termes d'une suite géométrique de raison $(1-p)$:

$$\begin{aligned}\sum_{k\in\mathcal{X}} \mathbb{P}(X = k) &= \sum_{k=1}^{+\infty} (1-p)^{k-1}\, p \\ &= p \sum_{k=1}^{+\infty} (1-p)^{k-1} \\ &= p\, \frac{1-(1-p)^{+\infty}}{1-(1-p)} = p\, \frac{1}{1-(1-p)} = \frac{p}{p} = 1\end{aligned}$$

On est donc bien en présence d'une loi de probabilités.

$\forall k \in [0, +\infty[$ et $\lambda > 0, \mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ (indication : $e^x = \sum_{k=0}^{+\infty} \frac{x^k}{k!}$).

C'est la loi de Poisson, ou loi des événements rares. Elle sert à dénombrer le nombre d'accidents sur une période de temps. On peut donc en avoir de 0 à une infinité (vraiment pas de chance).

Il est évident que $\mathbb{P}(X = k)$ est positif comme produit des termes positifs.

Ensuite, il faut montrer que $\sum_{k \in \mathcal{X}} \mathbb{P}(X = k) = 1$. En utilisant le développement limité de l'exponentielle :

$$\begin{aligned}\sum_{k \in \mathcal{X}} \mathbb{P}(X = k) &= \sum_{k=0}^{+\infty} \frac{\lambda^k e^{-\lambda}}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \times e^{\lambda} \\ &= e^{-\lambda+\lambda} = e^0 = 1\end{aligned}$$

On est donc bien en présence d'une loi de probabilités.

Posons : $F(x) = \mathbb{P}(X < x)$ et $F'(x) = f(x)$

$$\begin{aligned}
\mathbb{P}(a < X < b) &= \mathbb{P}(X < b) - \mathbb{P}(X < a) \\
&= F(b) - F(a) \\
&= [F(x)]_a^b \\
&= \int_a^b f(x)\ dx
\end{aligned}$$

$F(.)$ est la fonction de répartition (ou *cumulative density function*)

$f(.)$ est la fonction de densité telle que $f(x) = F'(x)$.

Définir la fonction de répartition $F(.)$ en termes de probabilités.

$$F(x) = \mathbb{P}(X < x)$$

La fonction de densité $f(.)$ est la dérivée de la fonction de répartition.

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

C'est la probabilité moyenne d'un intervalle infiniment petit autour de x :

$$\begin{aligned} f(x) &= F'(x) \\ &= \lim_{h\to 0} \frac{F(x+h) - F(x)}{h} \\ &= \lim_{h\to 0} \frac{\mathbb{P}(X < x+h) - \mathbb{P}(X < x)}{h} \\ &= \lim_{h\to 0} \frac{\mathbb{P}(x < X < x+h)}{h} \end{aligned}$$

Pour les VAC, son rôle est équivalent à celui de la loi de probabilités pour les VAD **MAIS** ce n'est pas une probabilité !

Soit X suit une variable aléatoire continue de densité de probabilités $f(x)$ pour $x \in \mathcal{X}$, l'ensemble des réalisables ou support.

Comment sait-on que $f(.)$ est une densité de probabilités ?

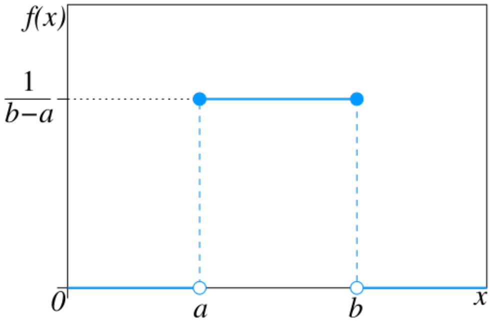
Une densité de probabilités vérifie les conditions suivantes :

- $\forall x \in \mathcal{X}, \quad f(x) \geqslant 0$. Donc la fonction de répartition est non-décroissante.

- C'est une fonction continue (sauf éventuellement en un nombre fini de points).

- $\int_{x \in \mathcal{X}} f(x) \, dx = 1$.

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



Positive (car $b > a$)
Continue sauf en a et b.

$$\begin{aligned}\int_{x \in \mathcal{X}} f(x)\, dx &= \int_a^b \frac{1}{b-a}\, dx \\ &= \left[\frac{x}{b-a}\right]_a^b \\ &= \frac{b}{b-a} - \frac{a}{b-a} \\ &= \frac{b-a}{b-a} \\ &= 1\end{aligned}$$

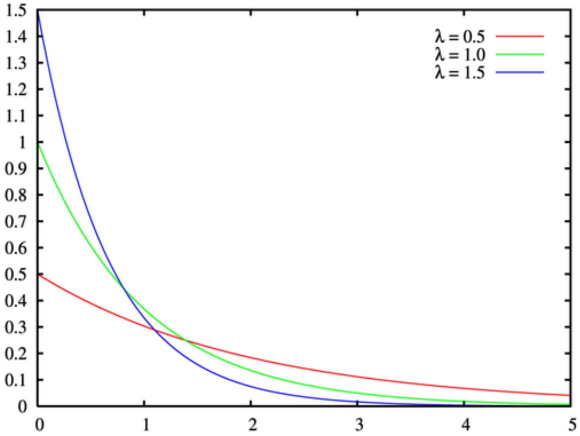
On est donc bien en présence d'une densité de probabilités. MAIS ce n'est pas une probabilité (posez $a = 0$ et $b = 0.5$: $f(x) = 2$.)

Une loi exponentielle modélise la durée de vie d'un phénomène sans mémoire, ou sans vieillissement, ou sans usure.

La probabilité que le phénomène dure au moins $t + s$ heures (ou n'importe quelle autre unité de temps) sachant qu'il a déjà duré t heures sera la même que la probabilité de durer s heures à partir de sa mise en fonction initiale.

En d'autres termes, le fait que le phénomène ait duré pendant t heures ne change rien à son espérance de vie à partir du temps t.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \forall x > 0, \lambda > 0 \\ 0 & \text{sinon} \end{cases}$$



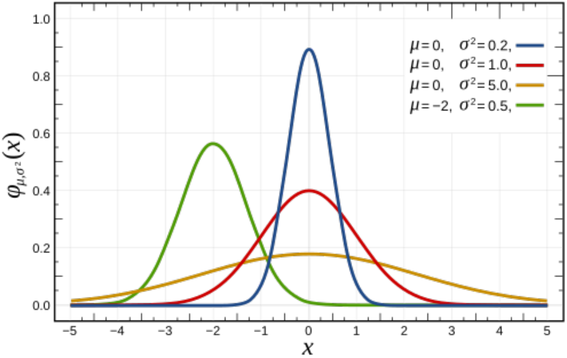
Positive (car l'exponentielle est toujours positive et λ aussi)
Continue sauf en 0.

$$\begin{aligned}
\int_{x \in \mathcal{X}} f(x)\, dx &= \int_{0}^{+\infty} \lambda e^{-\lambda x}\, dx \\
&= \left[-e^{-\lambda x}\right]_0^{+\infty} \\
&= \lim_{x \to +\infty} \left(-e^{-\lambda \times x}\right) - \left(-e^{-\lambda \times 0}\right) \\
&= 0 - (-1) \\
&= 1
\end{aligned}$$

On est donc bien en présence d'une densité de probabilités.

X suit une loi normale (m, σ^2) de densité :

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \forall x \in \mathbb{R}$$



Caractérisez les fonctions de répartition $F_1(x) = \mathbb{P}(X < x)$ et $F_2(x) = \mathbb{P}(X \leqslant x)$.

Ici pour une loi de probabilité donnée, $\forall x \in \mathcal{X},\ \ \mathbb{P}(X = x)$, deux définitions de la fonction de répartition sont possibles :

- $F_1(x) = \mathbb{P}(X < x)$: cumul des probabilités pour les événements inférieurs strictement à x. Ainsi, la probabilité de l'événement $X < x$ représente la probabilité de l'union des événements (incompatibles) pour lesquels X prend les valeurs strictement inférieures à x (donc x est exclu) dans l'ensemble des possibles de la loi :

$$F_1(x) = \mathbb{P}(X < x) = \sum_{k<x} \mathbb{P}(X = k)$$

- $F_2(x) = \mathbb{P}(X \leqslant x)$: cumul des probabilités pour les événements inférieurs ou égaux à x. Ainsi, la probabilité de l'événement $X \leqslant x$ représente la probabilité de l'union des événements (incompatibles) pour lesquels X prend les valeurs inférieures ou égales à x (donc x est inclus) dans l'ensemble des possibles de la loi :

$$F_2(x) = \mathbb{P}(X \leqslant x) = \sum_{k \leqslant x} \mathbb{P}(X = k)$$

Comme l'événement $(X \leqslant x) = (X < x) \cup (X = x)$, l'union de deux événements incompatibles, il vient que :

$$\begin{aligned} \mathbb{P}(X \leqslant x) &= \mathbb{P}\left[(X < x) \cup (X = x)\right] \\ &= \mathbb{P}(X < x) + \mathbb{P}(X = x) \end{aligned}$$

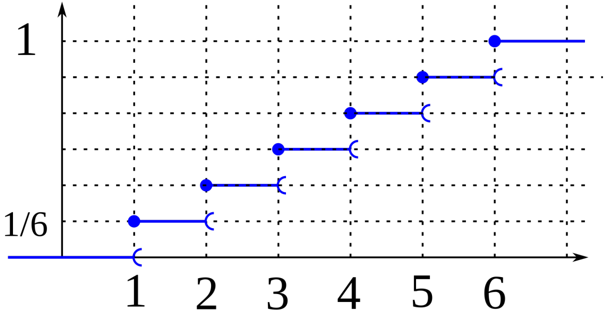
soit la relation entre les deux fonctions :

$$F_2(x) = F_1(x) + \mathbb{P}(X = x)$$

Il faut également faire attention au choix de l'indice de sommation : ici on ne peut plus utiliser x.

Les représenter graphiquement si X suit une loi de Bernoulli ou pour un tirage de dé standard.

Fonction de répartition discrète :
$\mathbb{P}(X \leqslant x) = \sum_{k \leqslant x} \mathbb{P}(X = k)$
avec $\mathbb{P}(X = k) = \frac{1}{6}$ et $k \in [1; 6]$



La fonction de densité $f(.)$ est la dérivée de la fonction de répartition $F(.)$.

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

$$\begin{aligned}
& F'(u) = f(u) \\
\Leftrightarrow & \int_{-\infty}^{x} F'(u)du = \int_{-\infty}^{x} f(u)du \\
\Leftrightarrow & [F(u)]_{-\infty}^{x} = \int_{-\infty}^{x} f(u)du \\
\Leftrightarrow & F(x) - \lim_{x \to -\infty} F(x) = \int_{-\infty}^{x} f(u)du \\
\Leftrightarrow & F(x) - \lim_{x \to -\infty} \mathbb{P}(X < x) = \int_{-\infty}^{x} f(u)du \\
\Leftrightarrow & F(x) = \int_{-\infty}^{x} f(u)du
\end{aligned}$$

C'est une fonction croissante (puisque sa dérivée est positive).

Elle prend des valeurs comprises entre $0 (= \lim_{x \to -\infty} F(x))$ et $1 (= \lim_{x \to +\infty} F(x))$ puisque c'est une probabilité.

Soit X une variable aléatoire continue définie sur $x \in \mathcal{X}$.

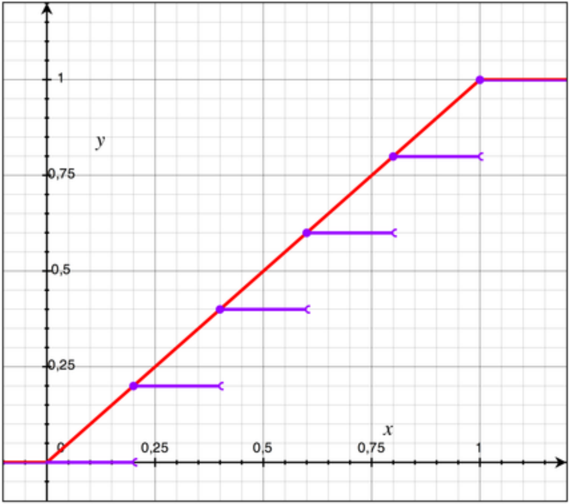
Caractérisez les fonctions de répartition $F_1(x) = \mathbb{P}(X < x)$ et $F_2(x) = \mathbb{P}(X \leqslant x)$.

Ici $F_1(x) = F_2(x)$ puisque $\mathbb{P}(X < x) = \mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} f(u)\, du$ dans le cas continu.

Donc, contrairement au cas discret, il est inutile de faire la distinction.

Fonctions de répartition de la loi uniforme : VAD/VAC

Lois uniformes discrète $\mathbb{P}(X = k) = \dfrac{1}{5}$ et continue $F(x) = \begin{cases} 0 & x < 0 \\ x & x \in [0, 1] \\ 1 & x > 1 \end{cases}$



Fonction de répartition de la loi uniforme continue

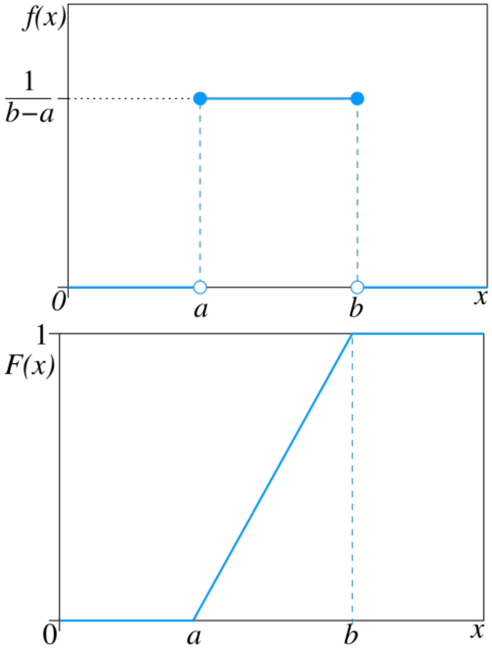
$$f(x) = \begin{cases} \frac{1}{b-a} & \forall x \in [a, b] \\ 0 & \text{sinon} \end{cases}$$

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(u)\, du \\ &= \int_{a}^{x} \frac{1}{b-a}\, du \qquad \text{si } x \in [a, b] \\ &= \left[\frac{u}{b-a}\right]_a^x \\ &= \frac{x}{b-a} - \frac{a}{b-a} \\ &= \frac{x-a}{b-a} \end{aligned}$$

Attention : la solution doit être déterminée sur $\mathbb{R}$ pour être complète.

$$F(x) = \mathbb{P}(X \leqslant x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & x \in [a, b] \\ 1 & x > b \end{cases}$$

Fonction de répartition de la loi uniforme continue



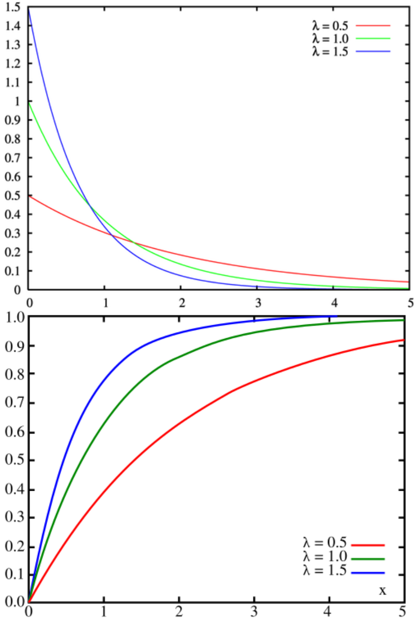
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \forall x > 0, \lambda > 0 \\ 0 & \text{sinon} \end{cases}$$

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(u)\, du \\ &= \int_{0}^{x} \lambda e^{-\lambda u}\, du \quad \text{si } x > 0 \\ &= \left[-e^{-\lambda u}\right]_0^x \\ &= -e^{-\lambda x} - \left(-e^{-\lambda 0}\right) \\ &= -e^{-\lambda x} + 1 \end{aligned}$$

Attention : la solution doit être déterminée sur $\mathbb{R}$ pour être complète.

$$F(x) = \mathbb{P}(X \leqslant x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & \text{sinon} \end{cases}$$

Fonction de répartition de la loi exponentielle

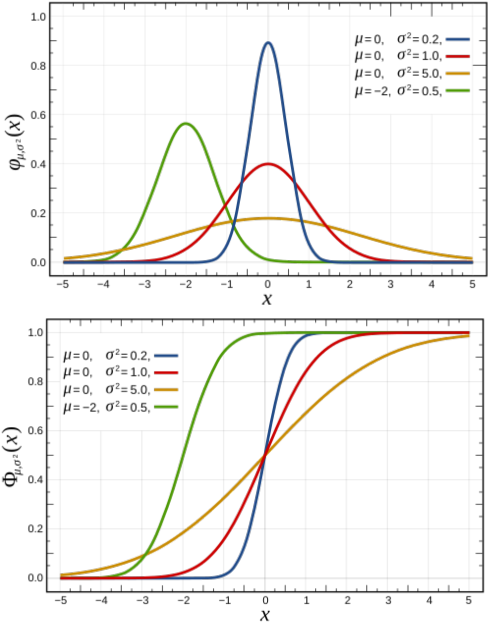


$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \forall x \in \mathbb{R}$$

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} \phi(u)\ du \\ &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(u-m)^2}{2\sigma^2}}\ du \end{aligned}$$

Pas de solution explicite.

Fonction de répartition de la loi Normale



Il faut utiliser la table de la loi normale centrée réduite $N(0,1)$.

On utilise la relation : si $Y \rightsquigarrow N(m,\sigma^2)$, si $X \rightsquigarrow N(0,1)$, alors :

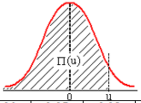
$$Y = m + \sigma X \Leftrightarrow X = \frac{Y-m}{\sigma}$$

Cette table est particulière car uniquement pour des valeurs de $x \geqslant 0$ car la loi est symétrique.

Les probabilités pour les valeurs $x < 0$ se déduisent par le raisonnement graphique utilisant la symétrie.

Calculer une probabilité avec la loi Normale

Table de Loi Normale
P(x<u)



	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	.0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8254	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	.0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0.9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	6,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	.0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

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

$$F(x) = \mathbb{P}(X \leqslant x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a,b] \\ 1 & x > b \end{cases}$$

$$\mathbb{P}(X < a) = F(a) = 0$$

$$\mathbb{P}(X > b) = 1 - \mathbb{P}(X < b) = 1 - F(b) = 1 - 1 = 0$$

$$\mathbb{P}(a < X < b) = \mathbb{P}(X < b) - \mathbb{P}(X < a) = F(b) - F(a) = \frac{b-a}{b-a} - \frac{a-a}{b-a} = 1$$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \forall x > 0, \lambda > 0 \\ 0 & \text{sinon} \end{cases}$$

$$F(x) = \mathbb{P}(X \leqslant x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & \text{sinon} \end{cases}$$

$$\mathbb{P}(X < a) = \begin{cases} 1 - e^{-\lambda a} & \text{si } a > 0 \\ 0 & \text{sinon} \end{cases}$$

$$\mathbb{P}(X > b) = 1 - \mathbb{P}(X < b) = 1 - F(b) = \begin{cases} 1 - (1 - e^{-\lambda b}) = e^{-\lambda b} & \text{si } b > 0 \\ 1 - 0 = 1 & \text{sinon} \end{cases}$$

Tout dépend de la position de a et de b par rapport à 0.

$$\begin{aligned}
\mathbb{P}(a < X < b) &= \mathbb{P}(X < b) - \mathbb{P}(X < a) \\
&= F(b) - F(a) \\
&= \begin{cases} 0 - 0 & \text{si } a < 0, b < 0 \\ (1 - e^{-\lambda b}) - 0 & \text{si } a < 0, b > 0 \\ (1 - e^{-\lambda b}) - (1 - e^{-\lambda a}) & \text{si } a > 0, b > 0 \end{cases} \\
&= \begin{cases} 0 & \text{si } a < 0, b < 0 \\ 1 - e^{-\lambda b} & \text{si } a < 0, b > 0 \\ e^{-\lambda a} - e^{-\lambda b} & \text{si } a > 0, b > 0 \end{cases}
\end{aligned}$$

VII- Rappels sur les principaux moments des variables aléatoires

$$\mathbb{E}(X) = \sum_{x \in \mathcal{X}} x \, \mathbb{P}(X = x)$$

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

Ce sont des constantes non aléatoires.

De manière générale, pour toute fonction $\varphi(.)$:

$$\mathbb{E}[\varphi(X)] = \sum_{x \in \mathcal{X}} \varphi(x)\ \mathbb{P}(X = x)$$

Ainsi par exemple :

$$\mathbb{E}[X^2] = \sum_{x \in \mathcal{X}} x^2\ \mathbb{P}(X = x)$$

$$\mathbb{E}[X^3] = \sum_{x \in \mathcal{X}} x^3\ \mathbb{P}(X = x)$$

$$\mathbb{E}[X^4] = \sum_{x \in \mathcal{X}} x^4\ \mathbb{P}(X = x)$$

$$\mathbb{E}[(X - \mathbb{E}(X))^2] = \sum_{x \in \mathcal{X}} [x - \mathbb{E}(X)]^2\ \mathbb{P}(X = x) = \mathbb{V}(X)$$

Application :

x	1	2	3	4	5
$\mathbb{P}(X = x)$	0.2	0.2	0.2	0.2	0.2

Ici on a : $x \in \mathcal{X} = \{1, \ldots, 5\}$ et $\mathbb{P}(X = x) = \frac{1}{5} \in [0, 1]$. De plus :

$$\begin{aligned}\sum_{x \in \mathcal{X}} \mathbb{P}(X = x) &= \sum_{x=1}^{5} \frac{1}{5} \\ &= \frac{1}{5} \times 5 = 1\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{x \in \mathcal{X}} x \, \mathbb{P}(X = x) \\
&= \sum_{x=1}^{5} x \times \frac{1}{5} \\
&= \frac{1}{5} \underbrace{\sum_{x=1}^{5} x} \\
&= \frac{1}{5} \frac{5 \times 6}{2} \\
&= 3
\end{aligned}
$$

si vous vous souvenez que $\sum_{x=1}^{N} x = \dfrac{N(N+1)}{2}$ (la somme des termes d'une suite arithmétique de raison 1).

$$
\begin{aligned}
\mathbb{V}(X) &= \sum_{x \in \mathcal{X}} [x - \mathbb{E}(X)]^2 \; \mathbb{P}(X = x) \\
&= \sum_{x=1}^{5} [x - 3]^2 \times \frac{1}{5} \\
&= \frac{1}{5} \sum_{x=1}^{5} [x - 3]^2 \\
&= \frac{1}{5} \left[(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2 \right] \\
&= \frac{1}{5} \, [4 + 1 + 0 + 1 + 4] \; = 2
\end{aligned}
$$

Ici on reconnaît la loi uniforme discrète : $x \in \mathcal{X} = \{1, \ldots, n\}$ et $\mathbb{P}(X = x) = \frac{1}{n}$.

Tous les résultats possibles de l'expérience ont la même probabilité.

Soit X une variable aléatoire discrète de loi de probabilités $\mathbb{P}(X = x)$ définie pour $x \in \mathcal{X}$. Soient a et b deux constantes réelles.

Montrez que l'espérance est un opérateur linéaire : $\mathbb{E}(aX + b) = a\,\mathbb{E}(X) + b$.

Utilisons :

$$\mathbb{E}[\varphi(X)] = \sum_{x \in \mathcal{X}} \varphi(x)\ \mathbb{P}(X = x)$$

en posant $\varphi(X) = aX + b$ (du coup $\varphi(x) = ax + b$).

$$\begin{aligned}
\mathbb{E}(aX + b) &= \sum_{x \in \mathcal{X}} (ax + b)\ \mathbb{P}(X = x) \\
&= \sum_{x \in \mathcal{X}} [ax \times \mathbb{P}(X = x) + b\mathbb{P}(X = x)] \\
&= \sum_{x \in \mathcal{X}} ax \times \mathbb{P}(X = x) + \sum_{x \in \mathcal{X}} b\mathbb{P}(X = x) \\
&= a \underbrace{\sum_{x \in \mathcal{X}} x \times \mathbb{P}(X = x)}_{=\mathbb{E}(X)} + b \underbrace{\sum_{x \in \mathcal{X}} \mathbb{P}(X = x)}_{=1} \\
&= a\, \mathbb{E}(X) + b
\end{aligned}$$

Calculez l'espérance de $X - \mathbb{E}(X)$.

