Probabilités & statistiques

Consignes :

- Vous disposez de **2 h** pour répondre aux **4 × 3** questions suivantes.
- **Calculatrice** non programmable **autorisée**, voire conseillée.
- Un **aide-mémoire** comportant certains faits à (ne pas) connaître est fourni en annexe.
- Soyez **précis** et **détaillé** dans vos réponses et justifications ; **exprimez**-vous !

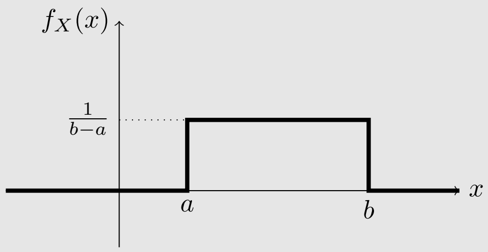
Exercice 1

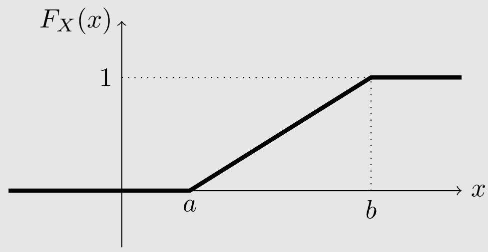
Soit X une variable aléatoire uniformément distribuée dans l'intervalle $[a, b]$.

a) Décrire et représenter la densité de probabilité f_X et la fonction de répartition F_X de X.

$$f_X(x) = \frac{1}{b-a}\Pi_{b-a}(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } a \leqslant x \leqslant b \\ 0 & \text{sinon} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{si } x < a \\ \dfrac{x-a}{b-a} & \text{si } a \leqslant x \leqslant b \\ 1 & \text{sinon} \end{cases}$$





b) Par calcul direct à partir de f_X et F_X, déterminer l'espérance, la variance et la médiane de X.

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x\, f_X(x)\,\mathrm{d}x = \int_a^b \frac{x}{b-a}\,\mathrm{d}x = \left[\frac{x^2}{2(b-a)}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

$$\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2\, f_X(x)\,\mathrm{d}x = \int_a^b \frac{x^2}{b-a}\,\mathrm{d}x = \left[\frac{x^3}{3(b-a)}\right]_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{4a^2+4ab+4b^2-3a^2-6ab-3b^2}{12} = \frac{(b-a)^2}{12}$$

Médiane m :

$$F_X(m) = \frac{m-a}{b-a} = \frac{1}{2} \implies 2(m-a) = b-a \implies m = \frac{a+b}{2}$$

c) Déterminer la fonction génératrice des moments g_X et l'utiliser pour retrouver l'espérance et la variance.

$$g_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{+\infty} e^{tx}\, f_X(x)\,\mathrm{d}x = \int_a^b \frac{e^{tx}}{b-a}\,\mathrm{d}x = \left[\frac{e^{tx}}{t(b-a)}\right]_a^b = \frac{e^{bt}-e^{at}}{t(b-a)}$$

En développant à l'ordre 2 :

$$\begin{aligned} g_X(t) &= \frac{1}{t(b-a)}\left(1 + bt + \frac{b^2t^2}{2} + \frac{b^3t^3}{6} - 1 - at - \frac{a^2t^2}{2} - \frac{a^3t^3}{6}\right) + \cdots \\ &= 1 + \frac{a+b}{2}\,t + \frac{a^2+ab+b^2}{6}\,t^2 + \cdots \end{aligned}$$

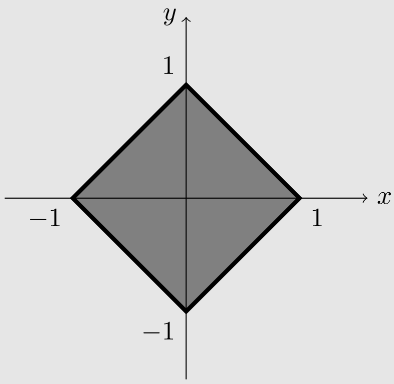
On retrouve donc $\mu = \dfrac{a+b}{2}$ et $\mu_2 = \dfrac{a^2+ab+b^2}{3}$ comme plus haut et donc $\mathrm{Var}(X) = \dfrac{(b-a)^2}{12}$.