Utilisons $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ en posant $a = 1$ et $b = -\mathbb{E}(X)$.

$$\begin{aligned}\mathbb{E}\left[X - \mathbb{E}(X)\right] &= \mathbb{E}(X) - \mathbb{E}(X)\\ &= 0\end{aligned}$$

$X - \mathbb{E}(X)$ est la variable aléatoire X qu'on a centrée. Son espérance est donc 0.

Montrez que la variance est un opérateur quadratique : $\mathbb{V}(aX+b) = a^2\mathbb{V}(X)$.

La variance de X s'écrit :

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

On a donc :

$$\mathbb{V}(aX+b) = \sum_{x \in \mathcal{X}} [ax + b - \mathbb{E}(aX+b)]^2 \, \mathbb{P}(X = x)$$

ou alors :

$$\begin{aligned} \mathbb{V}(aX+b) &= \mathbb{E}[aX + b - \mathbb{E}(aX+b)]^2 \\ &= \sum_{x \in \mathcal{X}} [ax + b - \mathbb{E}(aX+b)]^2 \, \mathbb{P}(X = x) \end{aligned}$$

en posant $\varphi(X) = [aX + b - \mathbb{E}(aX+b)]^2$ (du coup $\varphi(x) = [ax + b - \mathbb{E}(aX+b)]^2$).

Concentrons-nous sur le carré. Comme $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$:

$$\begin{aligned}
[ax + b - \mathbb{E}(aX + b)]^2 &= [ax + b - a\mathbb{E}(X) - b]^2 \\
&= [ax - a\mathbb{E}(X)]^2 \\
&= [a(x - \mathbb{E}(X))]^2 \\
&= a^2 [x - \mathbb{E}(X)]^2
\end{aligned}$$

On remplace :

$$\begin{aligned}
\mathbb{V}(aX + b) &= \sum_{x \in \mathcal{X}} [ax + b - \mathbb{E}(aX + b)]^2 \, \mathbb{P}(X = x) \\
&= \sum_{x \in \mathcal{X}} a^2 [x - \mathbb{E}(X)]^2 \, \mathbb{P}(X = x) \\
&= a^2 \underbrace{\sum_{x \in \mathcal{X}} [x - \mathbb{E}(X)]^2 \, \mathbb{P}(X = x)} \\
&= a^2 \, \mathbb{V}(X)
\end{aligned}$$

Calculez la variance de $\dfrac{X}{\sqrt{\mathbb{V}(X)}}$.

Utilisons $\mathbb{V}(aX+b) = a^2\mathbb{V}(X)$ en posant $a = \frac{1}{\sqrt{\mathbb{V}(X)}}$ et $b = 0$.

$$\begin{aligned} \mathbb{V}\left(\frac{X}{\sqrt{\mathbb{V}(X)}}\right) &= \left(\frac{1}{\sqrt{\mathbb{V}(X)}}\right)^2 \mathbb{V}(X) \\ &= \frac{1}{\mathbb{V}(X)}\mathbb{V}(X) \\ &= 1 \end{aligned}$$

$\dfrac{X}{\sqrt{\mathbb{V}(X)}}$ est la variable X réduite, c'est-à-dire divisée par son écart-type. La variance d'une variable réduite est donc égale à 1.

Montrez qu'on peut écrire : $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$.

La variance s'écrit :

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

Développons le carré et distribuons :

$$\begin{aligned}
\mathbb{V}(X) &= \sum_{x \in \mathcal{X}} [x - \mathbb{E}(X)]^2 \, \mathbb{P}(X = x) \\
&= \sum_{x \in \mathcal{X}} [x^2 - 2x\mathbb{E}(X) + \mathbb{E}^2(X)] \, \mathbb{P}(X = x) \\
&= \sum_{x \in \mathcal{X}} [x^2 \mathbb{P}(X = x) - 2x\mathbb{E}(X)\mathbb{P}(X = x) + \mathbb{E}^2(X)\mathbb{P}(X = x)] \\
&= \sum_{x \in \mathcal{X}} x^2 \, \mathbb{P}(X = x) + \sum_{x \in \mathcal{X}} -2x\mathbb{E}(X) \, \mathbb{P}(X = x) + \sum_{x \in \mathcal{X}} \mathbb{E}^2(X) \, \mathbb{P}(X = x) \\
&= \underbrace{\sum_{x \in \mathcal{X}} x^2 \, \mathbb{P}(X = x)}_{=\mathbb{E}(X^2)} - 2\mathbb{E}(X) \underbrace{\sum_{x \in \mathcal{X}} x \, \mathbb{P}(X = x)}_{=\mathbb{E}(X)} + \mathbb{E}^2(X) \underbrace{\sum_{x \in \mathcal{X}} \mathbb{P}(X = x)}_{=1} \\
&= \mathbb{E}(X^2) - 2\mathbb{E}^2(X) + \mathbb{E}^2(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)
\end{aligned}$$

Montrez qu'on peut également écrire : $\mathbb{V}(X) = \mathbb{E}\{[X(X-1)]\} + \mathbb{E}(X) - \mathbb{E}^2(X)$.
Repartons du dernier résultat :

$$\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}^2(X) \\
&= \mathbb{E}(X^2 \underbrace{-X+X}_{=0}) - \mathbb{E}^2(X) \\
&= \mathbb{E}(\underbrace{X^2 - X} + X) - \mathbb{E}^2(X) \\
&= \mathbb{E}\left[X(X-1) + X\right] - \mathbb{E}^2(X) \\
&= \mathbb{E}\left[X(X-1)\right] + \mathbb{E}(X) - \mathbb{E}^2(X)
\end{aligned}$$

Le passage de $\mathbb{E}\left[X(X-1)+X\right] = \mathbb{E}\left[X(X-1)\right] + \mathbb{E}(X)$ n'est pas trivial mais n'est pas difficile non plus en repassant par les sommes :

$$\begin{aligned}
\mathbb{E}\left[\underbrace{X(X-1)+X}_{\varphi(X)}\right] &= \sum_{x\in\mathcal{X}} [\underbrace{x(x-1)+x}_{=\varphi(x)}]\, \mathbb{P}(X=x) \\
&= \sum_{x\in\mathcal{X}} [x(x-1)\mathbb{P}(X=x) + x\mathbb{P}(X=x)] \\
&= \sum_{x\in\mathcal{X}} x(x-1)\mathbb{P}(X=x) + \sum_{x\in\mathcal{X}} x\mathbb{P}(X=x) \\
&= \mathbb{E}\left[X(X-1)\right] + \mathbb{E}(X)
\end{aligned}$$

L'espérance :

$$\text{VAD : } \mathbb{E}(X) = \sum_{x \in \mathcal{X}} x\mathbb{P}(X = x)$$

$$\text{VAC : } \mathbb{E}(X) = \int_{x \in \mathcal{X}} xf(x)\ dx$$

La variance :

$$\text{VAD : } \mathbb{V}(X) = \sum_{x \in \mathcal{X}} [x - \mathbb{E}(X)]^2\ \mathbb{P}(X = x)$$

$$\text{VAC : } \mathbb{V}(X) = \int_{x \in \mathcal{X}} [x - \mathbb{E}(X)]^2\ f(x)\ dx$$

Ce sont des constantes non aléatoires.

De manière générale, pour toute fonction $\varphi(.)$:

$$\mathbb{E}[\varphi(X)] = \int_{x \in \mathcal{X}} \varphi(x)\ f(x)\ dx$$

Ainsi par exemple :

$$\begin{aligned}
\mathbb{E}[X^2] &= \int_{x \in \mathcal{X}} x^2\ f(x)\ dx \\
\mathbb{E}[X^3] &= \int_{x \in \mathcal{X}} x^3\ f(x)\ dx \\
\mathbb{E}[X^4] &= \int_{x \in \mathcal{X}} x^4\ f(x)\ dx \\
\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] &= \int_{x \in \mathcal{X}} [x - \mathbb{E}(X)]^2\ f(x)\ dx
\end{aligned}$$

Soit X une variable aléatoire continue de densité de probabilités $f(x)$ définie pour $x \in \mathcal{X}$. Soient a et b deux constantes réelles.

La bonne nouvelle : ce sont les mêmes que pour les VAD (car une intégrale c'est une somme).

Montrez que l'espérance est un opérateur linéaire : $\mathbb{E}(aX + b) = a\,\mathbb{E}(X) + b$.

Utilisons :

$$\mathbb{E}[\varphi(X)] = \int_{x \in \mathcal{X}} \varphi(x)\ f(x)\ dx$$

en posant $\varphi(X) = aX + b$ (du coup $\varphi(x) = ax + b$).

$$
\begin{aligned}
\mathbb{E}(aX+b) &= \int_{x\in\mathcal{X}} (ax+b)\ f(x)\ dx \\
&= \int_{x\in\mathcal{X}} [ax\ f(x)\ dx + bf(x)\ dx] \\
&= \int_{x\in\mathcal{X}} ax\ f(x)\ dx + \int_{x\in\mathcal{X}} bf(x)\ dx \\
&= a \underbrace{\int_{x\in\mathcal{X}} x\ f(x)\ dx}_{=\mathbb{E}(X)} + b \underbrace{\int_{x\in\mathcal{X}} f(x)\ dx}_{=1} \\
&= a\,\mathbb{E}(X) + b
\end{aligned}
$$

Calculez l'espérance de $X - \mathbb{E}(X)$.

Utilisons $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ en posant $a = 1$ et $b = -\mathbb{E}(X)$.

$$\begin{aligned}\mathbb{E}\left[X - \mathbb{E}(X)\right] &= \mathbb{E}(X) - \mathbb{E}(X) \\ &= 0\end{aligned}$$

$X - \mathbb{E}(X)$ est la variable aléatoire X qu'on a centrée. Son espérance est donc 0.

Montrez que la variance est un opérateur quadratique : $\mathbb{V}(aX+b) = a^2\mathbb{V}(X)$.

La variance de X s'écrit :

$$\mathbb{V}(X) = \int_{x\in\mathcal{X}} [x - \mathbb{E}(X)]^2 \, f(x) \, dx$$

On a donc :

$$\mathbb{V}(aX+b) = \int_{x\in\mathcal{X}} [ax + b - \mathbb{E}(aX+b)]^2 \, f(x) \, dx$$

Comme $\mathbb{E}(aX+b) = a\,\mathbb{E}(X) + b$:

$$\begin{aligned}
[ax+b-\mathbb{E}(aX+b)]^2 &= [ax+b-a\mathbb{E}(X)-b]^2 \\
&= [ax-a\mathbb{E}(X)]^2 \\
&= [a(x-\mathbb{E}(X))]^2 \\
&= a^2\,[x-\mathbb{E}(X)]^2
\end{aligned}$$

On remplace :

$$\begin{aligned}
\mathbb{V}(aX+b) &= \int_{x\in\mathcal{X}} [ax+b-\mathbb{E}(aX+b)]^2\ f(x)\ dx \\
&= \int_{x\in\mathcal{X}} a^2\,[x-\mathbb{E}(X)]^2\ f(x)\ dx \\
&= a^2\ \underbrace{\int_{x\in\mathcal{X}} [x-\mathbb{E}(X)]^2\ f(x)\ dx} \\
&= a^2\ \mathbb{V}(X)
\end{aligned}$$

Calculez la variance de $\dfrac{X}{\sqrt{\mathbb{V}(X)}}$.

Utilisons $\mathbb{V}(aX+b) = a^2\mathbb{V}(X)$ en posant $a = \frac{1}{\sqrt{\mathbb{V}(X)}}$ et $b = 0$.

$$\begin{aligned}\mathbb{V}\left(\frac{X}{\sqrt{\mathbb{V}(X)}}\right) &= \left(\frac{1}{\sqrt{\mathbb{V}(X)}}\right)^2 \mathbb{V}(X)\\ &= \frac{1}{\mathbb{V}(X)}\mathbb{V}(X)\\ &= 1\end{aligned}$$

$\dfrac{X}{\sqrt{\mathbb{V}(X)}}$ est la variable X réduite, c'est-à-dire divisée par son écart-type. La variance d'une variable réduite est donc égale à 1.

Montrez qu'on peut écrire : $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$.

La variance s'écrit :

$$\mathbb{V}(X) = \int_{x \in \mathcal{X}} [x - \mathbb{E}(X)]^2 \ f(x) \ dx$$

Développons le carré et distribuons :

$$\begin{aligned}
\mathbb{V}(X) &= \int_{x \in \mathcal{X}} [x - \mathbb{E}(X)]^2 \ f(x) \ dx \\
&= \int_{x \in \mathcal{X}} [x^2 - 2x\mathbb{E}(X) + \mathbb{E}^2(X)] \ f(x) \ dx \\
&= \int_{x \in \mathcal{X}} [x^2 f(x) \ dx - 2x\mathbb{E}(X) f(x) \ dx + \mathbb{E}^2(X) f(x) \ dx] \\
&= \int_{x \in \mathcal{X}} x^2 \ f(x) \ dx + \int_{x \in \mathcal{X}} -2x\mathbb{E}(X) \ f(x) \ dx + \int_{x \in \mathcal{X}} \mathbb{E}^2(X) \ f(x) \ dx \\
&= \underbrace{\int_{x \in \mathcal{X}} x^2 \ f(x) \ dx}_{=\mathbb{E}(X^2)} - 2\mathbb{E}(X) \underbrace{\int_{x \in \mathcal{X}} x \ f(x) \ dx}_{=\mathbb{E}(X)} + \mathbb{E}^2(X) \underbrace{\int_{x \in \mathcal{X}} f(x) \ dx}_{=1} \\
&= \mathbb{E}(X^2) - 2\mathbb{E}^2(X) + \mathbb{E}^2(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)
\end{aligned}$$

Pour la loi uniforme $\forall k \in \mathcal{X} = [1, \ldots, n], \;\; \mathbb{P}(X = k) = \frac{1}{n}$:

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{k \in \mathcal{X}} k \; \mathbb{P}(X = k) \\
&= \sum_{k=1}^{n} k \; \frac{1}{n} \\
&= \frac{1}{n} \; \sum_{k=1}^{n} k \\
&= \frac{1}{n} \frac{n(n+1)}{2} \\
&= \frac{n+1}{2}
\end{aligned}
$$

Pour la loi de Bernoulli : $x = 0, 1 \quad \mathbb{P}(X = x) = p^x \ (1-p)^{(1-x)}$:

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x \in \mathcal{X}} x \, \mathbb{P}(X = x) \\
&= \sum_{x=0}^{1} x \, p^x \ (1-p)^{(1-x)} \\
&= \underbrace{0 \, p^0 \ (1-p)^{(1-0)}}_{\text{si } x=0} + \underbrace{1 \, p^1 \ (1-p)^{(1-1)}}_{\text{si } x=1} \\
&= p
\end{aligned}$$

Pour la loi Binomiale : $\forall k \in [0, \dots, n],\ \mathbb{P}(X = k) = C_n^k\ p^k\ (1-p)^{(n-k)}$
(indication : $kC_n^k = nC_{n-1}^{k-1}$) :

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k \in \mathcal{X}} k\ \mathbb{P}(X = k) \\
&= \sum_{k=0}^{n} k\ C_n^k\ p^k\ (1-p)^{(n-k)} \\
&= \sum_{\color{red}{k=1}}^{n} k\ C_n^k\ p^k\ (1-p)^{(n-k)}
\end{aligned}$$

Utilisons : $kC_n^k = nC_{n-1}^{k-1}$ Remarquons que :

$$\begin{aligned}
kC_n^k &= k\,\frac{n!}{k!(n-k)!} \\
&= k\,\frac{n!}{k \times (k-1)!(n-k)!} \\
&= \frac{n!}{(k-1)!(n-k)!} \\
&= \frac{n \times (n-1)!}{(k-1)!(n-k\underbrace{+1-1}_{=0})!} \\
&= n\,\underbrace{\frac{(n-1)!}{(k-1)!(n-1-(k-1))!}} \\
&= nC_{n-1}^{k-1}
\end{aligned}$$

Ce résultat est intéressant et pourra être réutilisé : $kC_n^k = nC_{n-1}^{k-1}$, $(k-1)C_{n-1}^{k-1} = (n-1)C_{n-2}^{k-2}$, $(k-2)C_{n-2}^{k-2} = (n-2)C_{n-3}^{k-3}$, ...

On remplace dans l'expression qui nous intéresse pour se rapprocher du binôme de Newton :

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=1}^{n} k\ C_n^k\ p^k\ (1-p)^{(n-k)} \\ &= \sum_{k=1}^{n} nC_{n-1}^{k-1}\ p^k\ (1-p)^{(n-k)} \\ &= n\sum_{k=1}^{n} C_{n-1}^{k-1}\ p^k\ (1-p)^{(n-k)}\end{aligned}$$

On y est presque :

$$\mathbb{E}(X) = n \sum_{k=1}^{n} C_{n-1}^{k-1}\ p^k\ (1-p)^{(n-k)}$$

Posons le changement d'indice suivant : $i = k - 1 \Leftrightarrow\ k = i + 1$.
Quand $k = 1$, $i = 0$.
Quand $k = n$, $i = n - 1$.
Il ne reste plus qu'à remplacer tous les $k - 1$ dans l'expression :

$$\begin{aligned}\mathbb{E}(X) &= n \sum_{i=0}^{n-1} C_{n-1}^{i}\ p^{i+1}\ (1-p)^{n-(1+i)} \\ &= np \underbrace{\sum_{i=0}^{n-1} C_{n-1}^{i}\ p^{i}\ (1-p)^{(n-1)-i}}\end{aligned}$$

On reconnaît le binôme de Newton réécrit pour la puissance $n - 1$. On a donc :

$$\begin{aligned}\mathbb{E}(X) &= np\,[p + (1-p)]^{n-1} \\ &= np \times 1^{n-1} \\ &= np\end{aligned}$$

Espérance de la loi de Poisson

Pour la loi de Poisson : $\forall k \in [0, +\infty[, \quad \mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$:

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k \in \mathcal{X}} k \, \mathbb{P}(X = k) \\
&= \sum_{k=0}^{+\infty} k \; e^{-\lambda} \frac{\lambda^k}{k!} \\
&= \sum_{k=1}^{+\infty} k \; e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=1}^{+\infty} k \frac{\lambda^k}{k \times (k-1)!} \\
&= e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} \\
&= e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda \times \lambda^{k-1}}{(k-1)!} \\
&= e^{-\lambda} \lambda \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!}
\end{aligned}$$

$$\mathbb{E}(X) = e^{-\lambda}\lambda \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

A nouveau petit changement d'indice : $i = k - 1$.
Quand $k = 1$, $i = 0$.
Quand $k = +\infty$, $i = +\infty$.
Il ne reste plus qu'à remplacer les $k - 1$ dans l'expression pour utiliser le DL :

$$\begin{aligned}\mathbb{E}(X) &= e^{-\lambda}\lambda \underbrace{\sum_{i=0}^{+\infty} \frac{\lambda^{i}}{i!}} \\ &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda\end{aligned}$$

On va commencer par utiliser $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$ car on a déjà calculé $\mathbb{E}(X)$.

Pour les deux dernières lois, on verra qu'il est plus simple d'utiliser $\mathbb{V}(X) = \mathbb{E}\left[X(X-1)\right] + \mathbb{E}(X) - \mathbb{E}^2(X)$.

Pour la loi uniforme $\forall k \in \mathcal{X} = [1, \ldots, n], \;\; \mathbb{P}(X = k) = \frac{1}{n}$:

Commençons par calculer $\mathbb{E}(X^2)$.

$$
\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k \in \mathcal{X}} k^2 \, \mathbb{P}(X = k) \\
&= \sum_{k=1}^{n} k^2 \, \frac{1}{n} \\
&= \frac{1}{n} \, \sum_{k=1}^{n} k^2 \\
&= \frac{1}{n} \frac{n(n+1)(2n+1)}{6} \\
&= \frac{(n+1)(2n+1)}{6}
\end{aligned}
$$

Pour la variance :

$$
\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}^2(X) \\
&= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 \\
&= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \\
&= \frac{2(n+1)(2n+1) - 3(n+1)^2}{12} \\
&= \frac{(n+1)\left[2(2n+1) - 3(n+1)\right]}{12} \\
&= \frac{(n+1)\left[4n+2-3n-3\right]}{12} \\
&= \frac{(n+1)(n-1)}{12} \\
&= \frac{n^2-1}{12}
\end{aligned}
$$

Pour la loi de Bernoulli : $x = 0, 1 \quad \mathbb{P}(X = x) = p^x\ (1-p)^{(1-x)}$:

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{x \in \mathcal{X}} x^2\ \mathbb{P}(X = x) \\
&= \sum_{x=0}^{1} x^2\ p^x\ (1-p)^{(1-x)} \\
&= 0^2\ p^0\ (1-p)^{(1-0)} + 1^2\ p^1\ (1-p)^{(1-1)} \\
&= p
\end{aligned}$$

Pour la variance :

$$\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}^2(X) \\
&= p - p^2 \\
&= p(1-p)
\end{aligned}$$

Pour la loi Binomiale : $\forall k \in [0, \ldots, n]$, $\mathbb{P}(X = k) = C_n^k\, p^k\, (1-p)^{(n-k)}$,
utilisons $\mathbb{V}(X) = \mathbb{E}\left[X(X-1)\right] + \mathbb{E}(X) - \mathbb{E}^2(X)$.

$$\begin{aligned}
\mathbb{E}\left[X(X-1)\right] &= \sum_{k \in \mathcal{X}} k(k-1)\, \mathbb{P}(X = k) \\
&= \sum_{k=0}^{n} k(k-1)\, C_n^k\, p^k\, (1-p)^{(n-k)} \\
&= \sum_{k=2}^{n} k(k-1)\, C_n^k\, p^k\, (1-p)^{(n-k)}
\end{aligned}$$

car pour $k = 0$ et $k = 1$, le terme générique s'annule.

Il est facile de réutiliser l'astuce de l'espérance et de montrer que :

$$kC_n^k = nC_{n-1}^{k-1}$$

$$(k-1)C_{n-1}^{k-1} = (n-1)C_{n-2}^{k-2}$$

Remplaçons :

$$\begin{aligned}
\mathbb{E}\left[X(X-1)\right] &= \sum_{k=2}^{n} k(k-1)\ C_n^k\ p^k\ (1-p)^{n-k} \\
&= \sum_{k=2}^{n} n(n-1)C_{n-2}^{k-2}\ p^k\ (1-p)^{n-k} \\
&= n(n-1)\sum_{k=2}^{n} C_{n-2}^{k-2}\ p^k\ (1-p)^{n-k}
\end{aligned}$$

$$\mathbb{E}\left[X(X-1)\right] = n(n-1)\sum_{k=2}^{n} C_{n-2}^{k-2}\ p^k\ (1-p)^{n-k}$$

Pour se ramener au binôme de Newton cette fois pour une puissance $n-2$, on fait le changement d'indice suivant : $i = k-2$. Quand $k = 2$, $i = 0$. Quand $k = n$, $i = n-2$. Il ne reste plus qu'à remplacer dans l'expression :

$$\begin{aligned}
\mathbb{E}\left[X(X-1)\right] &= n(n-1)\sum_{k=2}^{n} C_{n-2}^{k-2}\ p^k\ (1-p)^{n-k} \\
&= n(n-1)\sum_{i=0}^{n-2} C_{n-2}^{i}\ p^{i+2}\ (1-p)^{n-(i+2)} \\
&= n(n-1)\sum_{i=0}^{n-2} C_{n-2}^{i}\ p^{i}p^{2}\ (1-p)^{n-2-i} \\
&= n(n-1)p^2\underbrace{\sum_{i=0}^{n-2} C_{n-2}^{i}\ p^{i}\ (1-p)^{(n-2)-i}} \\
&= n(n-1)p^2 \times \underbrace{(p+1-p)^{n-2}}_{=1} \\
&= n(n-1)p^2
\end{aligned}$$

Reste à calculer la variance :

$$\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}\left[X(X-1)\right] + \mathbb{E}(X) - \mathbb{E}^2(X) \\
&= n(n-1)p^2 + np - (np)^2 \\
&= n^2p^2 - np^2 + np - n^2p^2 \\
&= np - np^2 \\
&= np(1-p)
\end{aligned}$$

Pour la loi de Poisson : $\forall k \in [0, +\infty[, \quad \mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$:

$$\begin{aligned}
\mathbb{E}\left[X(X-1)\right] &= \sum_{k \in \mathcal{X}} k(k-1)\, \mathbb{P}(X = k) \\
&= \sum_{k=0}^{+\infty} k(k-1)\, e^{-\lambda} \frac{\lambda^k}{k!} \\
&= \sum_{k=2}^{+\infty} k(k-1)\, e^{-\lambda} \frac{\lambda^k}{k!} \\
&= \sum_{k=2}^{+\infty} k(k-1)\, e^{-\lambda} \frac{\lambda^2 \lambda^{k-2}}{k \times (k-1) \times (k-2)!} \\
&= e^{-\lambda} \lambda^2 \sum_{k=2}^{+\infty} \frac{\lambda^{k-2}}{(k-2)!}
\end{aligned}$$

A nouveau petit changement d'indice : $i = k - 2$. Quand $k = 2$, $i = 0$. Quand $k = +\infty$, $i = +\infty$. Il ne reste plus qu'à remplacer dans l'expression.

$$\begin{aligned}
\mathbb{E}\left[X(X-1)\right] &= e^{-\lambda}\lambda^2 \sum_{k=2}^{+\infty} \frac{\lambda^{k-2}}{(k-2)!} \\
&= e^{-\lambda}\lambda^2 \sum_{i=0}^{+\infty} \frac{\lambda^{i}}{i!} \\
&= e^{-\lambda}\lambda^2 e^{\lambda} \\
&= \lambda^2
\end{aligned}$$

Reste à calculer la variance :

$$\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}\left[X(X-1)\right] + \mathbb{E}(X) - \mathbb{E}^2(X) \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda
\end{aligned}$$

Espérance de la loi uniforme continue

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

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

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \forall x > 0, \lambda > 0 \\ 0 & \text{sinon} \end{cases}$$

$$\mathbb{E}(X) = \int_{x \in \mathcal{X}} x \times f(x)\, dx$$
$$= \int_{0}^{+\infty} x \times \lambda e^{-\lambda x}\, dx$$

Pas de primitive évidente.

On va avoir recours à une intégration par parties.

$$
\begin{aligned}
& (u(x)\ v(x))' = u'(x) \times v(x) + u(x) \times v'(x) \\
\Leftrightarrow & \int_a^b (u(x)\ v(x))'\ dx = \int_a^b u'(x) \times v(x)\ dx + \int_a^b u(x) \times v'(x)\ dx \\
\Leftrightarrow & [u(x) \times v(x)]_a^b = \int_a^b u'(x) \times v(x)\ dx + \int_a^b u(x) \times v'(x)\ dx \\
\Leftrightarrow & \int_a^b u'(x) \times v(x)\ dx = [u(x) \times v(x)]_a^b - \int_a^b u(x) \times v'(x)\ dx
\end{aligned}
$$

On décompose la fonction de gauche qui n'a pas de primitive évidente en

- une fonction de primitive évidente : $u'(x)$;
- et telle que $u(x) \times v'(x)$ ait une primitive évidente.

Reste à définir qui est $u'(x)$ et qui est $v(x)$...

$$\int_a^b u'(x) \times v(x)\ dx = [u(x) \times v(x)]_a^b - \int_a^b u(x) \times v'(x)\ dx$$

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

Posons : $u'(x) = \lambda e^{-\lambda x}$ qui a une primitive évidente et donc $v(x) = x$.

On en déduit : $u(x) = -e^{-\lambda x}$ et $v'(x) = 1$.

$$\begin{aligned}
\mathbb{E}(X) &= \left[-xe^{-\lambda x}\right]_0^{+\infty} - \int_0^{+\infty} - e^{-\lambda x}\ dx \\
&= \lim_{x \to +\infty} \left(-xe^{-\lambda x}\right) - \left(-0 \times e^{-\lambda \times 0}\right) + \int_0^{+\infty} e^{-\lambda x}\ dx \\
&= \left[-\frac{1}{\lambda} e^{-\lambda x}\right]_0^{+\infty} \\
&= \frac{1}{\lambda} \left[-e^{-\lambda x}\right]_0^{+\infty} \\
&= \frac{1}{\lambda} \left[\lim_{x \to +\infty} \left(-e^{-\lambda x}\right) - \left(-e^{-\lambda \times 0}\right)\right] = \frac{1}{\lambda}
\end{aligned}$$

Variance de la loi uniforme continue

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

$$\begin{aligned}
\mathbb{E}(X^2) &= \int_{x \in \mathcal{X}} x^2 \times f(x)\; dx \\
&= \int_a^b x^2 \times \frac{1}{b-a}\; dx \\
&= \frac{1}{b-a} \int_a^b x^2\; dx \\
&= \frac{1}{b-a} \left[\frac{1}{3} \times x^3\right]_a^b \\
&= \frac{1}{3(b-a)} \left[b^3 - a^3\right] \\
&= \frac{(a^2 + ab + b^2)(b-a)}{3(b-a)} \\
&= \frac{a^2 + ab + b^2}{3}
\end{aligned}$$

On en déduit la variance :

$$\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}^2(X) \\
&= \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 \\
&= \frac{4\left(a^2 + ab + b^2\right) - 3\left(a+b\right)^2}{12} \\
&= \frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2}{12} \\
&= \frac{a^2 - 2ab + b^2}{12} \\
&= \frac{(b-a)^2}{12}
\end{aligned}$$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \forall x > 0, \lambda > 0 \\ 0 & \text{sinon} \end{cases}$$

$$\begin{aligned} \mathbb{E}(X^2) &= \int_{x \in \mathcal{X}} x^2 \times f(x)\, dx \\ &= \int_0^{+\infty} x^2 \times \lambda e^{-\lambda x}\, dx \end{aligned}$$

Pas de primitive évidente.

On va avoir recours à une intégration par parties.

$$\int_a^b u'(x) \times v(x)\ dx = [u(x) \times v(x)]_a^b - \int_a^b u(x) \times v'(x)\ dx$$

$$\mathbb{E}(X^2) = \int_0^{+\infty} x^2 \times \lambda e^{-\lambda x}\ dx$$

Posons : $u'(x) = \lambda e^{-\lambda x}$ qui a une primitive évidente et donc $v(x) = x^2$.

On en déduit : $u(x) = -e^{-\lambda x}$ et $v'(x) = 2x$.

$$\begin{aligned}
\mathbb{E}(X^2) &= \left[-x^2 e^{-\lambda x}\right]_0^{+\infty} - \int_0^{+\infty} -2xe^{-\lambda x}\ dx \\
&= \lim_{x \to +\infty} \left(-x^2 e^{-\lambda x}\right) - \left(-0^2 \times e^{-\lambda \times 0}\right) + \int_0^{+\infty} 2xe^{-\lambda x}\ dx \\
&= \int_0^{+\infty} 2xe^{-\lambda x}\ dx
\end{aligned}$$

pas de primitive connue. Donc soit on refait une intégration par parties, soit on remarque la ressemblance avec l'expression de l'espérance...

$$
\begin{aligned}
\mathbb{E}(X^2) &= \int_0^{+\infty} 2xe^{-\lambda x}\,dx \\
&= \frac{\lambda}{\lambda}\int_0^{+\infty} 2xe^{-\lambda x}\,dx \\
&= \frac{2}{\lambda}\int_0^{+\infty} x\lambda e^{-\lambda x}\,dx \\
&= \frac{2}{\lambda}\mathbb{E}(X) \\
&= \frac{2}{\lambda}\times\frac{1}{\lambda} \\
&= \frac{2}{\lambda^2}
\end{aligned}
$$

On en déduit la variance :

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}^2(X) \\ &= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 \\ &= \frac{1}{\lambda^2}\end{aligned}$$

VIII- Exercices de récapitulation

Pour déterminer k, supposons que f soit une densité, c'est-à-dire $\int_{-\infty}^{+\infty} f(x)dx = 1$. Manipulons le membre de droite :

$$\begin{aligned}
\int_{-\infty}^{+\infty} f(x)dx &= \int_{\theta}^{+\infty} kx^{-\alpha}dx \\
&= k\int_{\theta}^{+\infty} x^{-\alpha}dx \\
&= k\left[\frac{1}{-\alpha+1}x^{-\alpha+1}\right]_{\theta}^{+\infty} \\
&= \frac{k}{1-\alpha}\left[\frac{1}{x^{\alpha-1}}\right]_{\theta}^{+\infty} \\
&= \frac{k}{1-\alpha}\left[0-\frac{1}{\theta^{\alpha-1}}\right] \quad \text{puisque } \alpha-1>0 \\
&= \frac{k}{(\alpha-1)\theta^{\alpha-1}}
\end{aligned}$$

En égalisant à 1, on trouve donc :

$$\frac{k}{(\alpha-1)\theta^{\alpha-1}} = 1 \Leftrightarrow k = (\alpha-1)\theta^{\alpha-1} > 0, \text{ ce qui est cohérent.}$$

$$
\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{+\infty} x f(x)\,dx \\
&= \int_{\theta}^{+\infty} x(\alpha-1)\theta^{\alpha-1}x^{-\alpha}\,dx \\
&= (\alpha-1)\theta^{\alpha-1}\int_{\theta}^{+\infty} x^{-\alpha+1}\,dx \\
&= (\alpha-1)\theta^{\alpha-1}\left[\frac{1}{-\alpha+2}x^{-\alpha+2}\right]_{\theta}^{+\infty} \\
&= \frac{(\alpha-1)\theta^{\alpha-1}}{2-\alpha}\left[\frac{1}{x^{\alpha-2}}\right]_{\theta}^{+\infty} \\
&= \frac{(\alpha-1)\theta^{\alpha-1}}{2-\alpha}\left[0-\frac{1}{\theta^{\alpha-2}}\right] \qquad \text{puisque } \alpha-2>0 \\
&= \frac{(\alpha-1)\theta^{\alpha-1}}{(\alpha-2)\theta^{\alpha-2}} \\
&= \frac{\alpha-1}{\alpha-2}\theta^{\alpha-1-(\alpha-2)} \\
&= \frac{\alpha-1}{\alpha-2}\theta
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x)\,dx \\
&= \int_{\theta}^{+\infty} x^2(\alpha-1)\theta^{\alpha-1}x^{-\alpha}\,dx \\
&= (\alpha-1)\theta^{\alpha-1}\int_{\theta}^{+\infty} x^{-\alpha+2}\,dx \\
&= (\alpha-1)\theta^{\alpha-1}\left[\frac{1}{-\alpha+3}x^{-\alpha+3}\right]_{\theta}^{+\infty} \\
&= \frac{(\alpha-1)\theta^{\alpha-1}}{3-\alpha}\left[\frac{1}{x^{\alpha-3}}\right]_{\theta}^{+\infty} \\
&= \frac{(\alpha-1)\theta^{\alpha-1}}{3-\alpha}\left[0-\frac{1}{\theta^{\alpha-3}}\right] \qquad \text{puisque } \alpha-3>0 \\
&= \frac{(\alpha-1)\theta^{\alpha-1}}{(\alpha-3)\theta^{\alpha-3}} \\
&= \frac{\alpha-1}{\alpha-3}\theta^2
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}^2(X) \\
&= \frac{\alpha-1}{\alpha-3}\theta^2 - \left(\frac{\alpha-1}{\alpha-2}\right)^2\theta^2 \\
&= \left(\frac{1}{\alpha-3} - \frac{\alpha-1}{(\alpha-2)^2}\right)(\alpha-1)\theta^2 \\
&= \left(\frac{(\alpha-2)^2 - (\alpha-1)(\alpha-3)}{(\alpha-3)(\alpha-2)^2}\right)(\alpha-1)\theta^2 \\
&= \left(\frac{\alpha^2 - 4\alpha + 4 - (\alpha^2 - 4\alpha + 3)}{(\alpha-3)(\alpha-2)^2}\right)(\alpha-1)\theta^2 \\
&= \frac{\alpha-1}{(\alpha-3)(\alpha-2)^2}\theta^2
\end{aligned}
$$

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(u)du \\
&= \int_{\theta}^{x} (\alpha-1)\theta^{\alpha-1} u^{-\alpha} du \qquad \text{si } x \geqslant \theta \\
&= (\alpha-1)\theta^{\alpha-1} \int_{\theta}^{x} u^{-\alpha} du \\
&= (\alpha-1)\theta^{\alpha-1} \left[\frac{1}{-\alpha+1} u^{-\alpha+1} \right]_{\theta}^{x} \\
&= \frac{(\alpha-1)\theta^{\alpha-1}}{1-\alpha} \left[\frac{1}{u^{\alpha-1}} \right]_{\theta}^{x} \\
&= \frac{(\alpha-1)\theta^{\alpha-1}}{1-\alpha} \left[\frac{1}{x^{\alpha-1}} - \frac{1}{\theta^{\alpha-1}} \right] \\
&= \theta^{\alpha-1} \left[\frac{1}{\theta^{\alpha-1}} - \frac{1}{x^{\alpha-1}} \right] \\
&= 1 - \frac{\theta^{\alpha-1}}{x^{\alpha-1}}
\end{aligned}
$$

$$F(x) = \begin{cases} 0 & \text{si } x < \theta \\ 1 - \frac{\theta^{\alpha-1}}{x^{\alpha-1}} & \text{si x} \geqslant \theta \end{cases}$$

Exercice récapitulatif 2

$$f(x) = \begin{cases} 1+x & \forall x \in [-1,0] \\ 1-x & \forall x \in [0,1] \\ 0 & \text{sinon} \end{cases} \quad \text{est-elle une densité de probabilités ?}$$

1. La fonction doit être positive ou nulle :

$$\begin{cases} \text{si } x \in [-1,0], & f(x) = 1+x \in [0,1] \\ \text{si } x \in [0,1], & f(x) = 1-x \in [0,1] \\ \text{partout ailleurs}, & f(x) = 0 \end{cases}$$

2. La fonction doit être continue sauf éventuellement en un nombre fini de points : les cas à examiner ici sont les bornes des intervalles, c-à-d -1, 0 et $+1$:

$$\begin{cases} \text{si } x = -1, & f(-1^-) = 0 \text{ et } f(-1^+) = 1 + (-1) = 0 \text{ donc continue en } -1. \\ \text{si } x = 0, & f(0^-) = 1+0 = 1 \text{ et } f(0^+) = 1-0 = 1 \text{ donc continue en } 0. \\ \text{si } x = +1, & f(1^-) = 1-1 = 0 \text{ et } f(1^+) = 0 \text{ donc continue en } +1. \end{cases}$$

3. Il faut que $\int_{-\infty}^{+\infty} f(x)dx = 1$.

$$
\begin{aligned}
\int_{-\infty}^{+\infty} f(x)dx &= \int_{-\infty}^{-1} f(x)dx + \int_{-1}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{+\infty} f(x)dx \\
&= 0 + \int_{-1}^{0} (1+x)dx + \int_{0}^{1} (1-x)dx + 0 \\
&= \left[x + \frac{1}{2}x^2\right]_{-1}^{0} + \left[x - \frac{1}{2}x^2\right]_{0}^{1} \\
&= 0 - \left(-1 + \frac{1}{2}(-1)^2\right) + \left(1 - \frac{1}{2}(1)^2\right) - 0 \\
&= 1 - \frac{1}{2} + 1 - \frac{1}{2} = 1
\end{aligned}
$$

$$F(\boxed{x}) = \int_{-\infty}^{\boxed{x}} f(\boxed{t})d\boxed{t}$$

$$\text{si } x \in [-1, 0], \quad F(x) = \int_{-\infty}^{x} f(t)dt$$

$$= \int_{-1}^{x} f(t)dt$$

$$= \int_{-1}^{x} (1+t)dt$$

$$= \left[t + \frac{1}{2}t^2\right]_{-1}^{x}$$

$$= \left(x + \frac{1}{2}x^2\right) - \left((-1) + \frac{1}{2}(-1)^2\right)$$

$$= x + \frac{1}{2}x^2 + 1 - \frac{1}{2}$$

$$= x + \frac{1}{2}x^2 + \frac{1}{2}$$

$$
\begin{aligned}
\text{si } x \in [0,1], \quad F(x) &= \int_{-\infty}^{x} f(t)dt \\
&= \int_{-1}^{0} f(t)dt + \int_{0}^{x} f(t)dt \\
&= \int_{-1}^{0} (1+t)dt + \int_{0}^{x} (1-t)dt \\
&= \left[t + \frac{1}{2}t^2\right]_{-1}^{0} + \left[t - \frac{1}{2}t^2\right]_{0}^{x} \\
&= 0 - \left((-1) + \frac{1}{2}(-1)^2\right) + \left(x - \frac{1}{2}x^2\right) - 0 \\
&= +1 - \frac{1}{2} + x - \frac{1}{2}x^2 \\
&= x - \frac{1}{2}x^2 + \frac{1}{2}
\end{aligned}
$$

On a donc :

$$F(x) = \begin{cases} 0 & \text{si } x < -1 \\ x + \frac{1}{2}x^2 + \frac{1}{2} & \text{si } x \in [-1, 0] \\ x - \frac{1}{2}x^2 + \frac{1}{2} & \text{si } x \in [0, 1] \\ 1 & \text{si } x > 1 \end{cases}$$

$$
\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{+\infty} x f(x)\,dx \\
&= \int_{-1}^{0} x f(x)\,dx + \int_{0}^{1} x f(x)\,dx \\
&= \int_{-1}^{0} x(1+x)\,dx + \int_{0}^{1} x(1-x)\,dx \\
&= \left[\frac{1}{2}x^2 + \frac{1}{3}x^3\right]_{-1}^{0} + \left[\frac{1}{2}x^2 - \frac{1}{3}x^3\right]_{0}^{1} \\
&= 0 - \left(\frac{1}{2}(-1)^2 + \frac{1}{3}(-1)^3\right) + \left(\frac{1}{2}(1)^2 - \frac{1}{3}(1)^3\right) - 0 \\
&= -\frac{1}{2} + \frac{1}{3} + \frac{1}{2} - \frac{1}{3} = 0
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x)dx \\
&= \int_{-1}^{0} x^2 f(x)dx + \int_{0}^{1} x^2 f(x)dx \\
&= \int_{-1}^{0} x^2(1+x)dx + \int_{0}^{1} x^2(1-x)dx \\
&= \left[\frac{1}{3}x^3 + \frac{1}{4}x^4\right]_{-1}^{0} + \left[\frac{1}{3}x^3 - \frac{1}{4}x^4\right]_{0}^{1} \\
&= 0 - \left(\frac{1}{3}(-1)^3 + \frac{1}{4}(-1)^4\right) + \left(\frac{1}{3}(1)^3 - \frac{1}{4}(1)^4\right) - 0 \\
&= +\frac{1}{3} - \frac{1}{4} + \frac{1}{3} - \frac{1}{4} = \frac{1}{6}
\end{aligned}
$$

Donc comme $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$, $\mathbb{V}(X) = \dfrac{1}{6}$

Partie 1 : les couples de VA

I- Couples de VAD

Soit (X, Y) un couple de variables aléatoires discrètes de loi de probabilités jointe $\mathbb{P}(X = x, Y = y)$ définie pour un ensemble de définition $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

Pour une variable, on parlait d'une loi de probabilités jointe :

$$\forall x \in \mathcal{X}, \;\; \mathbb{P}(X = x) = \mathbb{P}(X = x) \geqslant 0$$
$$\sum_{x \in \mathcal{X}} \mathbb{P}(X = x) = 1$$

Pour un couple, on parle de loi de probabilités jointe. Comme précédemment, les probabilités sont positives et somment à 1 sur l'ensemble des possibles.

$$\forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \;\; \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x \cap Y = y) \geqslant 0$$
$$\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \mathbb{P}(X = x, Y = y) = 1$$

L'espérance :

$$\mathbb{E}(X) = \sum_{x \in \mathcal{X}} x \, \mathbb{P}(X = x)$$

$$\mathbb{E}(XY) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} xy \mathbb{P}(X = x, Y = y)$$

La variance :

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

$$\mathbb{V}(XY) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} [xy - \mathbb{E}(XY)]^2 \, \mathbb{P}(X = x, Y = y)$$

Définir les lois marginales de X et Y, leur espérance et leur variance.

Les lois marginales de X et Y :

$$\mathbb{P}(X = x) = \sum_{y \in \mathcal{Y}} \mathbb{P}(X = x, Y = y)$$
$$\mathbb{P}(Y = y) = \sum_{x \in \mathcal{X}} \mathbb{P}(X = x, Y = y)$$

Leurs espérances en repartant des formules déjà rencontrées :

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x \in \mathcal{X}} x \, \mathbb{P}(X = x) \\
&= \sum_{x \in \mathcal{X}} x \sum_{y \in \mathcal{Y}} \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} x \mathbb{P}(X = x, Y = y)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{y \in \mathcal{Y}} y \, \mathbb{P}(Y = y) \\
&= \sum_{y \in \mathcal{Y}} y \sum_{x \in \mathcal{X}} \mathbb{P}(X = x, Y = y) \\
&= \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} y \, \mathbb{P}(X = x, Y = y)
\end{aligned}$$

Leurs variances :

$$\begin{aligned}
\mathbb{V}(X) &= \sum_{x\in\mathcal{X}} [x - E(X)]^2 \, \mathbb{P}(X = x) \\
&= \sum_{x\in\mathcal{X}} [x - E(X)]^2 \, \sum_{y\in\mathcal{Y}} \mathbb{P}(X = x, Y = y) \\
&= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} [x - E(X)]^2 \, \mathbb{P}(X = x, Y = y)
\end{aligned}$$

$$\begin{aligned}
\mathbb{V}(Y) &= \sum_{y\in\mathcal{Y}} [y - E(Y)]^2 \, \mathbb{P}(Y = y) \\
&= \sum_{y\in\mathcal{Y}} [y - E(Y)]^2 \, \sum_{x\in\mathcal{X}} \mathbb{P}(X = x, Y = y) \\
&= \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} [y - E(Y)]^2 \, \mathbb{P}(X = x, Y = y)
\end{aligned}$$

		y			
		1	2	3	4
	-2	0.1	0.3	0.1	0
x	0	0.1	0	0.1	0.1
	2	0	0.1	c	0.1

Si c'est une loi de probabilités jointe, on va déterminer la valeur de c en utilisant la sommation à 1 des probabilités sur l'ensemble des possibles.

$$\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \mathbb{P}(X = x, Y = y) = 1$$

$$\Leftrightarrow 0.1 + 0.3 + 0.1 + 0 + 0.1 + 0 + 0.1 + 0.1 + 0 + 0.1 + c + 0.1 = 1$$

$$\Leftrightarrow c = 1 - 1 = 0$$

Caractérisez la loi de probabilités de $X + Y$.

Appelons Z la nouvelle VAD telle que $Z = X + Y$.
Construisons l'ensemble des possibles de Z. Pour cela, on calcule

$$\forall x \in \mathcal{X}, \forall y \in \mathcal{Y}, z = x + y$$

$+$		y			
		1	2	3	4
	-2	-1	0	1	2
x	0	1	2	3	4
	2	3	4	5	6

Les valeurs possibles de z (on ne répète pas celles qui apparaissent plusieurs fois) sont donc $\{-1, 0, 1, 2, 3, 4, 5, 6\}$.

Pour finir de construire la loi de Z, il faut maintenant calculer la loi de chaque occurrence.

$$\begin{aligned}
\mathbb{P}(Z=-1) &= \mathbb{P}(X=-2\cap Y=1)=0.1\\
\mathbb{P}(Z=0) &= \mathbb{P}(X=-2\cap Y=2)=0.3\\
\mathbb{P}(Z=1) &= \mathbb{P}\left[(X=-2\cap Y=3)\cup(X=0\cap Y=1)\right]\\
&= \mathbb{P}(X=-2\cap Y=3)+\mathbb{P}(X=0\cap Y=1)\\
&= 0.1+0.1=0.2\\
\mathbb{P}(Z=2) &= \mathbb{P}\left[(X=-2\cap Y=4)\cup(X=0\cap Y=2)\right]\\
&= \mathbb{P}(X=-2\cap Y=4)+\mathbb{P}(X=0\cap Y=2)\\
&= 0+0=0
\end{aligned}$$

$$\dots$$

On en déduit la loi de probabilités de Z :

z	-1	0	1	2	3	4	5	6
$\mathbb{P}(Z=z)$	0.1	0.3	0.2	0	0.1	0.2	0	0.1

Caractérisez la loi de probabilités conditionnelle $Y|X = 0$.

Repartons du tableau des probabilités jointes et faisons apparaître les probabilités marginales $\mathbb{P}(X = x) = \sum_{y \in \mathcal{Y}} \mathbb{P}(X = x, Y = y)$ et $\mathbb{P}(Y = y) = \sum_{x \in \mathcal{X}} \mathbb{P}(X = x, Y = y)$ en sommant par ligne et par colonne :

		y				
		1	2	3	4	$\mathbb{P}(X = x)$
	-2	0.1	0.3	0.1	0	0.5
x	0	0.1	0	0.1	0.1	0.3
	2	0	0.1	0	0.1	0.2
	$\mathbb{P}(Y = y)$	0.2	0.4	0.2	0.2	1

$\mathbb{P}(Y|X = 0)$ s'obtient par la formule de Bayes :

$$\mathbb{P}(Y = y|X = 0) = \frac{\mathbb{P}(Y = y \cap X = 0)}{\mathbb{P}(X = 0)}$$

Elle veut dire que maintenant, on ne s'intéresse plus à la distribution de Y que pour $X = 0$. Il vient :

$y\|X = 0$	1	2	3	4
$\mathbb{P}(Y = y\|X = 0)$	$\frac{0.1}{0.3}$	$\frac{0}{0.3}$	$\frac{0.1}{0.3}$	$\frac{0.1}{0.3}$

Caractérisez la loi de probabilités conditionnelle $X|Y = 1$.

		y				
		1	2	3	4	$\mathbb{P}(X = x)$
	-2	0.1	0.3	0.1	0	0.5
x	0	0.1	0	0.1	0.1	0.3
	2	0	0.1	0	0.1	0.2
	$\mathbb{P}(Y = y)$	0.2	0.4	0.2	0.2	1

Même raisonnement qu'à la question précédente. La formule de Bayes nous dit :

$$\mathbb{P}(X = x|Y = 1) = \frac{\mathbb{P}(X = x \cap Y = 1)}{\mathbb{P}(Y = 1)}$$

soit la distribution de X quand $Y = 1$. Il vient :

$x|Y = 1$	-2	0	2
$\mathbb{P}(X = x|Y = 1)$	$\frac{0.1}{0.2}$	$\frac{0.1}{0.2}$	$\frac{0}{0.2}$

Soient X et Y deux variables aléatoires discrètes de loi de probabilités jointe $\mathbb{P}(X = x, Y = y)$ avec $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

On définit l'opérateur Covariance comme :

$$\begin{aligned}\mathbb{C}ov(X, Y) &= \mathbb{E}\big\{[X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]\big\} \\ &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} [x - \mathbb{E}(X)][y - \mathbb{E}(Y)]\mathbb{P}(X = x, Y = y)\end{aligned}$$

Montrez que $\mathbb{C}ov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y)$.

$$
\begin{aligned}
\mathbb{C}ov(X, Y) &= \mathbb{E}\{[X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]\} \\
&= \mathbb{E}\left[XY - Y\mathbb{E}(X) - X\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y)\right] \\
&= \mathbb{E}(XY) - \mathbb{E}\left[Y\mathbb{E}(X)\right] - \mathbb{E}\left[X\mathbb{E}(Y)\right] + \mathbb{E}\left[\mathbb{E}(X)\mathbb{E}(Y)\right] \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(Y)\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y) \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{C}ov(X,Y) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}[x-\mathbb{E}(X)][y-\mathbb{E}(Y)]\mathbb{P}(X=x,Y=y)\\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}[xy\mathbb{P}(X=x,Y=y)-y\mathbb{E}(X)\mathbb{P}(X=x,Y=y)\\
&\qquad -x\mathbb{E}(Y)\mathbb{P}(X=x,Y=y)+\mathbb{E}(X)\mathbb{E}(Y)\mathbb{P}(X=x,Y=y)]\\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}xy\mathbb{P}(X=x,Y=y)-\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}y\mathbb{E}(X)\mathbb{P}(X=x,Y=y)\\
&\qquad -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}x\mathbb{E}(Y)\mathbb{P}(X=x,Y=y)+\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\mathbb{E}(X)\mathbb{E}(Y)\mathbb{P}(X=x,Y=y)\\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}xy\mathbb{P}(X=x,Y=y)-\mathbb{E}(X)\sum_{y\in\mathcal{Y}}y\underbrace{\sum_{x\in\mathcal{X}}\mathbb{P}(X=x,Y=y)}_{=\mathbb{P}(Y=y)}\\
&\qquad -\mathbb{E}(Y)\sum_{x\in\mathcal{X}}x\underbrace{\sum_{y\in\mathcal{Y}}\mathbb{P}(X=x,Y=y)}_{=\mathbb{P}(X=x)}+\mathbb{E}(X)\mathbb{E}(Y)\underbrace{\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\mathbb{P}(X=x,Y=y)}_{=1}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{C}ov(X, Y) &= \underbrace{\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} xy\mathbb{P}(X = x, Y = y)}_{=\mathbb{E}(XY)} - \mathbb{E}(X)\underbrace{\sum_{y\in\mathcal{Y}} y\mathbb{P}(Y = y)}_{=\mathbb{E}(Y)} \\
&\quad - \mathbb{E}(Y)\underbrace{\sum_{x\in\mathcal{X}} x\mathbb{P}(X = x)}_{=\mathbb{E}(X)} + \mathbb{E}(X)\mathbb{E}(Y) \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \underbrace{- \mathbb{E}(Y)\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y)}_{=0} \\
&= \mathbb{E}(XY) - \mathbb{E}(Y)\mathbb{E}(X)
\end{aligned}
$$

Que vaut $\mathbb{C}ov(X, X)$?

En repartant de $\mathbb{C}ov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(Y)\mathbb{E}(X)$ et en posant $X = Y$:

$$\begin{aligned}
\mathbb{C}ov(X, X) &= \mathbb{E}(XX) - \mathbb{E}(X)\mathbb{E}(X) \\
&= \mathbb{E}(X^2) - \mathbb{E}^2(X) \\
&\equiv \mathbb{V}(X)
\end{aligned}$$

Que vaut $\mathbb{C}ov(aX, bY)$?

$$\begin{aligned}
\mathbb{C}ov(aX, bY) &= \mathbb{E}\{[aX - \mathbb{E}(aX)][bY - \mathbb{E}(bY)]\} \\
&= \mathbb{E}\{[aX - a\mathbb{E}(X)][bY - b\mathbb{E}(Y)]\} \\
&= \mathbb{E}\{a[X - \mathbb{E}(X)]b[Y - \mathbb{E}(Y)]\} \\
&= ab\ \underbrace{\mathbb{E}\{[X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]\}} \\
&= ab\,\mathbb{C}ov(X, Y)
\end{aligned}$$

Ou en replongeant dans les sommes :

$$\begin{aligned}
\mathbb{C}ov(aX, bY) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}[ax-\mathbb{E}(aX)][by-\mathbb{E}(bY)]\ \mathbb{P}(X=x, Y=y) \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}[ax-a\mathbb{E}(X)][by-b\mathbb{E}(Y)]\ \mathbb{P}(X=x, Y=y) \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} ab[x-\mathbb{E}(X)][y-\mathbb{E}(Y)]\ \mathbb{P}(X=x, Y=y) \\
&= ab\ \underbrace{\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}[x-\mathbb{E}(X)][y-\mathbb{E}(Y)]\ \mathbb{P}(X=x, Y=y)} \\
&= ab\, \mathbb{C}ov(X, Y)
\end{aligned}$$

Que vaut $\mathbb{C}ov(X, Y)$ si les deux variables aléatoires X et Y sont indépendantes en probabilités ?

L'indépendance en probabilités implique :

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

Donc :

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} xy \ \mathbb{P}(X = x, Y = y) \\
&= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} xy \ \mathbb{P}(X = x) \ \mathbb{P}(Y = y) \\
&= \sum_{x\in\mathcal{X}} x\mathbb{P}(X = x) \sum_{y\in\mathcal{Y}} y\mathbb{P}(Y = y) \\
&= \mathbb{E}(X) \ \mathbb{E}(Y)
\end{aligned}$$

Comme :

$$\begin{aligned}
\mathbb{C}ov(X, Y) &= \mathbb{E}(XY) - \mathbb{E}(Y)\mathbb{E}(X) \\
&= 0 \ \text{ si indépendance}
\end{aligned}$$

Autre façon avec les sommes :

$$
\begin{aligned}
\mathbb{C}ov(X, Y) &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} [x - \mathbb{E}(X)][y - \mathbb{E}(Y)] \ \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} [x - \mathbb{E}(X)][y - \mathbb{E}(Y)] \ \mathbb{P}(X = x) \ \mathbb{P}(Y = y) \\
&= \sum_{x \in \mathcal{X}} [x - \mathbb{E}(X)] \mathbb{P}(X = x) \ \sum_{y \in \mathcal{Y}} [y - \mathbb{E}(Y)] \mathbb{P}(Y = y) \\
&= \mathbb{E}[X - \mathbb{E}(X)] \ \mathbb{E}[Y - \mathbb{E}(Y)] \\
&= [\mathbb{E}(X) - \mathbb{E}(X)] \ [\mathbb{E}(Y) - \mathbb{E}(Y)] \\
&= 0
\end{aligned}
$$

Toujours dans ce cas, calculez $\mathbb{V}(X+Y)$. Comme

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

il vient :

$$\begin{aligned}
\mathbb{V}(X+Y) &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} [x + y - \mathbb{E}(X+Y)]^2 \, \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \{[x + y - \mathbb{E}(X) - \mathbb{E}(Y)]\}^2 \, \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \{[x - \mathbb{E}(X)] + [y - \mathbb{E}(Y)]\}^2 \, \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \{[x - \mathbb{E}(X)]^2 + [y - \mathbb{E}(Y)]^2 + 2[x - \mathbb{E}(X)][y - \mathbb{E}(Y)]\} \, \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} [x - \mathbb{E}(X)]^2 \mathbb{P}(X = x, Y = y) + \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} [y - \mathbb{E}(Y)]^2 \mathbb{P}(X = x, Y = y) \\
&\quad + \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} 2[x - \mathbb{E}(X)][y - \mathbb{E}(Y)] \mathbb{P}(X = x, Y = y)
\end{aligned}$$

$$
\begin{aligned}
\mathbb{V}(X+Y) &= \sum_{x\in\mathcal{X}} [x-\mathbb{E}(X)]^2 \overbrace{\sum_{y\in\mathcal{Y}} \mathbb{P}(X=x,Y=y)}^{=\mathbb{P}(X=x)} + \sum_{y\in\mathcal{Y}} [y-\mathbb{E}(Y)]^2 \overbrace{\sum_{x\in\mathcal{X}} \mathbb{P}(X=x,Y=y)}^{=\mathbb{P}(Y=y)} \\
&\qquad + 2\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} [x-\mathbb{E}(X)][y-\mathbb{E}(Y)]\mathbb{P}(X=x,Y=y) \\
&= \overbrace{\sum_{x\in\mathcal{X}} [x-\mathbb{E}(X)]^2 \mathbb{P}(X=x)}^{=\mathbb{V}(X)} + \overbrace{\sum_{y\in\mathcal{Y}} [y-\mathbb{E}(Y)]^2 \mathbb{P}(Y=y)}^{\mathbb{V}(Y)} \\
&\qquad \underbrace{+ 2\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} [x-\mathbb{E}(X)][y-\mathbb{E}(Y)]\mathbb{P}(X=x,Y=y)}_{=\mathbb{C}ov(X,Y)} \\
&= \mathbb{V}(X)+\mathbb{V}(Y)+2\,\mathbb{C}ov(X,Y)
\end{aligned}
$$

Si les deux VA sont indépendantes en probabilités : $\mathbb{C}ov(X, Y) = 0$ donc

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

II- Couples de VAC

Comment sait-on qu'on est en présence d'une densité de probabilités d'un couple de VAC ?

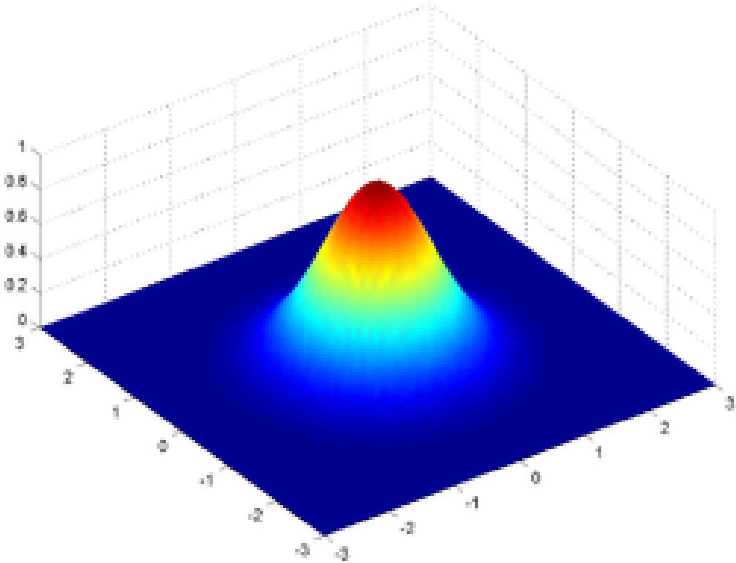
Loi de probabilités pour un couple de VAD :

- $\forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \;\; \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x \cap Y = y) \geqslant 0.$
- $\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \mathbb{P}(X = x, Y = y) = 1.$

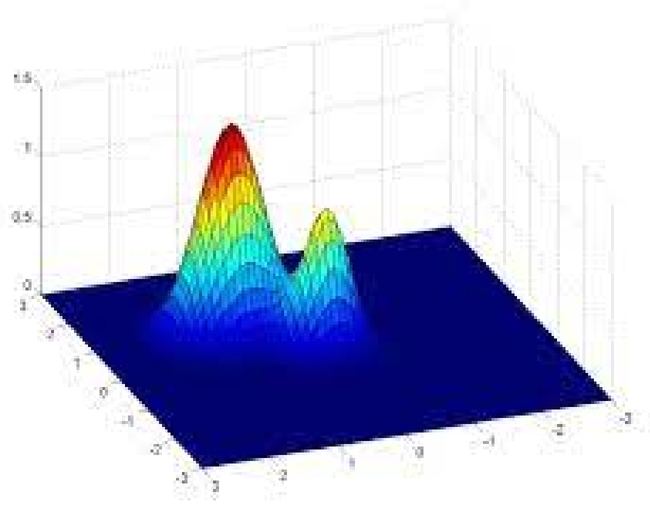
Une densité de probabilités de couple de VAC vérifie les conditions suivantes :

- $\forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \;\; f_{X,Y}(x, y) \geqslant 0.$
- C'est une fonction continue (sauf éventuellement en un nombre fini de points).
- $\int_{x \in \mathcal{X}} \int_{y \in \mathcal{Y}} f_{X,Y}(x, y) \; dxdy = 1.$

Exemple



Exemple



Soit (X, Y) un couple de variables aléatoires continues de densité de probabilités jointe $f_{X,Y}(x, y) = \begin{cases} cxy & \forall (x, y) \in [0, 4] \times [1, 5] \\ 0 & \text{sinon} \end{cases}$.

Pour déterminer c (une inconnue), il faut une équation :

$$
\begin{aligned}
\int_{x\in\mathcal{X}} \int_{y\in\mathcal{Y}} f_{X,Y}(x, y)\ dxdy = 1 &\Leftrightarrow \int_{x=0}^{4} \int_{y=1}^{5} cxy\ dxdy = 1 \\
&\Leftrightarrow c \int_{x=0}^{4} \int_{y=1}^{5} xy\ dxdy = 1 \\
&\Leftrightarrow c \int_{0}^{4} x\ dx \int_{1}^{5} y\ dy = 1 \\
&\Leftrightarrow c \left[\frac{1}{2}x^2\right]_0^4 \left[\frac{1}{2}y^2\right]_1^5 = 1 \\
&\Leftrightarrow c \times \frac{1}{2}(4^2 - 0^2) \times \frac{1}{2}(5^2 - 1^2) = 1 \\
&\Leftrightarrow c \times 8 \times 12 = 1 \\
&\Leftrightarrow c = \frac{1}{96}
\end{aligned}
$$

Définir l'espérance.

$$\text{VAD} : \mathbb{E}(XY) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} xy \mathbb{P}(X = x, Y = y)$$

$$\text{VAC} : \mathbb{E}(XY) = \int_{x \in \mathcal{X}} \int_{y \in \mathcal{Y}} xy f_{X,Y}(x, y) dxdy$$

Définir la variance.

$$\text{VAD} : \mathbb{V}(XY) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \left[xy - \mathbb{E}(XY)\right]^2 \mathbb{P}(X = x, Y = y)$$

$$\text{VAC} : \mathbb{V}(XY) = \int_{x \in \mathcal{X}} \int_{y \in \mathcal{Y}} \left[xy - \mathbb{E}(XY)\right]^2 f_{X,Y}(x, y)dxdy$$

Définir les lois marginales de X et Y, leur espérance et leur variance.

Les lois marginales de X et Y :

$$\text{VAD} : \mathbb{P}(X = x) = \sum_{y \in \mathcal{Y}} \mathbb{P}(X = x, Y = y)$$

$$\text{VAC} : f_X(x) = \int_{y \in \mathcal{Y}} f_{X,Y}(x, y) dy$$

$$\text{VAD} : \mathbb{P}(Y = y) = \sum_{x \in \mathcal{X}} \mathbb{P}(X = x, Y = y)$$

$$\text{VAC} : f_Y(y) = \int_{x \in \mathcal{X}} f_{X,Y}(x, y) dx$$

La fonction de répartition $F(.)$.

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)dudv$$

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f_{X,Y}(u,y)dudy$$

$$F_Y(y) = \int_{-\infty}^{y} f_Y(u)du = \int_{-\infty}^{+\infty} \int_{-\infty}^{y} f_{X,Y}(x,u)dxdu$$

$$\text{VAD} : \mathbb{C}ov(X, Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} [x - \mathbb{E}(X)][y - \mathbb{E}(Y)]\mathbb{P}(X = x, Y = y)$$

$$\text{VAC} : \mathbb{C}ov(X, Y) = \int_{x \in \mathcal{X}} \int_{y \in \mathcal{Y}} [x - \mathbb{E}(X)][y - \mathbb{E}(Y)] f_{X,Y}(x, y) dx dy$$

Montrez que $\mathbb{C}ov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

$$\begin{aligned}
\mathbb{C}ov(X,Y) &= \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} [x-\mathbb{E}(X)][y-\mathbb{E}(Y)] f_{X,Y}(x,y)dxdy \\
&= \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} [xyf_{X,Y}(x,y)dxdy - y\mathbb{E}(X)f_{X,Y}(x,y)dxdy \\
&\qquad - x\mathbb{E}(Y)f_{X,Y}(x,y)dxdy + \mathbb{E}(X)\mathbb{E}(Y)f_{X,Y}(x,y)dxdy] \\
&= \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} xyf_{X,Y}(x,y)dxdy - \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} y\mathbb{E}(X)f_{X,Y}(x,y)dxdy \\
&- \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} x\mathbb{E}(Y)f_{X,Y}(x,y)dxdy + \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} \mathbb{E}(X)\mathbb{E}(Y)f_{X,Y}(x,y)dxdy \\
&= \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} xyf_{X,Y}(x,y)dxdy - \mathbb{E}(X)\int_{y\in\mathcal{Y}} y\underbrace{\int_{x\in\mathcal{X}} f_{X,Y}(x,y)dx}_{=f_Y(y)}\,dy \\
&- \mathbb{E}(Y)\int_{x\in\mathcal{X}} x\underbrace{\int_{y\in\mathcal{Y}} f_{X,Y}(x,y)dy}_{=f_X(x)}\,dx + \mathbb{E}(X)\mathbb{E}(Y)\underbrace{\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} f_{X,Y}(x,y)dxdy}_{=1}
\end{aligned}$$

$$
\begin{aligned}
\mathbb{C}ov(X, Y) &= \underbrace{\int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} xyf_{X,Y}(x, y)dxdy}_{=\mathbb{E}(XY)} - \mathbb{E}(X)\underbrace{\int_{y\in\mathcal{Y}} yf_Y(y)dy}_{=\mathbb{E}(Y)} \\
&\quad - \mathbb{E}(Y)\underbrace{\int_{x\in\mathcal{X}} xf_X(x)dx}_{=\mathbb{E}(X)} + \mathbb{E}(X)\mathbb{E}(Y) \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \underbrace{-\mathbb{E}(Y)\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y)}_{=0} \\
&= \mathbb{E}(XY) - \mathbb{E}(Y)\mathbb{E}(X)
\end{aligned}
$$

Que vaut $\mathbb{C}ov(X, Y)$ si les deux variables aléatoires X et Y sont indépendantes en probabilités ?

L'indépendance en probabilités implique :

$$f_{X,Y}(x, y) = f_X(x)\ f_Y(y)$$

Donc :

$$\begin{aligned}
\mathbb{E}(XY) &= \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} xy\,f_{X,Y}(x, y)dxdy \\
&= \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} xy\,f_X(x)f_Y(y)dxdy \\
&= \int_{x\in\mathcal{X}} xf_X(x)dx \int_{y\in\mathcal{Y}} yf_Y(y)dy \\
&= \mathbb{E}(X)\mathbb{E}(Y)
\end{aligned}$$

Comme :

$$\begin{aligned}
\mathbb{C}ov(X, Y) &= \mathbb{E}(XY) - \mathbb{E}(Y)\mathbb{E}(X) \\
&= 0 \ \text{ si indépendance}
\end{aligned}$$

$$
\begin{aligned}
& f_{X,Y}(u,v) = f_X(u)\ f_Y(v) \\
\Leftrightarrow & \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)dudv = \int_{-\infty}^{x}\int_{-\infty}^{y} f_X(u)\ f_Y(v)dudv \\
\Leftrightarrow & \underbrace{\int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)dudv}_{F_{X,Y}(x,y)} = \underbrace{\int_{-\infty}^{x} f_X(u)du}_{F_X(x)} \underbrace{\int_{-\infty}^{y} f_Y(v)dv}_{F_Y(y)}
\end{aligned}
$$

Partie 2 : Inférence statistique

I- Construction d’un estimateur

Vous jouez à pile ou face et vous comptez le nombre de pile. Vous disposez d'un échantillon aléatoire de N tirages $(X_1, X_2, \ldots, X_N)$ pour lesquels vous observez la réalisation de pile (succès codé 1) $(x_1, x_2, \ldots, x_N)$.

- la loi de cette expérience est la loi de **Bernoulli** : donne la probabilité de succès ($p \in [0, 1]$) ou d'échec ($1 - p$) de l'expérience.
- On suppose que chaque tirage ou variable aléatoire de l'échantillon suit la même loi de Bernoulli de paramètre $p \in [0, 1]$ et quelles sont toutes indépendantes.
- Loi de probabilités de Bernoulli de paramètre $p \in [0, 1]$:

$$\mathbb{P}(X_i = k) = p^k (1 - p)^{1-k}$$

avec $k = 0, 1$ (resp. échec, succès).

Vous êtes assureur et vous êtes intéressés par **le nombre des sinistres** sur une certaine période. Vous disposez d'un échantillon aléatoire de N clients $(X_1, X_2, \ldots, X_N)$ pour lesquels vous observez le nombre de sinistres réalisés $(x_1, x_2, \ldots, x_N)$.

- On suppose que la loi sous-jacente est la loi de **Poisson** (loi des événements rares comme les accidents).
- On suppose que chaque variable aléatoire de l'échantillon suit la même loi de Poisson et qu'elles sont toutes indépendantes.
- On suppose donc ici que **tout le monde a la même probabilité de sinistres**...!??
- Sa loi de probabilités :

$$\mathbb{P}(X_i = x_i) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

avec $x_i \in \mathbb{N}$.

- **Echantillon** = 1 groupe d'observations tirées "au hasard" dans la population
- **iid** = **i**dentiquement et **i**ndépendamment **d**istribués
- une observation de l'échantillon x_i est une **réalisation** particulière d'une Variable Aléatoire (**VA**) X_i

i	1	2	...	N
X	x_1	x_2	...	x_N

- Une VA mesure **quantitativement** le résultat d'une expérience aléatoire grâce à sa **loi de probabilités** caractérisée par un ou plusieurs **paramètres**.
- Ces paramètres sont **inconnus**, on va donc chercher à les **estimer** avec des outils **fiables**.

- L’**estimateur** est un outil général issu d’un critère d’estimation
- sert à **”deviner”** la valeur du paramètre de la loi caractérisant une **population** grâce à l’**échantillon**.
- l’échantillon $(X_1, X_2, \ldots, X_N)$ = ensemble de VA supposées **iid**.
- le **maximum de vraisemblance** : ”quelle valeur des paramètres de la loi permet de maximiser le log de la probabilité de l’échantillon ?”

$$\max_{?} \ln \mathbb{P}\{(X_1 = x_1) \cap (X_2 = x_2) \cap \cdots \cap (X_N = x_N)\}$$

- Permet de déduire une formule générale pour calculer la valeur du ou des paramètres de la loi sous-jacente au phénomène étudié et suivie par les VA : **l’estimateur**
- Permet de calculer une valeur quand appliqué à l’échantillon : **l’estimation**

Soit une suite de n VAD iid $(X_1, X_2, \ldots, X_n)$.

On a vu que pour deux variables aléatoires X et Y indépendantes en probabilités, on a :

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

On généralise :

$$\begin{aligned}
\mathbb{P}\{(X_1 = x_1) \cap (X_2 = x_2) \cap \cdots \cap (X_n = x_n)\} \\
&= \mathbb{P}(X_1 = x_1) \times \mathbb{P}(X_2 = x_2) \times \cdots \times \mathbb{P}(X_n = x_n) \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) = \prod_{i=1}^{n} \mathbb{P}(X = x_i)
\end{aligned}$$

En appliquant le logarithme :

$$\ln\left\{\prod_{i=1}^{n} \mathbb{P}(X = x_i)\right\} = \sum_{i=1}^{n} \ln[\mathbb{P}(X = x_i)] \equiv \mathcal{L}(.)$$

Soit une suite de n VAC iid $(X_1, X_2, \ldots, X_n)$.

On a vu que pour deux VAC X et Y indépendantes en probabilités, on a :

$$f_{X,Y}(x, y) = f_X(x)\; f_Y(y)$$

Comme indépendantes et identiques,

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} f_X(x_i)$$

En appliquant le logarithme :

$$\ln\left\{\prod_{i=1}^{n} f_X(x_i)\right\} = \sum_{i=1}^{n} \ln[f_X(x_i)] \equiv \mathcal{L}(.)$$

On notera de façon générique dans ce qui suit $\mathcal{L}(\theta; x)$ la fonction à maximiser.

Elle dépend

- du ou des paramètres inconnus θ qui est soit un scalaire, soit un vecteur de k paramètres $\theta = (\theta_1, \theta_2, \ldots, \theta_k)'$;

- des observations de l'échantillon notée $x = (x_1, x_2, \ldots, x_n)'$.

Le programme s'écrit :

$$\widehat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta; x)$$

Trouver le maximum revient à calculer la dérivée de la fonction par rapport à l'inconnue. En la solution $\widehat{\theta}$, cette dérivée est égale à 0. Cela nous donne ici une équation à une inconnue ou un système de *k* équations à *k* inconnues $\widehat{\lambda}$.

Comme cette condition est valable pour un maximum ou un minimum, il faut s'assurer qu'on est bien à un maximum en vérifiant le signe de la dérivée seconde.

1. On calcule $\frac{\partial \mathcal{L}}{\partial \theta}(\theta; x)$, la dérivée à l'ordre 1, qui doit s'annuler pour un maximum :

$$\frac{\partial \mathcal{L}}{\partial \theta}(\widehat{\theta}; x) = 0$$

2. On calcule $\frac{\partial^2 \mathcal{L}}{\partial \theta^2}(\theta; x)$, la dérivée du second ordre, pour étudier son signe :

   si $\frac{\partial^2 \mathcal{L}}{\partial \theta^2}(\theta; x) < 0, \ \forall \theta$, on est à un maximum global ;

   si $\frac{\partial^2 \mathcal{L}}{\partial \theta^2}(\widehat{\theta}; x) < 0$ pour le ou les points candidats trouvés, on est à un maximum local.

1. On calcule le système des dérivées partielles premières (ou gradient) et on cherche les θ qui vérifient :

$$g(\widehat{\theta}; x) = 0 \Leftrightarrow \begin{cases} \dfrac{\partial \mathcal{L}}{\partial \theta_1}(\widehat{\theta}; x) = 0 \\ \vdots \\ \dfrac{\partial \mathcal{L}}{\partial \theta_k}(\widehat{\theta}; x) = 0 \end{cases}$$

2. On calcule les dérivées partielles secondes (ou matrice Hessienne) :

$$H(\theta; x) = \begin{pmatrix} \dfrac{\partial^2 \mathcal{L}}{\partial \theta_1^2}(\theta; x) & \dfrac{\partial^2 \mathcal{L}}{\partial \theta_1 \partial \theta_2}(\theta; x) & \ldots & \dfrac{\partial^2 \mathcal{L}}{\partial \theta_1 \partial \theta_k}(\theta; x) \\ \vdots & \vdots & \ldots & \vdots \\ \dfrac{\partial^2 \mathcal{L}}{\partial \theta_k \partial \theta_1}(\theta; x) & \dfrac{\partial^2 \mathcal{L}}{\partial \theta_k \partial \theta_2}(\theta; x) & \ldots & \dfrac{\partial^2 \mathcal{L}}{\partial \theta_k^2}(\theta; x) \end{pmatrix}$$

si les mineurs de $H(\theta; x)$ alternent de signes en commençant par le négatif, on est à un maximum global ;

si les mineurs de $H(\widehat{\theta}; x)$ alternent de signes en commençant par le négatif, on est à un maximum local ;

Pour k paramètres inconnus, il faut k dérivées partielles premières.

Il ne faut pas confondre $g(\theta; x)$, qu'on a nommé gradient et qui dépend des données, avec $g(\widehat{\theta}; X)$, qui dépend des VA et qu'on nommera score.

Le score est aléatoire comme fonction des VA. Une propriété importante est :

$$\mathbb{E}(g(\theta; X)) = 0$$

De même pour la Hessienne déterministe $H(\theta; x)$ et la Hessienne stochastique $H(\theta; X)$.

La matrice Hessienne est carrée, de format k et symétrique.

Application : à la loi de Poisson de paramètre $\lambda > 0$.

On sait que $\forall x_i \in [0, +\infty[,\ \mathbb{P}(X = x_i) = \dfrac{\lambda^{x_i} e^{-\lambda}}{x_i!}$. Prenons le $\ln(.)$:

$$
\begin{aligned}
\ln[\mathbb{P}(X = x_i)] &= \ln\left(\frac{\lambda^{x_i} e^{-\lambda}}{x_i!}\right) \\
&= \ln\left(\lambda^{x_i} e^{-\lambda}\right) - \ln\left(x_i!\right) \\
&= \ln\left(\lambda^{x_i}\right) + \ln\left(e^{-\lambda}\right) - \ln\left(x_i!\right) \\
&= \ln\left(\lambda^{x_i}\right) - \lambda \ln(e) - \ln\left(x_i!\right) \\
&= x_i \ln(\lambda) - \lambda - \ln\left(x_i!\right)
\end{aligned}
$$

On peut alors déduire la log-vraisemblance de l'échantillon :

$$
\begin{aligned}
\mathcal{L}(\lambda; x) &= \sum_{i=1}^{n} \ln[\mathbb{P}(X = x_i)] \\
&= \sum_{i=1}^{n} \left(x_i \ln(\lambda) - \lambda - \ln(x_i!)\right) \\
&= \sum_{i=1}^{n} x_i \ln(\lambda) - \sum_{i=1}^{n} \lambda - \sum_{i=1}^{n} \ln(x_i!) \\
&= \ln(\lambda) \sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} \ln(x_i!)
\end{aligned}
$$

Calculons la dérivée première en partant de l'expression trouvée à la question précédente :

$$
\begin{aligned}
\mathcal{L}(\lambda; x) &= \ln(\lambda) \sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} \ln(x_i!) \\
\frac{\partial \mathcal{L}}{\partial \lambda}(\lambda; x) &= \frac{\partial \left[\ln(\lambda) \sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} \ln(x_i!)\right]}{\partial \lambda} \\
&= \frac{\partial \left[\ln(\lambda) \sum_{i=1}^{n} x_i\right]}{\partial \lambda} - \frac{\partial (n\lambda)}{\partial \lambda} - \underbrace{\frac{\partial \left[\sum_{i=1}^{n} \ln(x_i!)\right]}{\partial \lambda}}_{=0} \\
&= \left(\sum_{i=1}^{n} x_i\right) \frac{\partial \left[\ln(\lambda)\right]}{\partial \lambda} - n \frac{\partial (\lambda)}{\partial \lambda} \\
&= \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n
\end{aligned}
$$

Calculons la dérivée première en la solution :

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \lambda}(\widehat{\lambda}; x) &= 0 \\ \Leftrightarrow \frac{1}{\widehat{\lambda}} \sum_{i=1}^{n} x_i - n = 0 \\ \Leftrightarrow n = \frac{1}{\widehat{\lambda}} \sum_{i=1}^{n} x_i \\ \Leftrightarrow \widehat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i \equiv \overline{x} \end{aligned}$$

Calculons la dérivée seconde comme la dérivée de la dérivée première :

$$
\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \lambda^2}(\lambda; x) &= \frac{\partial \left[\dfrac{\partial \mathcal{L}}{\partial \lambda}(\lambda; x) \right]}{\partial \lambda} \\
&= \frac{\partial \left(\dfrac{1}{\lambda} \sum_{i=1}^{n} x_i \right)}{\partial \lambda} \\
&= \left(\sum_{i=1}^{n} x_i \right) \frac{\partial \left(\frac{1}{\lambda} \right)}{\partial \lambda} \\
&= - \underbrace{\frac{1}{\lambda^2}}_{>0} \underbrace{\sum_{i=1}^{n} x_i}_{>0} < 0
\end{aligned}
$$

C'est donc un maximum global.

Remarque

On notera :

$$\widehat{\lambda} = \widehat{\lambda}(x) = \overline{x} \text{ : l'estimation = un nombre}$$

$$\widehat{\lambda} = \widehat{\lambda}(X) = \overline{X} \text{ : l'estimateur = une VA}$$

On sait que $x_i = 0, 1 \quad \mathbb{P}(X = x_i) = p^{x_i}\ (1-p)^{(1-x_i)}$. On calcule son $\ln(.)$:

$$\begin{aligned} \ln[\mathbb{P}(X = x_i)] &= \ln\left(p^{x_i}\ (1-p)^{(1-x_i)}\right) \\ &= \ln\left(p^{x_i}\right) + \ln\left((1-p)^{(1-x_i)}\right) \\ &= x_i \ln(p) + (1-x_i)\ln(1-p) \end{aligned}$$

On peut alors déduire :

$$\begin{aligned} \mathcal{L}(p; x) &= \sum_{i=1}^{n} \ln[\mathbb{P}(X = x_i)] \\ &= \sum_{i=1}^{n} \left(x_i \ln(p) + (1-x_i)\ln(1-p)\right) \\ &= \sum_{i=1}^{n} x_i \ln(p) + \sum_{i=1}^{n} (1-x_i)\ln(1-p) \\ &= \ln(p)\sum_{i=1}^{n} x_i + \ln(1-p)\sum_{i=1}^{n}(1-x_i) \end{aligned}$$

Calculons la dérivée première :

$$
\begin{aligned}
\mathcal{L}(p;x) &= \ln(p)\sum_{i=1}^{n} x_i + \ln(1-p)\sum_{i=1}^{n}(1-x_i) \\
\frac{\partial \mathcal{L}}{\partial p}(p;x) &= \frac{\partial\left[\ln(p)\sum_{i=1}^{n} x_i + \ln(1-p)\sum_{i=1}^{n}(1-x_i)\right]}{\partial p} \\
&= \frac{\partial\left[\ln(p)\sum_{i=1}^{n} x_i\right]}{\partial p} + \frac{\partial\left[\ln(1-p)\sum_{i=1}^{n}(1-x_i)\right]}{\partial p} \\
&= \left[\sum_{i=1}^{n} x_i\right] \times \frac{\partial\left[\ln(p)\right]}{\partial p} + \left[\sum_{i=1}^{n}(1-x_i)\right] \times \frac{\partial\left[\ln(1-p)\right]}{\partial p} \\
&= \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}\sum_{i=1}^{n}(1-x_i)
\end{aligned}
$$

Posons la dérivée première en la solution :

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial p}(\widehat{p}; x) &= 0 \\
\Leftrightarrow \frac{1}{\widehat{p}} \sum_{i=1}^{n} x_i - \frac{1}{1-\widehat{p}} \sum_{i=1}^{n} (1-x_i) = 0 \\
\Leftrightarrow \frac{1}{\widehat{p}} \sum_{i=1}^{n} x_i = \frac{1}{1-\widehat{p}} \sum_{i=1}^{n} (1-x_i) \\
\Leftrightarrow \frac{1-\widehat{p}}{\widehat{p}} = \frac{\sum_{i=1}^{n} (1-x_i)}{\sum_{i=1}^{n} x_i} \\
\Leftrightarrow \frac{1}{\widehat{p}} - 1 = \frac{n - \sum_{i=1}^{n} (x_i)}{\sum_{i=1}^{n} x_i} \\
\Leftrightarrow \frac{1}{\widehat{p}} - 1 = \frac{n}{\sum_{i=1}^{n} x_i} - 1 \\
\Leftrightarrow \frac{1}{\widehat{p}} = \frac{n}{\sum_{i=1}^{n} x_i} \\
\Leftrightarrow \widehat{p} = \frac{\sum_{i=1}^{n} x_i}{n} \equiv \bar{x}
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial p^2}(p;x) &= \frac{\partial \left(\dfrac{\partial \mathcal{L}}{\partial p}(p;x) \right)}{\partial p} \\
&= \frac{\partial \left[\dfrac{1}{p} \sum_{i=1}^{n} x_i - \dfrac{1}{1-p} \sum_{i=1}^{n} (1-x_i) \right]}{\partial p} \\
&= \frac{\partial \left[\dfrac{1}{p} \sum_{i=1}^{n} x_i \right]}{\partial p} - \frac{\partial \left[\frac{1}{1-p} \sum_{i=1}^{n} (1-x_i) \right]}{\partial p} \\
&= \left(\sum_{i=1}^{n} x_i \right) \times \frac{\partial \left(\frac{1}{p} \right)}{\partial p} - \left(\sum_{i=1}^{n} (1-x_i) \right) \times \frac{\partial \left(\frac{1}{1-p} \right)}{\partial p} \\
&= \left(\sum_{i=1}^{n} x_i \right) \times \frac{-1}{p^2} - \left(\sum_{i=1}^{n} (1-x_i) \right) \times \frac{1}{(1-p)^2} \\
&= - \left(\underbrace{\frac{\sum_{i=1}^{n} x_i}{p^2} + \frac{\sum_{i=1}^{n} (1-x_i)}{(1-p)^2}}_{>0} \right) < 0
\end{aligned}
$$

C'est donc un maximum global.

Remarque

On notera :

$$\widehat{p} = \widehat{p}(x) = \overline{x} \text{ : l'estimation = un nombre}$$

$$\widehat{/p} = \widehat{p}(X) = \overline{X} \text{ : l'estimateur = une VA}$$

Soit $(X_1, X_2, \ldots, X_n)$ une suite de n variables aléatoires continues identiquement et indépendamment distribuées selon la densité de probabilités

$$f(x_i;\theta) = \begin{cases} \frac{x_i^{\frac{1}{\theta}-1}}{\theta c^{\frac{1}{\theta}}} & \text{si } x_i \in [0, c] \\ 0 & \text{sinon} \end{cases}$$

avec c une constante connue.

On veut estimer le paramètre θ par maximum de vraisemblance :

$$\widehat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta; x) = \sum_{i=1}^{n} \ln[f(x_i;\theta)].$$

Calculons $\ln[f(x_i;\theta)]$ en utilisant les propriétés de $\ln(.)$ à savoir :

$$\ln\left(\frac{A}{B}\right) = \ln(A) - \ln(B)$$
$$\ln(AB) = \ln(A) + \ln(B)$$
$$\ln(A^\alpha) = \alpha \ln(A)$$

$$\begin{aligned}
\ln f(x_i;\theta) &= \ln\left(\frac{x_i^{\frac{1}{\theta}-1}}{\theta c^{\frac{1}{\theta}}}\right) \\
&= \ln\left(x_i^{\frac{1}{\theta}-1}\right) - \ln\left(\theta c^{\frac{1}{\theta}}\right) \\
&= \left(\frac{1}{\theta}-1\right)\ln(x_i) - \ln(\theta) - \ln(c^{\frac{1}{\theta}}) \qquad \text{attention à la priorité !} \\
&= \left(\frac{1}{\theta}-1\right)\ln(x_i) - \ln(\theta) - \frac{1}{\theta}\ln(c).
\end{aligned}$$

En utilisant les propriétés des sommes, on en déduit l'expression du critère à maximiser $\mathcal{L}(\theta)$:

$$\begin{aligned}
\mathcal{L}(\theta; x) &= \sum_{i=1}^{n} \ln f(x_i; \theta) \\
&= \sum_{i=1}^{n} \left[\left(\frac{1}{\theta} - 1 \right) \ln(x_i) - \ln(\theta) - \frac{1}{\theta} \ln(c) \right] \\
&= \left(\frac{1}{\theta} - 1 \right) \sum_{i=1}^{n} \ln(x_i) - \sum_{i=1}^{n} \ln(\theta) - \sum_{i=1}^{n} \frac{1}{\theta} \ln(c) \\
&= \left(\frac{1}{\theta} - 1 \right) \sum_{i=1}^{n} \ln(x_i) - n \ln(\theta) - \frac{n}{\theta} \ln(c).
\end{aligned}$$

Calculons la dérivée première :

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \theta}(\theta; x) &= \frac{\partial \left[\left(\frac{1}{\theta} - 1\right) \sum_{i=1}^{n} \ln(x_i) - n \ln(\theta) - \frac{n}{\theta} \ln(c)\right]}{\partial \theta} \\
&= \frac{\partial \left[\frac{1}{\theta} \sum_{i=1}^{n} \ln(x_i)\right]}{\partial \theta} - \underbrace{\frac{\partial \left[\sum_{i=1}^{n} \ln(x_i)\right]}{\partial \theta}}_{=0} - \frac{\partial \left[n \ln(\theta)\right]}{\partial \theta} - \frac{\partial \left[\frac{n}{\theta} \ln(c)\right]}{\partial \theta} \\
&= \left(\sum_{i=1}^{n} \ln(x_i)\right) \frac{\partial \left[\frac{1}{\theta}\right]}{\partial \theta} - n \frac{\partial \left[\ln(\theta)\right]}{\partial \theta} - n \ln(c) \frac{\partial \left[\frac{1}{\theta}\right]}{\partial \theta} \\
&= -\frac{1}{\theta^2} \sum_{i=1}^{n} \ln(x_i) - \frac{n}{\theta} + \frac{n \ln(c)}{\theta^2}
\end{aligned}
$$

On en déduit la CPO :

$$
\begin{aligned}
&\frac{\partial \mathcal{L}}{\partial \theta}(\widehat{\theta}; x) \quad = 0 \\
\Leftrightarrow & -\frac{1}{\widehat{\theta}^2} \sum_{i=1}^{n} \ln(x_i) - \frac{n}{\widehat{\theta}} + \frac{n \ln(c)}{\widehat{\theta}^2} = 0 \\
\Leftrightarrow & -\frac{1}{\widehat{\theta}^2} \sum_{i=1}^{n} \ln(x_i) - \frac{n\widehat{\theta}}{\widehat{\theta}^2} + \frac{n \ln(c)}{\widehat{\theta}^2} = 0 \\
\Leftrightarrow & -\sum_{i=1}^{n} \ln(x_i) - n\widehat{\theta} + n \ln(c) = 0 \\
\Leftrightarrow & \; n\widehat{\theta} = n \ln(c) - \sum_{i=1}^{n} \ln(x_i) \\
\Leftrightarrow & \; \widehat{\theta} = \ln(c) - \frac{1}{n} \sum_{i=1}^{n} \ln(x_i)
\end{aligned}
$$

Repartons de :

$$\frac{\partial \mathcal{L}}{\partial \theta}(\theta; x) = -\frac{1}{\theta^2}\sum_{i=1}^{n} \ln(x_i) - \frac{n}{\theta} + \frac{n\ln(c)}{\theta^2}$$

et dérivons à nouveau par rapport à θ :

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \theta^2}(\theta; x) &= \frac{\partial\left(-\frac{1}{\theta^2}\sum_{i=1}^{n} \ln(x_i) - \frac{n}{\theta} + \frac{n\ln(c)}{\theta^2}\right)}{\partial \theta} \\
&= \frac{\partial\left(-\theta^{-2}\sum_{i=1}^{n} \ln(x_i) - n\theta^{-1} + n\ln(c)\theta^{-2}\right)}{\partial \theta} \\
&= +2\theta^{-3}\sum_{i=1}^{n} \ln(x_i) + n\theta^{-2} - 2n\ln(c)\theta^{-3} \\
&= \frac{2}{\theta^3}\sum_{i=1}^{n} \ln(x_i) + \frac{n}{\theta^2} - \frac{2n\ln(c)}{\theta^3} \\
&= \frac{2\sum_{i=1}^{n} \ln(x_i) - 2n\ln(c)}{\theta^3} + \frac{n}{\theta^2} \\
&= 2n\frac{\frac{1}{n}\sum_{i=1}^{n} \ln(x_i) - \ln(c)}{\theta^3} + \frac{n}{\theta^2}
\end{aligned}$$

Le signe n'est pas clair. Evaluons l'expression en $\theta = \widehat{\theta}$:

$$
\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \theta^2}(\widehat{\theta}; x) &= 2n \frac{\overbrace{\frac{1}{n}\sum_{i=1}^{n} \ln(x_i) - \ln(c)}^{=-\widehat{\theta}}}{\widehat{\theta}^3} + \frac{n}{\widehat{\theta}^2} \\
&= -2n\frac{\widehat{\theta}}{\widehat{\theta}^3} + \frac{n}{\widehat{\theta}^2} \\
&= -\frac{2n}{\widehat{\theta}^2} + \frac{n}{\widehat{\theta}^2} \\
&= -\frac{n}{\widehat{\theta}^2} < 0.
\end{aligned}
$$

La dérivée seconde en le point candidat étant forcément négative ($n > 0$, $\widehat{\theta}^2 > 0$), on est donc bien à un maximum local.

Soit une suite de n variables aléatoires $(X_1, X_2, \ldots, X_n)$ iid selon une loi normale $\mathcal{N}(m, \sigma^2)$ de densité de probabilité

$$\phi_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

On veut estimer les paramètres m et σ^2 en résolvant le programme du maximum de vraisemblance :

$$\widehat{m}, \widehat{\sigma}^2 = \arg\max_{m,\sigma^2} \mathcal{L}(m, \sigma^2; x) = \sum_{i=1}^{n} \ln \phi_{m,\sigma^2}(x_i)$$

$$\max_{m,\sigma^2} \mathcal{L}(m,\sigma^2;x) = \sum_{i=1}^{n} \ln \phi_{m,\sigma^2}(x_i)$$

Calculons d'abord le log de la fonction :

$$\begin{aligned}
\ln \phi_{m,\sigma^2}(x_i) &= \ln\left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-m)^2}{2\sigma^2}}\right] \\
&= \ln\left[\frac{1}{\sqrt{2\pi\sigma^2}}\right] + \ln\left[e^{-\frac{(x_i-m)^2}{2\sigma^2}}\right] \\
&= -\ln\left[\left(2\pi\sigma^2\right)^{\frac{1}{2}}\right] - \frac{(x_i-m)^2}{2\sigma^2} \\
&= -\frac{1}{2}\ln\left(2\pi\sigma^2\right) - \frac{(x_i-m)^2}{2\sigma^2} \\
&= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\sigma^2\right) - \frac{(x_i-m)^2}{2\sigma^2}
\end{aligned}$$

Mettons le résultat dans la somme pour obtenir la fonction à maximiser (et donc à dériver) :

$$\begin{aligned}
\mathcal{L}(m, \sigma^2; x) &= \sum_{i=1}^{n} \ln \phi_{m,\sigma^2}(x_i) \\
&= \sum_{i=1}^{n} \left[-\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln\left(\sigma^2\right) - \frac{(x_i - m)^2}{2\sigma^2} \right] \\
&= \sum_{i=1}^{n} -\frac{1}{2} \ln(2\pi) - \sum_{i=1}^{n} \frac{1}{2} \ln\left(\sigma^2\right) - \sum_{i=1}^{n} \frac{(x_i - m)^2}{2\sigma^2} \\
&= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln\left(\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - m)^2
\end{aligned}$$

$$\widehat{m}, \widehat{\sigma}^2 = \arg\max_{m,\sigma^2} \mathcal{L}(m, \sigma^2; x)$$

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial m}(\widehat{m}, \widehat{\sigma}^2; x) = 0 \\[2ex] \dfrac{\partial \mathcal{L}}{\partial (\sigma^2)}(\widehat{m}, \widehat{\sigma}^2; x) = 0 \end{cases}$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial m}(m, \sigma^2; x) &= \frac{\partial \left[-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2\right]}{\partial m} \\
&= \underbrace{\frac{\partial \left[-\frac{n}{2}\ln(2\pi)\right]}{\partial m}}_{=0} - \underbrace{\frac{\partial \left[\frac{n}{2}\ln\left(\sigma^2\right)\right]}{\partial m}}_{=0} - \frac{\partial \left[\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2\right]}{\partial m} \\
&= -\frac{1}{2\sigma^2}\frac{\partial \left[\sum_{i=1}^{n}(x_i - m)^2\right]}{\partial m} \\
&= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{\partial \left[(x_i - m)^2\right]}{\partial m} \\
&= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}2(-1)(x_i - m) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - m)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial(\sigma^2)}(m, \sigma^2; x) &= \frac{\partial\left[-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2\right]}{\partial(\sigma^2)} \\
&= \underbrace{\frac{\partial\left[-\frac{n}{2}\ln(2\pi)\right]}{\partial(\sigma^2)}}_{=0} - \frac{\partial\left[\frac{n}{2}\ln\left(\sigma^2\right)\right]}{\partial(\sigma^2)} - \frac{\partial\left[\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2\right]}{\partial(\sigma^2)} \\
&= -\frac{n}{2}\frac{\partial\left[\ln\left(\sigma^2\right)\right]}{\partial(\sigma^2)} - \left[\frac{1}{2}\sum_{i=1}^{n}(x_i - m)^2\right]\frac{\partial\left[\frac{1}{\sigma^2}\right]}{\partial(\sigma^2)} \\
&= -\frac{n}{2}\frac{1}{\sigma^2} - \left[\frac{1}{2}\sum_{i=1}^{n}(x_i - m)^2\right]\left(-\frac{1}{\sigma^4}\right) \\
&= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - m)^2
\end{aligned}
$$

$$
\begin{cases}
\quad \frac{\partial \mathcal{L}}{\partial m}(\widehat{m}, \widehat{\sigma}^2; x) = 0 \\
\quad \frac{\partial \mathcal{L}}{\partial (\sigma^2)}(\widehat{m}, \widehat{\sigma}^2; x) = 0
\end{cases}
$$

$$
\Leftrightarrow \begin{cases}
\frac{1}{\widehat{\sigma}^2} \sum_{i=1}^{n} (x_i - \widehat{m}) = 0 \\
-\frac{n}{2} \frac{1}{\widehat{\sigma}^2} + \frac{1}{2\widehat{\sigma}^4} \sum_{i=1}^{n} (x_i - \widehat{m})^2 = 0
\end{cases}
$$

$$
\Leftrightarrow \begin{cases}
\sum_{i=1}^{n} (x_i - \widehat{m}) = 0 \\
\frac{1}{2\widehat{\sigma}^2} \left[-n + \frac{1}{\widehat{\sigma}^2} \sum_{i=1}^{n} (x_i - \widehat{m})^2 \right] = 0
\end{cases}
$$

$$
\Leftrightarrow \begin{cases}
\sum_{i=1}^{n} (x_i - \widehat{m}) = 0 \\
-n + \frac{1}{\widehat{\sigma}^2} \sum_{i=1}^{n} (x_i - \widehat{m})^2 = 0
\end{cases}
$$

La première équation peut se résoudre seule !

$$
\begin{aligned}
& \sum_{i=1}^{n} (x_i - \widehat{m}) = 0 \\
\Leftrightarrow\ & \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \widehat{m} = 0 \\
\Leftrightarrow\ & \sum_{i=1}^{n} x_i - n\widehat{m} = 0 \\
\Leftrightarrow\ & \sum_{i=1}^{n} x_i = n\widehat{m} \\
\Leftrightarrow\ & \widehat{m} = \frac{1}{n} \sum_{i=1}^{n} x_i \equiv \bar{x}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial(\sigma^2)}(\widehat{m},\widehat{\sigma}^2;x)=0 &\Leftrightarrow -n+\frac{1}{\widehat{\sigma}^2}\sum_{i=1}^{n}(x_i-\widehat{m})^2=0\\
&\Leftrightarrow -n+\frac{1}{\widehat{\sigma}^2}\sum_{i=1}^{n}(x_i-\bar{x})^2=0\\
&\Leftrightarrow \frac{1}{\widehat{\sigma}^2}\sum_{i=1}^{n}(x_i-\bar{x})^2=n\\
&\Leftrightarrow \widehat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\equiv V(x)
\end{aligned}
$$

$$H(m,\sigma^2;x) = \begin{pmatrix} \frac{\partial \mathcal{L}^2}{\partial m^2}(m,\sigma^2;x) & \frac{\partial \mathcal{L}^2}{\partial m \partial(\sigma^2)}(m,\sigma^2;x) \\ \frac{\partial \mathcal{L}^2}{\partial(\sigma^2)\partial m}(m,\sigma^2;x) & \frac{\partial \mathcal{L}^2}{\partial(\sigma^2)^2}(m,\sigma^2;x) \end{pmatrix}$$

$$\begin{aligned} \frac{\partial \mathcal{L}^2}{\partial m^2}(m,\sigma^2;x) = \frac{\partial\left(\frac{\partial \mathcal{L}}{\partial m}(m,\sigma^2;x)\right)}{\partial m} &= \frac{\partial\left(\frac{1}{\sigma^2}\sum_{i=1}^n (x_i - m)\right)}{\partial m} \\ &= \frac{1}{\sigma^2}\frac{\partial\left(\sum_{i=1}^n (x_i - m)\right)}{\partial m} \\ &= \frac{1}{\sigma^2}\sum_{i=1}^n \frac{\partial\,(x_i - m)}{\partial m} \\ &= \frac{1}{\sigma^2}\sum_{i=1}^n -1 \\ &= -\frac{n}{\sigma^2} \end{aligned}$$

$$\frac{\partial \mathcal{L}^2}{\partial m^2}(\widehat{m},\widehat{\sigma}^2;x) = -\frac{n}{\widehat{\sigma}^2}$$

$$H(m,\sigma^2;x) = \begin{pmatrix} \frac{\partial \mathcal{L}^2}{\partial m^2}(m,\sigma^2;x) & \frac{\partial \mathcal{L}^2}{\partial m \partial(\sigma^2)}(m,\sigma^2;x) \\ \frac{\partial \mathcal{L}^2}{\partial(\sigma^2)\partial m}(m,\sigma^2;x) & \frac{\partial \mathcal{L}^2}{\partial(\sigma^2)^2}(m,\sigma^2;x) \end{pmatrix}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}^2}{\partial m \partial(\sigma^2)}(m,\sigma^2;x) = \frac{\partial\left(\frac{\partial \mathcal{L}}{\partial m}(m,\sigma^2;x)\right)}{\partial(\sigma^2)} &= \frac{\partial\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - m)\right)}{\partial(\sigma^2)} \\
&= \left[\sum_{i=1}^{n}(x_i - m)\right]\frac{\partial\left(\frac{1}{\sigma^2}\right)}{\partial(\sigma^2)} \\
&= \left[\sum_{i=1}^{n}(x_i - m)\right]\left(-\frac{1}{(\sigma^2)^2}\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}^2(m,\sigma^2)}{\partial m \partial(\sigma^2)}(\widehat{m},\widehat{\sigma}^2;x) &= \left[\underbrace{\sum_{i=1}^{n}(x_i - \bar{x})}_{=0}\right]\left(-\frac{1}{\widehat{\sigma}^4}\right) \\
&= 0
\end{aligned}$$

$$H(m, \sigma^2; x) = \begin{pmatrix} \frac{\partial \mathcal{L}^2}{\partial m^2}(m, \sigma^2) & \frac{\partial \mathcal{L}^2}{\partial m \partial(\sigma^2)}(m, \sigma^2) \\ \color{red}{\frac{\partial \mathcal{L}^2}{\partial(\sigma^2)\partial m}(m, \sigma^2)} & \frac{\partial \mathcal{L}^2}{\partial(\sigma^2)^2(m, \sigma^2)} \end{pmatrix}$$

$$\frac{\partial \mathcal{L}^2}{\partial(\sigma^2)\partial m}(m, \sigma^2) = \frac{\partial \mathcal{L}^2}{\partial m \partial(\sigma^2)}(m, \sigma^2)$$

$$H(m,\sigma^2;x) = \begin{pmatrix} \frac{\partial \mathcal{L}^2}{\partial m^2}(m,\sigma^2) & \frac{\partial \mathcal{L}^2}{\partial m \partial(\sigma^2)}(m,\sigma^2) \\ \frac{\partial \mathcal{L}^2}{\partial(\sigma^2)\partial m}(m,\sigma^2) & \frac{\partial \mathcal{L}^2}{\partial(\sigma^2)^2}(m,\sigma^2) \end{pmatrix}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}^2}{\partial(\sigma^2)^2}(m,\sigma^2) = \frac{\partial\left(\frac{\partial \mathcal{L}}{\partial(\sigma^2)}(m,\sigma^2)\right)}{\partial(\sigma^2)} &= \frac{\partial\left(-\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (x_i-m)^2\right)}{\partial(\sigma^2)} \\
&= \frac{\partial\left(-\frac{n}{2}\frac{1}{\sigma^2}\right)}{\partial(\sigma^2)} + \frac{\partial\left(\frac{1}{2\sigma^4}\sum_{i=1}^n (x_i-m)^2\right)}{\partial(\sigma^2)} \\
&= -\frac{n}{2}\frac{\partial\left(\frac{1}{\sigma^2}\right)}{\partial(\sigma^2)} + \frac{1}{2}\left[\sum_{i=1}^n (x_i-m)^2\right]\frac{\partial\left(\frac{1}{(\sigma^2)^2}\right)}{\partial(\sigma^2)} \\
&= -\frac{n}{2}\left(-\frac{1}{\sigma^4}\right) + \frac{1}{2}\left[\sum_{i=1}^n (x_i-m)^2\right]\left(\frac{-2}{(\sigma^2)^3}\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}^2}{\partial(\sigma^2)^2}(\widehat{m},\widehat{\sigma}^2;x) &= \frac{n}{2\widehat{\sigma}^4} - \frac{1}{\widehat{\sigma}^6}\sum_{i=1}^n (x_i-\bar{x})^2 = \frac{n}{2\widehat{\sigma}^4} - \frac{n\widehat{\sigma}^2}{\widehat{\sigma}^6} \\
&= \frac{n}{2\widehat{\sigma}^4} - \frac{2n}{2\widehat{\sigma}^4} = -\frac{n}{2\widehat{\sigma}^4}
\end{aligned}$$

$$H(\widehat{m}, \widehat{\sigma}^2; x) = \begin{pmatrix} -\dfrac{n}{\widehat{\sigma}^2} & 0 \\ 0 & -\dfrac{n}{2\widehat{\sigma}^4} \end{pmatrix}$$

$$\det(H(\widehat{m}, \widehat{\sigma}^2; x)_1) = -\frac{n}{\widehat{\sigma}^2} < 0$$
$$\det(H(\widehat{m}, \widehat{\sigma}^2; x)_2) = -\frac{n}{\widehat{\sigma}^2} \times -\frac{n}{2\widehat{\sigma}^4} = \frac{n^2}{2\widehat{\sigma}^6} > 0$$

Alternance de signes des mineurs en commençant par le négatif.

Donc la matrice est définie négative.

Donc on est bien à un maximum local...

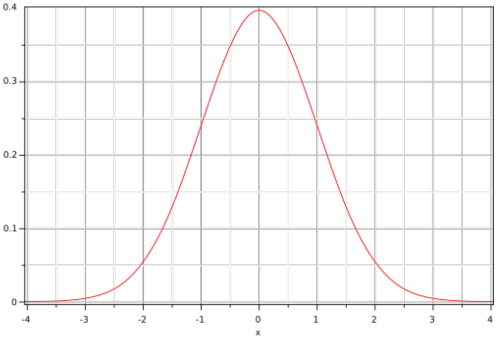
II- Propriétés d’un estimateur

La loi normale

$$Y \hookrightarrow \mathcal{N}(\mu, \sigma^2)$$

$$X = \frac{Y - \mu}{\sigma} \hookrightarrow \mathcal{N}(0, 1)$$

$$Y = \mu + \sigma X$$



Soit n lois normales indépendantes $X_i \rightsquigarrow \mathcal{N}(\mu_i, \sigma_i^2)$:

$$\sum_{i=1}^{n} X_i \rightsquigarrow \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Définition d'une loi du Chi2

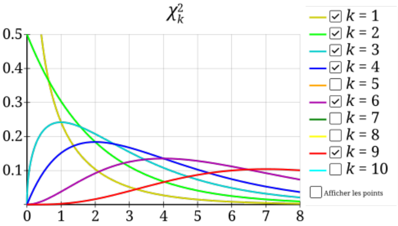
$$\forall i \in [1, \ldots, n],\ X_i \underset{iid}{\hookrightarrow} \mathcal{N}(0, 1)$$

$$Z = \sum_{i=1}^{n} X_i^2 \hookrightarrow \chi^2(n)$$

Propriétés intéressantes :

$$\mathbb{E}\left(\chi^2(n)\right) = n$$

$$\mathbb{V}\left(\chi^2(n)\right) = 2n$$



On dit que la suite de VA X_i converge en probabilité vers une VA X si $\forall\epsilon > 0$:

$$\mathbb{P}\left(|X_n - X| < \epsilon\right) \underset{n\to+\infty}{\longrightarrow} 1$$

ou encore

$$\mathbb{P}\left(|X_n - X| > \epsilon\right) \underset{n\to+\infty}{\longrightarrow} 0$$

On écrit alors

$$X_n \xrightarrow{p} X$$

Pour le montrer, il suffit d'utiliser la condition suffisante suivante :

$$\mathbb{E}(X_n) \underset{n\to+\infty}{\longrightarrow} a$$

$$\mathbb{V}(X_n) \underset{n\to+\infty}{\longrightarrow} 0$$

$$\Rightarrow X_n \xrightarrow{p} a$$

On dit que la suite de VA X_i de fonction de répartition F_i converge en loi vers une VA X de fonction de répartition F si la suite $F_i(x)$ converge vers $F(x)$ en tout point où F est continue. On écrit alors

$$X_n \xrightarrow{L} X$$

Pour le montrer, il suffit d'utiliser la convergence en probabilités :

$$X_n \xrightarrow{p} X \Rightarrow \; X_n \xrightarrow{L} X$$

Soit une suite de n VAC iid $(X_1, X_2, \ldots, X_n)$, chacune d'espérance m et de variance σ^2.

Soit $\overline{X} = \dfrac{1}{n} \sum_{i=1}^{n} X_i$.

La loi faible des grands nombres :

$$\left.\begin{array}{l} \mathbb{E}(\overline{X}) = \mathbb{E}(X) = m \\ \mathbb{V}(\overline{X}) = \frac{\sigma^2}{n} \underset{n \to +\infty}{\longrightarrow} 0 \end{array}\right. \Rightarrow \overline{X} \xrightarrow{p} \mathbb{E}(X) = m$$

En dilatant l'estimateur pour obtenir une VA avec une variance finie :

$$\mathbb{E}\left(\sqrt{n}\,\overline{X}\right) = \sqrt{n}\,\mathbb{E}(\overline{X}) = \sqrt{n}\,m$$
$$\mathbb{V}\left(\sqrt{n}\,\overline{X}\right) = n\,\mathbb{V}(\overline{X}) = \sigma^2$$

Le théorème central-limite :

$$\frac{\sqrt{n}\,\overline{X} - \mathbb{E}(\sqrt{n}\,\overline{X})}{\sqrt{\mathbb{V}(\sqrt{n}\,\overline{X})}} = \frac{\sqrt{n}(\overline{X} - m)}{\sigma} \overset{L}{\rightsquigarrow} \mathcal{N}(0, 1)$$

En utilisant les propriétés de la loi normale, on peut approximer les lois :

$$\frac{\sqrt{n}\,\overline{X} - \mathbb{E}(\sqrt{n}\,\overline{X})}{\sqrt{\mathbb{V}(\sqrt{n}\,\overline{X})}} = \frac{\sqrt{n}(\overline{X} - m)}{\sigma} \overset{L}{\rightsquigarrow} \mathcal{N}(0,1)$$

$$\Leftrightarrow \frac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}} \overset{L}{\rightsquigarrow} \mathcal{N}(0,1)$$

$$\Leftrightarrow \overline{X} \overset{L}{\rightsquigarrow} \mathcal{N}(m, \frac{\sigma^2}{n})$$

Soit la loi Binomiale $Z \rightsquigarrow B(n, p)$. Elle peut s'écrire comme la somme de n lois de Bernoulli iid $X_i \rightsquigarrow B(p)$.

$$Z = \sum_{i=1}^{n} X_i = n\overline{X_n}$$

$$\begin{array}{ll} \mathbb{E}(X_i) = p & \mathbb{V}(X_i) = p(1-p) \\ \mathbb{E}(Z) = np & \mathbb{V}(Z) = np(1-p) \\ \mathbb{E}(\overline{X_n}) = p & \mathbb{V}(\overline{X_n}) = \frac{p(1-p)}{n} \end{array}$$

On a donc : $\overline{X_n} \xrightarrow{p} p$ puisque $\mathbb{V}(\overline{X_n}) \underset{n\to+\infty}{\longrightarrow} 0$.

Le théorème central-limite : $\sqrt{n}\, \frac{\overline{X_n}-p}{\sqrt{p(1-p)}} \xrightarrow{L} \mathcal{N}(0,1)$. Il vient :

$$\sqrt{n}\, \frac{Z/n - p}{\sqrt{p(1-p)}} = n\, \frac{Z/n - p}{\sqrt{np(1-p)}} = \frac{Z - np}{\sqrt{np(1-p)}} \xrightarrow{L} \mathcal{N}(0,1)$$

On peut donc approximer une loi $Z \rightsquigarrow B(n, p)$ par une loi normale $\mathcal{N}(np, np(1-p))$ (généralement si $n \geqslant 30$, $np \geqslant 5$, $n(1-p) \geqslant 5$.)

Soit la loi Poisson $Z \rightsquigarrow P(\lambda)$. Elle peut s'écrire comme la somme de λ lois de Poisson iid $X_i \rightsquigarrow P(1)$.

$$Z = \sum_{i=1}^{\lambda} X_i = \lambda \overline{X_\lambda}$$

$$\begin{array}{rr} \mathbb{E}(X_i) = 1 & \mathbb{V}(X_i) = 1 \\ \mathbb{E}(Z) = \lambda & \mathbb{V}(Z) = \lambda \\ \mathbb{E}(\overline{X_n}) = 1 & \mathbb{V}(\overline{X_n}) = \frac{1}{\lambda} \end{array}$$

On a donc : $\overline{X_\lambda} \xrightarrow{p} 1$ puisque $\mathbb{V}(\overline{X_n}) \underset{\lambda \to +\infty}{\longrightarrow} 0$.

Le théorème central-limite : $\sqrt{\lambda}\, \frac{\overline{X_\lambda} - 1}{1} \xrightarrow{L} \mathcal{N}(0, 1)$. Il vient :

$$\sqrt{\lambda}\, \frac{Z/\lambda - 1}{1} = \lambda\, \frac{Z/\lambda - 1}{\sqrt{\lambda}} = \frac{Z - \lambda}{\sqrt{\lambda}} \xrightarrow{L} \mathcal{N}(0, 1)$$

On peut donc approximer une loi $Z \rightsquigarrow P(\lambda)$ par une loi normale $\mathcal{N}(\lambda, \lambda)$ (généralement si λ assez grand).

Propriétés théoriques d’un estimateur

- $\widehat{\lambda}$ est un estimateur de λ

- C’est une variable aléatoire.

- Il doit posséder des propriétés théoriques intéressantes

1. **Sans biais** (= propriété théorique de l'espérance de la VA) :

$$\mathbb{E}(\widehat{\lambda}) = \lambda \text{ (le paramètre inconnu)}$$

2. **Précision** (= propriété théorique de la variance de la VA) :

$$\mathbb{V}(\widehat{\lambda}) = \frac{\sigma^2}{n}$$

3. **Convergent en proba** (= propriété théorique de la distribution de la VA) :

$$\mathbb{V}(\widehat{\lambda}) \underset{n \longrightarrow +\infty}{\longrightarrow} 0 \Rightarrow \widehat{\lambda} \xrightarrow{p} \lambda \Rightarrow \sqrt{n}(\widehat{\lambda} - \lambda) \overset{L}{\rightsquigarrow} \mathcal{N}(0, \sigma^2)$$

Remarque : ces propriétés concernent tous les estimateurs, quelque que soit la loi sous-jacente.

4. **Efficace si** sa variance = la borne CRFD :

$$\mathbb{V}(\widehat{\lambda}) = \frac{1}{I(\lambda; X)}$$

avec

$$I(\lambda; X) = \mathbb{E}\left(-\frac{\partial^2 \mathcal{L}}{\partial \lambda^2}(\lambda; X)\right) = \mathbb{V}\left(\frac{\partial \mathcal{L}}{\partial \lambda}(\lambda; X)\right) = \mathbb{E}\left(\frac{\partial \mathcal{L}}{\partial \lambda}(\lambda; X)\right)^2$$

On appelle $I(\lambda; X)$ **l'information de Fisher** et $\frac{1}{I(\lambda; X)}$ **la borne de Cramer-Rao-Frechet-Darmois**.

Pour cette propriété, il est nécessaire de faire l'hypothèse d'une loi sous-jacente pour l'échantillon pour écrire la vraisemblance et en déduire l'information de Fisher et la borne CRFD.

Exemple

La densité sous-jacente : $f_X(\theta; x_i) = \dfrac{1}{\theta} \exp\left(-\dfrac{x_i}{\theta}\right)$

La log-vraisemblance de l'échantillon : $\mathcal{L}(\theta; x) = -n \ln(\theta) - \dfrac{1}{\theta} \sum_{i=1}^n x_i$

Le gradient : $g(\theta; x) = \dfrac{\partial \mathcal{L}}{\partial \theta}(\theta; x) = -\dfrac{n}{\theta} + \dfrac{1}{\theta^2} \sum_{i=1}^n x_i$

L'estimateur : $\widehat{\theta}(X) = \overline{X}$, d'espérance θ et de variance $\dfrac{\theta^2}{n}$

Le score : $g(\theta; X) = \dfrac{\partial \mathcal{L}}{\partial \theta}(\theta; X) = -\dfrac{n}{\theta} + \dfrac{1}{\theta^2} \sum_{i=1}^n X_i$ avec $\mathbb{E}\left(g(\theta; X)\right) = 0$.

La Hessienne : $H(\theta; X) = \dfrac{\partial^2 \mathcal{L}}{\partial \theta^2}(\theta; X) = \dfrac{n}{\theta^2} + \dfrac{2}{\theta^3} \sum_{i=1}^n X_i$

L'information de Fisher : $I(\theta; X) = \mathbb{E}\left(-H(\theta; X)\right) = \mathbb{V}\left(g(\theta; X)\right) = \dfrac{n}{\theta^2}$

Comme $\mathbb{V}(\widehat{\theta}(X)) = \dfrac{1}{I(\theta; X)}$, $\widehat{\theta}(X)$ est efficace.

Soit une suite de n variables aléatoires iid $(X_1, X_2, \ldots, X_n)$.

Supposons que leur espérance est m et leur variance σ^2.

Cela signifie que ces n variables aléatoires sont toutes identiques entre elles et ont toutes la même loi, la même espérance et la même variance qu'une variable X par exemple. C'est comme si on avait répété l'expérience aléatoire liée à X, N fois de façon indépendante.

On a vu que dans beaucoup de cas, quand on voulait estimer un paramètre inconnu qui est l'espérance de la loi sous-jacente $\mathbb{E}(X) = m$, l'estimateur est souvent la moyenne arithmétique des variables aléatoires sous-jacentes $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

$\overline{X}$ est une variable aléatoire comme combinaison linéaire de variables aléatoires.

Il est facile de montrer que cet estimateur est sans biais et convergent en probabilités.

Espérance de $\overline{X}$

$$\begin{aligned}
\mathbb{E}\left(\overline{X}\right) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(X_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(X\right) \quad \text{(car VA identiques)} \\
&= \frac{1}{n} n\, \mathbb{E}\left(X\right) \\
&= \mathbb{E}\left(X\right) = m
\end{aligned}$$

Variance de $\overline{X}$

$$
\begin{aligned}
\mathbb{V}\left(\overline{X}\right) &= \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\left(X_i\right) \quad \text{(car VA indépendantes)} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\left(X\right) \quad \text{(car VA identiques)} \\
&= \frac{1}{n^2}\, n\,\mathbb{V}\left(X\right) \\
&= \frac{\mathbb{V}\left(X\right)}{n} = \frac{\sigma^2}{n}
\end{aligned}
$$

Que se passe-t-il lorsque $n \to \infty$?

$$\lim_{n \to +\infty} \mathbb{V}\left(\overline{X}\right) = \lim_{n \to +\infty} \frac{\sigma^2}{n} = 0$$

On en déduit donc que :

$$\overline{X} \xrightarrow{p} \mathbb{E}(X) = m$$

Plus la taille de l'échantillon augmente, plus la variance diminue jusqu'à valoir 0.

Cela implique que la variable aléatoire $\overline{X}$ tend vers la valeur de son espérance, sans aucun aléa possible.

Donc $\overline{X}$ est le meilleur estimateur de l'espérance de la variable aléatoire X, jusqu'à lui être égal si la taille de l'échantillon augmente indéfiniment.

Ces résultats dépendent-ils d'une loi de probabilités particulière ?

Ce résultat est indépendant d'un choix de variable aléatoire particulier. Il ne vaut ici que pour des VA iid.

C'est **la loi faible des grands nombres**, qu'on note : $\overline{X} \xrightarrow{p} \mathbb{E}(X)$. **Si on veut estimer l'espérance d'une VA X, il faut la répéter un très grand nombre de fois de façon indépendante et calculer la moyenne des résultats.**

Illustration de la convergence en proba de $\overline{X}$ avec la loi de Poisson et Julia

La convergence en loi résulte de la convergence en probabilités.

Comme on sait comment dégénère la loi, il suffit de la multiplier par $\sqrt{n}$ pour la VA $\overline{X}$ ait une variance qui ne tende plus vers 0 quand n tend vers $+\infty$.

On en déduit donc que $\sqrt{n}\,\overline{X}$ suit approximativement une loi normale (donc une loi non dégénérée car elle a une variance non nulle si n est assez grand).

C'est le **théorème central-limite**.

$$\frac{\sqrt{n}\,\overline{X} - \mathbb{E}(\sqrt{n}\,\overline{X})}{\sqrt{\mathbb{V}(\sqrt{n}\,\overline{X})}} \overset{L}{\rightsquigarrow} \mathcal{N}(0,1)$$

Pour s'en convaincre :

$$\mathbb{E}\left(\sqrt{n}\,\overline{X}\right) = \sqrt{n}\,\mathbb{E}(\overline{X}) = \sqrt{n}\,m$$
$$\mathbb{V}\left(\sqrt{n}\,\overline{X}\right) = N\,\mathbb{V}(\overline{X}) = n\,\mathbb{V}(\overline{X}) = \sigma^2$$

On en déduit donc :

$$\frac{\sqrt{n}\,\overline{X} - \mathbb{E}(\sqrt{n}\,\overline{X})}{\sqrt{\mathbb{V}(\sqrt{n}\,\overline{X})}} = \frac{\sqrt{n}(\overline{X} - m)}{\sigma} \overset{L}{\rightsquigarrow} \mathcal{N}(0, 1)$$
$$\Leftrightarrow \frac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}} \overset{L}{\rightsquigarrow} \mathcal{N}(0, 1)$$
$$\Leftrightarrow \sqrt{n}(\overline{X} - m) \overset{L}{\rightsquigarrow} \mathcal{N}(0, \sigma^2)$$
$$\Leftrightarrow \overline{X} \overset{L}{\rightsquigarrow} \mathcal{N}(m, \frac{\sigma^2}{n})$$

Résultat encore une fois indépendant d'un choix de variable aléatoire particulier.

Si la suite de n variables aléatoires identiquement et indépendamment distribuées $(X_1, X_2, \ldots, X_n)$ est $\mathcal{N}(m, \sigma^2)$, alors cela nous débloque un résultat important.

En utilisant les propriétés de la loi normale, il vient que $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ **suit une loi normale comme combinaison linéaire de lois normales**.

On a caractérisé son espérance et sa variance :

$$\mathbb{E}\left(\overline{X}\right) = m \quad \mathbb{V}\left(\overline{X}\right) = \frac{\sigma^2}{n}$$

On en déduit donc

$$\frac{\overline{X} - \mathbb{E}(\overline{X})}{\sqrt{\mathbb{V}(\overline{X})}} \rightsquigarrow \mathcal{N}(0, 1) \Leftrightarrow \overline{X} \rightsquigarrow \mathcal{N}(m, \frac{\sigma^2}{n})$$

Ce n'est plus une approximation !

Définissons la variance théorique :

$$\widehat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$$

Cet outil est intéressant (on verra pourquoi plus tard) mais inutilisable puisque m est inconnu. On ne peut donc pas calculer cette quantité.

Il faut donc définir un outil plus opérationnel.

Définissons à présent la variance empirique :

$$\widehat{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

On veut calculer $\mathbb{E}(S_n^2)$ pour voir s'il est sans biais.

$$\begin{aligned}\mathbb{E}(\widehat{S}_n^2) &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(X_i - \overline{X}\right)^2\end{aligned}$$

Essayons de voir ce que vaut $\mathbb{E}\left(X_i - \overline{X}\right)^2$.

Comment estimer la variance ?

$$
\begin{aligned}
\mathbb{E}\left(X_i - \overline{X}\right)^2 &= \mathbb{E}\left(X_i - m + m - \overline{X}\right)^2 \\
&= \mathbb{E}\left[(X_i - m) - (\overline{X} - m)\right]^2 \\
&= \mathbb{E}\left[(X_i - m)^2 + (\overline{X} - m)^2 - 2(X_i - m)(\overline{X} - m)\right] \\
&= \underbrace{\mathbb{E}(X_i - m)^2}_{=\mathbb{V}(X_i)} + \underbrace{\mathbb{E}(\overline{X} - m)^2}_{=\mathbb{V}(\overline{X})} - 2\mathbb{E}\left[(X_i - m)(\overline{X} - m)\right]
\end{aligned}
$$

Comment estimer la variance ?

$$\begin{aligned}
\mathbb{E}\left(X_i - \overline{X}\right)^2 &= \mathbb{V}(X_i) + \mathbb{V}(\overline{X}) - 2\mathbb{E}\left[(X_i - m)\left(\frac{1}{n}\sum_{j=1}^{n} X_j - m\right)\right] \\
&= \mathbb{V}(X_i) + \mathbb{V}(\overline{X}) - 2\mathbb{E}\left[(X_i - m)\frac{1}{n}\sum_{j=1}^{n}(X_j - m)\right] \\
&= \mathbb{V}(X_i) + \mathbb{V}(\overline{X}) - \frac{2}{n}\mathbb{E}\left[\sum_{j=1}^{n}(X_i - m)\left(X_j - m\right)\right] \\
&= \mathbb{V}(X_i) + \mathbb{V}(\overline{X}) - \frac{2}{n}\mathbb{E}\left[\underbrace{(X_i - m)^2}_{\text{si } i=j} + \sum_{j=1, j\neq i}^{n}(X_i - m)(X_j - m)\right] \\
&= \mathbb{V}(X_i) + \mathbb{V}(\overline{X}) - \frac{2}{n}\underbrace{\mathbb{E}\left[(X_i - m)^2\right]}_{=\mathbb{V}(X_i)} - \frac{2}{n}\mathbb{E}\left[\sum_{j=1, j\neq i}^{n}(X_i - m)(X_j - m)\right] \\
&= \mathbb{V}(X_i) + \mathbb{V}(\overline{X}) - \frac{2}{n}\mathbb{V}(X_i) - \frac{2}{n}\overbrace{\sum_{j=1, j\neq i}^{n}\underbrace{\mathbb{E}\left[(X_i - m)(X_j - m)\right]}_{=\mathbb{C}ov(X_i, X_j)=0}}^{=0}
\end{aligned}$$

Comment estimer la variance ?

$$\begin{aligned}
\mathbb{E}\left(X_i - \overline{X}\right)^2 &= \mathbb{V}(X_i) + \mathbb{V}(\overline{X}) - \frac{2}{n}\mathbb{V}(X_i) \\
&= \sigma^2 + \frac{\sigma^2}{n} - \frac{2}{n}\sigma^2 \\
&= \sigma^2 - \frac{1}{n}\sigma^2 \\
&= \left(1 - \frac{1}{n}\right)\sigma^2 \\
&= \left(\frac{n-1}{n}\right)\sigma^2
\end{aligned}$$

D'où

$$\mathbb{E}(\widehat{S}_n^2) = \left(\frac{n-1}{n}\right)\sigma^2 \neq \sigma^2$$

La variance empirique $\widehat{S}_n^2$ est donc un estimateur biaisé à distance finie de σ^2...

... mais asymptotiquement sans biais :

$$\mathbb{E}(\widehat{S}_n^2) = \left(\frac{n-1}{n}\right)\sigma^2 \underset{n\to\infty}{\to} \sigma^2$$

Ce résultat décevant permet cependant de trouver un estimateur sans biais pour la variance. Repartons de l'égalité trouvée et utilisons les propriétés de l'espérance :

$$
\begin{aligned}
& \mathbb{E}(\widehat{S}_n^2) = \left(\frac{n-1}{n}\right)\sigma^2 \\
\Leftrightarrow\ & \left(\frac{n}{n-1}\right)\mathbb{E}(\widehat{S}_n^2) = \sigma^2 \\
\Leftrightarrow\ & \left(\frac{n}{n-1}\right)\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right) = \sigma^2 \\
\Leftrightarrow\ & \mathbb{E}\left[\left(\frac{n}{n-1}\right)\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right] = \sigma^2 \\
\Leftrightarrow\ & \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right] = \sigma^2 \\
\Leftrightarrow\ & \mathbb{E}\left(\widehat{S}_{n-1}^2\right) = \sigma^2
\end{aligned}
$$

La variance empirique corrigée $\widehat{S}_{n-1}^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ est donc un estimateur sans biais de σ^2.

Ce résultat ne dépend pas de la loi suivie par les VA.

Si on réintroduit l'hypothèse que la suite de n variables aléatoires identiquement et indépendamment distribuées $(X_1, X_2, \ldots, X_n)$ est $\mathcal{N}(m, \sigma^2)$, cela nous permet d'obtenir les lois des estimateurs qu'on vient de voir.

$$\widehat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$$

Cela peut paraître un peu inutile puisqu'on ne peut pas l'utiliser dans la pratique mais il nous servira plus tard comme résultat théorique. Posons :

$$\frac{n}{\sigma^2}\widehat{S}^2 = \frac{n}{\sigma^2}\frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2 = \sum_{i=1}^{n}\left(\underbrace{\frac{X_i - m}{\sigma}}_{\rightsquigarrow iid\mathcal{N}(0,1)}\right)^2 \rightsquigarrow \chi^2(n)$$

car on reconnaît ici une somme de n lois normales centrées réduites indépendantes au carré. C'est donc une loi $\chi^2(n)$. Il vient :

$$\mathbb{E}\left(\frac{n}{\sigma^2}\widehat{S}^2\right) = n \Leftrightarrow \mathbb{E}\left(\widehat{S}^2\right) = \sigma^2$$

$$\mathbb{V}\left(\frac{n}{\sigma^2}\widehat{S}^2\right) = 2n \Leftrightarrow \mathbb{V}\left(\widehat{S}^2\right) = \frac{2\sigma^4}{n}$$

Loi suivie par $\widehat{S}_n^2$

$$\widehat{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

Posons :

$$\frac{n}{\sigma^2}\widehat{S}_n^2 = \frac{n}{\sigma^2}\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma}\right)^2 \rightsquigarrow \chi^2(n-1).$$

$n-1$ car dans la suite des n VA $(X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_n - \overline{X})$, on n'a seulement $n-1$ VA indépendantes pour pouvoir construire $\overline{X}$ à partir de $(X_1, X_2, \ldots, X_n)$.
On en déduit :

$$\mathbb{E}\left(\frac{n}{\sigma^2}\widehat{S}_n^2\right) = n-1 \Leftrightarrow \mathbb{E}\left(\widehat{S}_n^2\right) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$
$$\mathbb{V}\left(\frac{n}{\sigma^2}\widehat{S}_n^2\right) = 2(n-1) \Leftrightarrow \mathbb{V}\left(\widehat{S}_n^2\right) = \frac{2(n-1)}{n^2}\sigma^4$$

Par bonheur, on retrouve bien le résultat d'estimateur baisé qu'on a démontré précédemment.

Loi suivie par $\widehat{S}^2_{n-1}$

$$\widehat{S}^2_{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

Posons :

$$\frac{(n-1)}{\sigma^2}\widehat{S}^2_{n-1} = \frac{n-1}{\sigma^2}\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma}\right)^2 \rightsquigarrow \chi^2(n-1)$$

Il vient :

$$\mathbb{E}\left(\frac{(n-1)}{\sigma^2}\widehat{S}^2_{n-1}\right) = n-1 \Leftrightarrow \mathbb{E}\left(\widehat{S}^2_{n-1}\right) = \sigma^2$$
$$\mathbb{V}\left(\frac{(n-1)}{\sigma^2}\widehat{S}^2_{n-1}\right) = 2(n-1) \Leftrightarrow \mathbb{V}\left(\widehat{S}^2_{n-1}\right) = \frac{2\sigma^4}{n-1}$$

Par bonheur, on retrouve bien le résultat d'estimateur sans biais démontré précédemment.
Avec la variance de l'estimateur, on peut facilement déduire sa convergence en probabilités : $\mathbb{V}\left(\widehat{S}^2_{n-1}\right) = \dfrac{2\sigma^4}{n-1} \underset{N\to\infty}{\longrightarrow} 0 \Rightarrow \widehat{S}^2_{n-1} \xrightarrow{p} \sigma^2$.

III- Intervalles de confiance

$$X \hookrightarrow \mathcal{N}(0, 1)$$

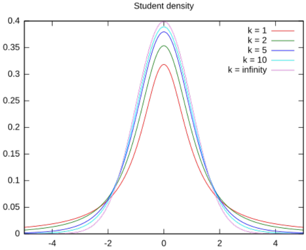
$$Z \hookrightarrow \chi^2(n)$$

X et Z indépendantes

$$\frac{X}{\sqrt{\frac{Z}{n}}} \hookrightarrow T(n)$$

Propriété intéressante :

$$T(n) \underset{n \longrightarrow \infty}{\longrightarrow} \mathcal{N}(0, 1)$$



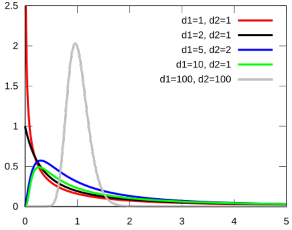
La loi de Fisher-Snedecor

$$X_1 \hookrightarrow \chi^2(\nu_1)$$

$$X_2 \hookrightarrow \chi^2(\nu_2)$$

X_1 et X_2 indépendantes

$$\frac{\frac{X_1}{\nu_1}}{\frac{X_2}{\nu_2}} \hookrightarrow F(\nu_1, \nu_2)$$



Loi de Fisher

(Valeurs de x telles que $Prob(F_{p,q} \leq x) = 95\%$)

$q \backslash p$	1	2	3	4	5	6	7	8	9	10	12
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	243,91
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2.69
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2.60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2.53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2.48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2.42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2.38
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2.34
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2.31
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2.28
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2.25
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2.23
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2.20
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2.18
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2.16
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2.09
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2.00
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1.95
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1.92
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1.88
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1.85
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1.83
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1.75

On sait construire un estimateur $\widehat{\theta}(X)$ pour les paramètres de lois θ.

On connaît leurs propriétés (sans biais, précision, convergences, efficacité).

On peut donc en déduire une estimation ponctuelle $\widehat{\theta}(x)$.

Ne pourrait-on pas fournir plutôt une fourchette ou un intervalle pour le paramètre inconnu, incluant la précision de l'estimation ?

La démarche à suivre est toujours la même, en deux temps.

- trouver une fonction de l'estimateur et du paramètre inconnu dont on connaît la loi, classique de préférence.

- utiliser ce résultat et la table de la loi pour un niveau de confiance donné pour construire l'intervalle de confiance.

$$\widehat{\theta}(X) \hookrightarrow \mathcal{N}(\theta, \mathbb{V}(\widehat{\theta}(X)))$$

$\widehat{\theta}(X)$ est l'estimateur sans biais et efficace du paramètre inconnu θ.

$\mathbb{V}(\widehat{\theta}(X))$ est la variance théorique de l'estimateur.

Qu'il soit normalement distribué (ou asymptotiquement normal), les propriétés de la loi normale nous donnent :

$$\frac{\widehat{\theta}(X) - \theta}{\sqrt{\mathbb{V}(\widehat{\theta}(X))}} \hookrightarrow \mathcal{N}(0, 1)$$

Il est en théorie possible de construire un IC pour ce paramètre inconnu à partir de ce résultat.

Définissons l'intervalle de confiance à partir de la probabilité $(1-\alpha)$ que les réalisations de cette loi normale centrée réduite appartiennent à l'intervalle $[-t_\alpha, t_\alpha]$ (inconnu pour le moment) :

$$\mathbb{P}\left(-t_\alpha < \frac{\widehat{\theta}(X)-\theta}{\sqrt{\mathbb{V}(\widehat{\theta}(X))}} < t_\alpha\right) = 1-\alpha$$

L'intervalle pour la loi de $\widehat{\theta}(X)$ est ici symétrique puisque la loi normale est symétrique autour de 0.

La valeur de t_α dépend donc de la valeur choisie pour $(1-\alpha)$, la surface de l'intervalle.

Exemple : si on désire qu'il y ait 95% de chances que les réalisations de la loi normale appartiennent à l'intervalle de confiance, on choisit dans la table de la loi normale centrée réduite la valeur de t_α associée à la surface à gauche de 97,5%, soit $t_\alpha = 1,95$. Pour 90% de chances, on choisira $t_\alpha = 1,64$.

t_α étant connu, il est possible de déduire l'intervalle de confiance du paramètre inconnu θ :

$$\Rightarrow -t_\alpha < \frac{\widehat{\theta}(x) - \theta}{\sqrt{\mathbb{V}(\widehat{\theta}(X))}} < t_\alpha$$

$$\Leftrightarrow \; -t_\alpha \sqrt{\mathbb{V}(\widehat{\theta}(X))} < \widehat{\theta}(x) - \theta < t_\alpha \sqrt{\mathbb{V}(\widehat{\theta}(X))}$$

$$\Leftrightarrow \widehat{\theta}(x) - t_\alpha \sqrt{\mathbb{V}(\widehat{\theta}(X))} < \theta < \widehat{\theta}(x) + t_\alpha \sqrt{\mathbb{V}(\widehat{\theta}(X))}$$

On peut donc en déduire l'expression de l'intervalle de confiance :

$$\theta \in \left[\widehat{\theta}(x) - t_\alpha \sqrt{\mathbb{V}(\widehat{\theta}(X))}; \widehat{\theta}(x) + t_\alpha \sqrt{\mathbb{V}(\widehat{\theta}(X))}\right]$$

au niveau de confiance de $(1 - \alpha)$.

L'IC est symétrique autour de l'estimation ponctuelle $\widehat{\theta}(x)$.

La largeur de l'intervalle dépend :

- du niveau de confiance $(1 - \alpha)$: plus $(1 - \alpha)$ est grand, plus t_α sera important et donc plus l'intervalle sera large ;

- de la précision de l'estimation ponctuelle mesurée par l'écart-type $\sqrt{\mathbb{V}(\widehat{\theta}(X))}$: plus l'estimation ponctuelle est imprécise, plus l'intervalle sera large. Une façon de réduire cette imprécision est d'augmenter la taille de l'échantillon puisque l'estimateur est convergent en probabilités (augmenter N réduit la variance de l'estimateur).

On ne connaît pas la valeur de $\mathbb{V}(\widehat{\theta}(X))$ car c'est la variance théorique.

L'approche précédente n'est donc pas applicable.

$\mathbb{V}(\widehat{\theta}(X))$ étant inconnu, on va l'estimer en utilisant un estimateur de $\widehat{\mathbb{V}(\widehat{\theta}(X))}$.

D'où l'utilité d'avoir un estimateur de variance et de connaître ses lois ! ! ! !

Car la nouvelle variable aléatoire formée à présent par :

$$\frac{\widehat{\theta}(X) - \theta}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}}$$

ne suit plus une loi normale centrée réduite !

En effet, on peut réécrire cette variable aléatoire comme :

$$\frac{\widehat{\theta}(X)-\theta}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} = \frac{\frac{\widehat{\theta}(X)-\theta}{\sqrt{\mathbb{V}(\widehat{\theta}(X))}}}{\sqrt{\frac{\widehat{\mathbb{V}(\widehat{\theta}(X))}}{\mathbb{V}(\widehat{\theta}(X))}\frac{n-1}{n-1}}} = \frac{\frac{\widehat{\theta}(X)-\theta}{\sqrt{\mathbb{V}(\widehat{\theta}(X))}}}{\sqrt{\frac{(n-1)\frac{\widehat{\mathbb{V}(\widehat{\theta}(X))}}{\mathbb{V}(\widehat{\theta}(X))}}{n-1}}} = \frac{\mathcal{N}(0,1)}{\sqrt{\frac{\chi^2(n-1)}{n-1}}} = T(n-1)$$

On voit bien qu'il s'agit du rapport entre une loi normale centrée réduite et la racine carrée d'une loi du chi2 rapportée à son nombre de degrés de liberté.

On est donc en présence d'une loi de Student à $(n-1)$ degrés de liberté si ces deux lois sont indépendantes (à admettre).

$$\mathbb{P}\left(-t_\alpha < \frac{\widehat{\theta}(X)-\theta}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} < t_\alpha\right) = 1-\alpha$$

$$\Rightarrow -t_\alpha < \frac{\widehat{\theta}(x)-\theta}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} < t_\alpha$$

$$\Leftrightarrow\ -t_\alpha\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}} < \widehat{\theta}(x)-\theta < t_\alpha\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}$$

$$\Leftrightarrow \widehat{\theta}(x) - t_\alpha\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}} < \theta < \widehat{\theta}(x) + t_\alpha\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}$$

On peut donc en déduire l'intervalle de confiance :

$$\theta \in \left[\widehat{\theta}(x) - t_\alpha\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}};\widehat{\theta}(x) + t_\alpha\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}\right]$$

au niveau de confiance de $(1-\alpha)$.

L'IC est symétrique autour de l'estimation ponctuelle $\widehat{\theta}(x)$ puisque la loi de Student est symétrique.

La largeur de l'intervalle 2 $t_\alpha \sqrt{\widehat{\mathbb{V}(\widehat{\theta}(x))}}$ dépend :

- du niveau de confiance fixé $(1-\alpha)$: plus $(1-\alpha)$ est grand, plus t_α sera important et donc plus l'intervalle sera large ;

- de la précision de l'estimation ponctuelle mesurée par la variance estimée $\widehat{\mathbb{V}(\widehat{\theta}(x))}$: plus l'estimation ponctuelle est imprécise, plus l'intervalle sera large. Augmenter la taille de l'échantillon peut encore une fois permettre de corriger ce point.

- La loi de Student traduit une incertitude plus grande que la loi Normale car la variance de l'estimateur est estimée alors que dans le cas normal, on faisait comme si elle était connue. Dès lors, les valeurs critiques de la loi de Student sont plus grandes que celles de la loi Normale.

Exemple : si on désire qu'il y ait 95% de chances que les réalisations d'une loi de Student à 25 degrés de liberté appartiennent à l'intervalle de confiance, on choisit dans la table de la loi de Student la valeur de t_α associée à la surface à gauche de 97,5% ou de 2,5% à droite, soit $t_\alpha = 2,06$.

La table de Student

LOI de STUDENT

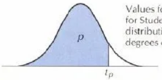


TABLE A.3 Values of t for Various Probability Levels, p

ν	$t_{.60}$	$t_{.75}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.9975}$	$t_{.999}$	$t_{.9995}$
1	0.325	1.000	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.289	0.816	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.277	0.765	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.271	0.741	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	0.256	0.685	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.255	0.679	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.254	0.679	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
70	0.254	0.678	1.044	1.294	1.667	1.994	2.381	2.648	2.899	3.211	3.435
80	0.254	0.678	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
90	0.254	0.677	1.042	1.291	1.662	1.987	2.368	2.632	2.878	3.183	3.402
100	0.254	0.677	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
110	0.254	0.677	1.041	1.289	1.659	1.982	2.361	2.621	2.865	3.166	3.381
120	0.254	0.677	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Soit une suite de n variables aléatoires iid $(X_1, X_2, \ldots, X_n)$, d'espérance m et de variance σ^2.

Un bon estimateur de m :

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

On a montré qu'il avait toutes les propriétés + avec la normalité : l'efficacité et la normalité : $\Leftrightarrow \ \overline{X} \rightsquigarrow \mathcal{N}(m, \frac{\sigma^2}{n})$.

Un bon estimateur de σ^2 :

$$\widehat{S}^2_{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

On connaît sa loi et on sait qu'il est sans biais et convergent.

$$\frac{(n-1)}{\sigma^2}\widehat{S}^2_{n-1} \rightsquigarrow \chi^2(n-1)$$

$$\mathbb{E}\left(\widehat{S}^2_{n-1}\right) = \sigma^2$$

$$\mathbb{V}\left(\widehat{S}^2_{n-1}\right) = \frac{2\sigma^4}{n-1}$$

Comme

$$\overline{X} \rightsquigarrow \mathcal{N}(m, \frac{\sigma^2}{n}) \Leftrightarrow \frac{\overline{X} - \mathbb{E}(\overline{X})}{\sqrt{\mathbb{V}(\overline{X})}} = \frac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}} \hookrightarrow \mathcal{N}(0, 1)$$

mais σ^2 étant inconnu, cela ne sert à rien en pratique.

On va l'estimer avec

$$\frac{(n-1)\widehat{S}^2_{n-1}}{\sigma^2} \hookrightarrow \chi^2(n-1)$$

On peut donc poser :

$$\frac{\overline{X} - m}{\frac{\widehat{S}_{n-1}}{\sqrt{n}}} = \frac{\sqrt{n}(\overline{X} - m)}{\widehat{S}_{n-1}} = \frac{\frac{\sqrt{n}(\overline{X}-m)}{\sigma}}{\sqrt{\frac{n-1}{n-1}\frac{\widehat{S}^2_{n-1}}{\sigma^2}}} = \frac{\mathcal{N}(0,1)}{\sqrt{\frac{\chi^2(n-1)}{n-1}}} \hookrightarrow T(n-1)$$

Après avoir trouvé la valeur de t_α dans la table de Student à $(n-1)$ degrés de liberté pour le niveau de confiance $1-\alpha$, l'IC s'écrit :

$$m \in \left[\overline{x} - t_\alpha \frac{\widehat{s}_{n-1}}{\sqrt{n}} ; \overline{x} + t_\alpha \frac{\widehat{s}_{n-1}}{\sqrt{n}}\right]$$

Utilisons le résultat :

$$\frac{(n-1)\widehat{S}^2_{n-1}}{\sigma^2} \hookrightarrow \chi^2(n-1)$$

Définissons l'IC comme la probabilité $(1-\alpha)$ que les réalisations appartiennent à $[\chi^2_{inf}; \chi^2_{sup}]$:

$$\mathbb{P}\left(\chi^2_{inf} < \frac{(n-1)\widehat{S}^2_{n-1}}{\sigma^2} < \chi^2_{sup}\right) = 1-\alpha$$

La loi du χ^2 n'étant pas symétrique, il n'y a aucune relation entre les valeurs formant l'intervalle de confiance. La surface à l'extérieur de l'intervalle est égale à α, supposée répartie en deux surfaces égales $\frac{\alpha}{2}$.

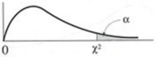
χ^2_{inf} dans la table du χ^2 à $(n-1)$ degrés de liberté : surface à sa gauche égale à $\frac{\alpha}{2}$ ou surface à sa droite égale à $1-\frac{\alpha}{2}$.

χ^2_{sup} dans la table du χ^2 à $(n-1)$ degrés de liberté : surface à sa gauche égale à $1-\frac{\alpha}{2}$ ou surface à sa droite égale à $\frac{\alpha}{2}$.

Exemple : pour une loi du chi à 30 degrés de liberté, l'intervalle de confiance à 90% est $[18,49; 43,77]$.

La table du chi2

Table χ^2 : points de pourcentage supérieurs de la distribution χ^2



dl	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.82	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.54	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.66	23.59
10	2.15	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.75
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.21	28.30
13	3.56	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.69	26.12	29.14	31.31
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.15
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.56	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.93	36.78	40.29	42.80
23	9.26	10.19	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.88	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.37	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.32	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.80	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.20	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.78	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.67	22.14	24.42	26.51	29.06	33.67	39.34	45.61	51.80	55.75	59.34	63.71	66.80
50	27.96	29.68	32.35	34.76	37.69	42.95	49.34	56.33	63.16	67.50	71.42	76.17	79.52
60	35.50	37.46	40.47	43.19	46.46	52.30	59.34	66.98	74.39	79.08	83.30	88.40	91.98
70	43.25	45.42	48.75	51.74	55.33	61.70	69.34	77.57	85.52	90.53	95.03	100.44	104.24
80	51.14	53.52	57.15	60.39	64.28	71.15	79.34	88.13	96.57	101.88	106.63	112.34	116.35
90	59.17	61.74	65.64	69.13	73.29	80.63	89.33	98.65	107.56	113.14	118.14	124.13	128.32
100	67.30	70.05	74.22	77.93	82.36	90.14	99.33	109.14	118.49	124.34	129.56	135.82	140.19

On en déduit :

$$\Rightarrow \chi^2_{inf} < \frac{(n-1)\widehat{s}^2_{n-1}}{\sigma^2} < \chi^2_{sup}$$
$$\Leftrightarrow \frac{1}{\chi^2_{sup}} < \frac{\sigma^2}{(n-1)\widehat{s}^2_{n-1}} < \frac{1}{\chi^2_{inf}}$$
$$\Leftrightarrow \frac{(n-1)\widehat{s}^2_{n-1}}{\chi^2_{sup}} < \sigma^2 < \frac{(n-1)\widehat{s}^2_{n-1}}{\chi^2_{inf}}$$

On peut donc en déduire l'intervalle de confiance :

$$\sigma^2 \in \left[\frac{(n-1)\widehat{s}^2_{n-1}}{\chi^2_{sup}} ; \frac{(n-1)\widehat{s}^2_{n-1}}{\chi^2_{inf}}\right]$$

au niveau de confiance $(1-\alpha)$.

Ici, IC non symétrique autour de l'estimation ponctuelle du paramètre inconnu. Comme précédemment, la largeur de l'intervalle dépend du niveau de confiance choisi.

IV- Les tests

Le ou les paramètres inconnus sont-ils statistiquement égaux ou non à telle valeur supposée ?

La théorie des tests le permet, en se ramenant uniquement au choix entre deux hypothèses antagonistes, notées H_0 et H_a (ou H_1). L'hypothèse H_0, encore appelée hypothèse nulle, est privilégiée jusqu'au moment où elle est infirmée par l'observation.

Ainsi, le test a pour but de mesurer l'adéquation d'une hypothèse à la réalité observée à travers l'information apportée par l'échantillon.

On peut voir cela comme une distance entre l'hypothèse et l'observation

On retient plusieurs étapes dans la démarche des tests.

Il est d'abord nécessaire de formuler les hypothèses à tester, et par voie de conséquence, les erreurs de décision associées à ces hypothèses.

Supposons que l'on fasse un test sur le paramètre inconnu θ. On distinguera différents types de tests sur un paramètre.

Dans le cas des **tests unilatères**, on teste l'égalité du paramètre inconnu θ à la valeur θ^{H_0} supposée dans l'hypothèse nulle contre la stricte supériorité (respectivement infériorité) à cette même valeur sous l'hypothèse alternative :

$$H_0 : \ \theta = \theta^{H_0}$$

$$H_a : \ \theta > \theta^{H_0} \ \text{ ou } \ H_a : \ \theta < \theta^{H_0}$$

Dans le cas des **tests bilatères**, on teste l'égalité du paramètre inconnu θ à la valeur θ^{H_0} supposée dans l'hypothèse nulle contre la différence (c'est-à-dire la stricte supériorité ou infériorité) à cette même valeur sous l'hypothèse alternative :

$$H_0 : \ \theta = \theta^{H_0}$$
$$H_a : \ \theta \neq \theta^{H_0}$$

Deux actions possibles et donc deux possibilités de se tromper.

	H_0 vraie	H_0 fausse
Accepter H_0	-	erreur 2de espèce
Rejeter H_0	erreur 1ère espèce	-

Deux types de risques d’erreur :

- **le risque d’erreur de première espèce** : risque de refuser l’hypothèse nulle alors qu’elle est vraie. Sa probabilité est $\boldsymbol{\alpha} = \mathbb{P}(\overline{\boldsymbol{H_0}}||\boldsymbol{H_0})$.

- Le risque d’erreur de seconde espèce : risque d’accepter l’hypothèse nulle alors qu’elle est fausse. Sa probabilité est $\beta = \mathbb{P}(H_0||\overline{H_0})$.

avec l’opérateur $\mathbb{P}(.||.)$ ayant pour premier argument **une décision** et pour second **un état de la nature** (inobservable).

$\boldsymbol{\alpha} = \mathbb{P}(\overline{\boldsymbol{H_0}}||\boldsymbol{H_0})$ est donc la probabilité de décider de ne pas accepter H_0 alors que H_0 est vraie.

Conditionnement **par rapport à un état de la nature (H_0 vraie ou non) que l'on n'observe pas** et non un événement. Ce n'est donc pas une probabilité conditionnelle.

On accepte ou on rejette l'hypothèse nulle et rien d'autre ! L'hypothèse alternative ne permet que de définir la zone de rejet.

Trouver une statistique, permettant de mesurer l'adéquation entre l'hypothèse formulée et ce que disent les données pour répondre à la question posée.

- **une fonction discriminante (ou pivotale)**
- construite sous l'hypothèse nulle,
- une "distance" entre l'hypothèse et les données,
- à comparer à une valeur théorique calculée à partir de la loi de probabilités et définie à partir de la zone de rejet de l'hypothèse nulle construite à partir de l'hypothèse alternative pour une probabilité de risque de première espèce α donnée (Neymann-Pearson).

La **règle de décision** qui en découle permet ensuite de conclure.

$$H_0 : \ \theta = \theta^{H_0}$$
$$H_a : \ \theta \neq \theta^{H_0}$$

soit encore $H_0 : \ \theta - \theta^{H_0} = 0$.

Du point de vue de l'estimation, mesurée par la distance $\widehat{\theta}(x) - \theta^{H_0} \lessgtr 0$.

Idéalement la distance $\widehat{\theta}(x) - \theta^{H_0}$ devrait être 0 si les données confirment H_0.

Mais il faut tenir compte de l'incertitude donc pas forcément 0 mais assez petit pour confirmer.

On va utiliser une statistique fondée sur $\widehat{\theta}(X) - \theta^{H_0}$ dont on connaît la loi.

Sous H_0, l'espérance de cette loi doit être 0 puisque $\mathbb{E}(\widehat{\theta}(X)) = \theta = \theta^{H_0}$.

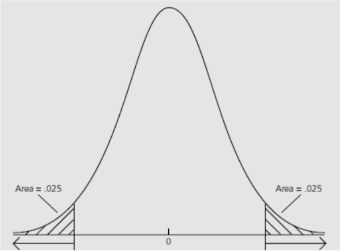


Figure – Zone de rejet test bilatère pour $\alpha = 5\%$

Ici, la zone d'acceptation va se situer autour de 0 (sous H_0, $\widehat{\theta}(X)$ est centrée sur θ^{H_0}), jusqu'à deux valeurs situées de part et d'autre de 0 (du fait de H_a).

- si $\widehat{\theta}(x) - \theta^{H_0}$ tombe dans cette zone, on peut statistiquement accepter $H_0 : \ \theta = \theta^{H_0}$.

- Sinon, au-delà de cette marge (d'un coté ou de l'autre), H_0 n'est plus statistiquement acceptable.

En appliquant la définition de la probabilité du risque de première espèce, et en adaptant la règle de rejet de l'hypothèse nulle :

$$\alpha = \mathbb{P}(\overline{H_0} \mid H_0) = \mathbb{P}\left(\frac{\widehat{\theta}(X) - \theta}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} \notin \underbrace{[-t_\alpha; t_\alpha]}_{ZA} \parallel \theta = \theta^{H_0}\right)$$

En utilisant l'événement contraire :

$$1 - \alpha = \mathbb{P}\left(\frac{\widehat{\theta}(X) - \theta}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} \in [-t_\alpha; t_\alpha] \parallel \theta = \theta^{H_0}\right)$$

$$= \mathbb{P}\left(-t_\alpha \leqslant \underbrace{\frac{\widehat{\theta}(X) - \theta^{H_0}}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}}}_{\rightsquigarrow T(n-1)} \leqslant t_\alpha\right)$$

Il ne reste alors plus qu'à trouver la valeur de t_α associée à la probabilité centrale $(1 - \alpha)$ dans la table de Student $T(n - 1)$.

Une fois trouvée t_α dans la table, on calcule $\dfrac{\widehat{\theta}(x) - \theta^{H_0}}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}}$:

- si $\dfrac{\widehat{\theta}(x) - \theta^{H_0}}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} \in [-t_\alpha ; t_\alpha]$, on est dans la zone d'acceptation du test et on accepte H_0 ;

- si $\dfrac{\widehat{\theta}(x) - \theta^{H_0}}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} \notin [-t_\alpha ; t_\alpha]$, on est dans la zone de rejet du test et on ne peut pas accepter H_0.

On peut aussi réinterpréter les choses avec les intervalles de confiance.

$$\frac{\widehat{\theta}(x)-\theta^{H_0}}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(x))}}} \in [-t_\alpha; t_\alpha] \Leftrightarrow \theta^{H_0} \in \left[\widehat{\theta}(x) - t_\alpha\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(x))}}; \widehat{\theta}(x) + t_\alpha\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(x))}}\right]$$

on est dans la zone d'acceptation du test ou la valeur supposée θ^{H_0} est dans l'IC, on accepte H_0.

$$\frac{\widehat{\theta}(x)-\theta^{H_0}}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} \notin [-t_\alpha; t_\alpha] \Leftrightarrow \theta^{H_0} \notin \left[\widehat{\theta}(x) - t_\alpha\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}; \widehat{\theta}(x) + t_\alpha\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}\right]$$

on est dans la zone de rejet du test ou la valeur supposée θ^{H_0} n'est pas dans l'IC, et on n'accepte pas H_0.

Attention : ne fonctionne que pour les tests bilatéraux au risque α et un IC au niveau de confiance $1-\alpha$.

$$H_0 : \ \theta = \theta^{H_0}$$
$$H_a : \ \theta > \theta^{H_0}$$

on utilisera comme fonction discriminante la loi suivie par l'estimateur de ce paramètre en se plaçant sous l'hypothèse nulle. Ainsi :

$$\frac{\widehat{\theta}(X) - \theta}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} \hookrightarrow T(n-1)$$

et en se plaçant sous H_0 vraie (c'est-à-dire $\theta = \theta^{H_0}$) :

$$\frac{\widehat{\theta}(X) - \theta^{H_0}}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} \hookrightarrow T(n-1)$$

L'hypothèse alternative permet de définir la zone critique, c'est à dire la zone de rejet de l'hypothèse nulle connaissant la probabilité du risque de première espèce α.

On raisonne de la manière suivante : du fait de l'incertitude due au modèle, à l'échantillonnage,. . ., on ne pourra probablement pas observer exactement l'égalité à la valeur θ^{H_0} même si c'est bien le cas. On va donc se laisser une certaine marge pour accepter H_0.

Posons que cette marge d'acceptation va jusqu'à une valeur critique t_α, inconnue mais supérieure à la fonction pivotale (du fait de l'hypothèse alternative). L'action d'accepter ou de rejeter H_0 va donc se traduire dans la position observée de la statistique relativement à cette valeur critique :

- à l'intérieur de cette marge, on peut statistiquement accepter l'hypothèse d'égalité du paramètre inconnu θ à la valeur θ^{H_0} supposée dans H_0.
- Au delà de cette marge, H_0 n'est plus statistiquement acceptable.

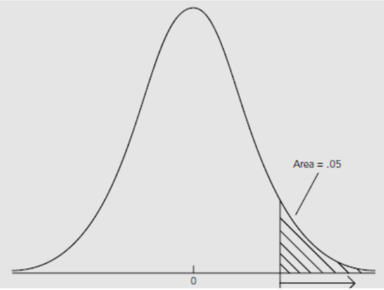


Figure – Zone de rejet test unilatère droit pour $\alpha = 5\%$

C’est ainsi que l’on va confronter notre hypothèse aux données.

La probabilité du risque de première espèce, à savoir rejeter l’hypothèse nulle alors qu’elle est vraie, se traduit par la probabilité que la fonction pivotale tombe au delà de la marge acceptable avec la probabilité α :

$$\begin{aligned}
\alpha &= \mathbb{P}(\overline{H_0} \mid H_0) \\
&= \mathbb{P}\left(\frac{\widehat{\theta}(X) - \theta}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} \geqslant t_\alpha \parallel \theta = \theta^{H_0}\right) \\
&= \mathbb{P}\left(\frac{\widehat{\theta}(X) - \theta^{H_0}}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} \geqslant t_\alpha\right)
\end{aligned}$$

où t_α est la valeur critique dans la table de Student pour la probabilité α au delà de laquelle il n’est plus tenable de défendre l’hypothèse nulle et où on doit donc la rejeter.

Une fois trouvée t_α dans la table, on calcule $\dfrac{\widehat{\theta}(x) - \theta^{H_0}}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}}$:

- si $\dfrac{\widehat{\theta}(x) - \theta^{H_0}}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} < t_\alpha$, on est dans la zone d'acceptation du test et on accepte H_0 ;

- si $\dfrac{\widehat{\theta}(x) - \theta^{H_0}}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} > t_\alpha$, on est dans la zone de rejet du test et on ne peut pas accepter H_0.

Prenons l'autre test unilatère sur le paramètre inconnu θ :

$$H_0 : \ \theta = \theta^{H_0}$$
$$H_a : \ \theta < \theta^{H_0}$$

La zone d'acceptation va jusqu'à une valeur critique $-t_\alpha$ inférieure à 0 (du fait de l'hypothèse alternative). L'action d'accepter ou de rejeter l'hypothèse nulle va donc se traduire dans la position observée de la fonction pivotale relativement à cette valeur critique :

- à l'intérieur de cette marge, on peut statistiquement accepter l'hypothèse d'égalité du paramètre inconnu θ à la valeur θ^{H_0} supposée dans H_0.
- En deçà de cette marge, H_0 n'est plus statistiquement acceptable.

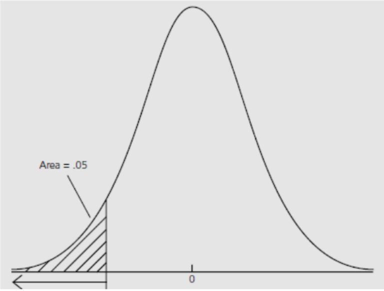


Figure – Zone de rejet test unilatère gauche pour $\alpha = 5\%$

Comme précédemment, calculons la probabilité du risque de première espèce :

$$\begin{aligned}
\alpha &= \mathbb{P}(\overline{H_0} \mid H_0) \\
&= \mathbb{P}\left(\frac{\widehat{\theta}(X) - \theta}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} \leqslant -t_\alpha \parallel \theta = \theta^{H_0}\right) \\
&= \mathbb{P}\left(\frac{\widehat{\theta}(X) - \theta^{H_0}}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} \leqslant -t_\alpha\right)
\end{aligned}$$

où $-t_\alpha$ est la valeur critique de la table de Student pour la probabilité α en deçà de laquelle il n'est plus tenable de défendre l'hypothèse nulle et où on doit donc la rejeter.

Attention, ici il faudra utiliser les propriétés de symétrie de la loi de Student pour trouver la valeur de $-t_\alpha$ qui sera ici **négatif**.

Une fois trouvée t_α dans la table, on calcule $\dfrac{\widehat{\theta}(x) - \theta^{H_0}}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}}$:

- si $\dfrac{\widehat{\theta}(x) - \theta^{H_0}}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} > -t_\alpha$, on est dans la zone d'acceptation du test et on accepte H_0 ;

- si $\dfrac{\widehat{\theta}(x) - \theta^{H_0}}{\sqrt{\widehat{\mathbb{V}(\widehat{\theta}(X))}}} < -t_\alpha$, on est dans la zone de rejet du test et on ne peut pas accepter H_0.

- Statistique et Probabilités, J.P. Lecoutre, Dunod.
- Statistique et Probabilités, C. Hurlin et V. Mignon, Dunod.
- Probabilités et Statistiques, A. Combrouze, Presses Universitaires de France.