Exercice 2

Un couple aléatoire (X, Y) est uniformément distribué dans le carré $\mathcal{C} \subset \mathbb{R}^2$ de sommets $(\pm 1, 0)$, $(0, \pm 1)$.

a) Explicitez la densité conjointe $f(x, y)$ du couple ainsi que les densités marginales $f_X(x)$ et $f_Y(y)$.

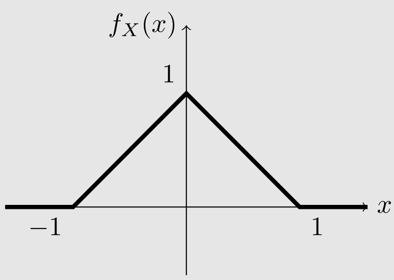
On a $f(x,y) = \dfrac{1}{\mathrm{aire}(\mathcal{C})} = \dfrac{1}{2}$ pour $(x,y) \in \mathcal{C}$, c'est-à-dire $|x \pm y \leqslant 1$; et $f(x,y) = 0$ sinon.

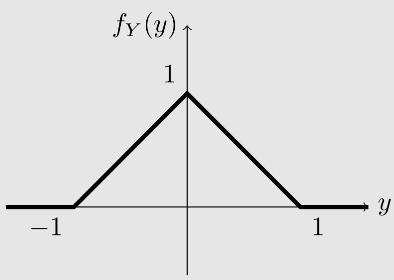


Pour les densités marginales, on intègre dans chaque direction :

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y = \int_{-1+|x|}^{1-|x|} \frac{1}{2}\,\mathrm{d}y = 1 - |x| \qquad \text{sur } [-1, 1]$$

et de même pour Y.





b) Déterminer $\mathrm{Cov}(X, Y)$. Les variables X et Y sont-elles indépendantes ?

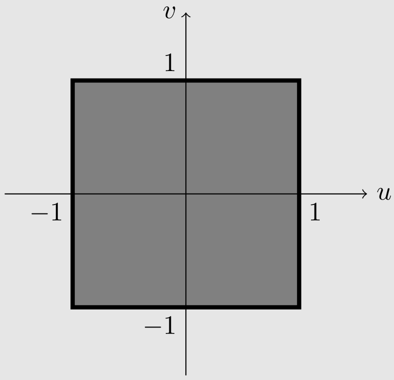
$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0 - 0 \cdot 0 = 0$$

Par contre X et Y ne sont pas indépendantes puisqu'on voit bien que $f(x,y) \neq f_X(x)\, f_Y(y)$. Ou alors on peut vérifier par exemple que

$$\underbrace{\mathbb{P}[|X|, |Y| \leqslant \tfrac{1}{2}]}_{\frac{1}{2}} \neq \underbrace{\mathbb{P}[|X| \leqslant \tfrac{1}{2}]}_{\frac{3}{4}} \cdot \underbrace{\mathbb{P}[|Y| \leqslant \tfrac{1}{2}]}_{\frac{3}{4}}.$$

c) Mêmes questions avec $U := X + Y$ et $V := X - Y$.

Dans le plan (u,v), le couple est supporté sur le carré $[-1,1]^2$:



et la densité conjointe est

$$\frac{1}{4}\, \Pi_2(u)\, \Pi_2(v) = \frac{1}{2}\, \Pi_2(u) \cdot \frac{1}{2}\, \Pi_2(v).$$

On a toujours $\mathrm{Cov}(U,V) = 0$ mais cette fois les variables sont indépendantes.

Exercice 3

Modélisons la durée d'un match de tennis (en heures) par $X = X_1 + ... + X_k$ où chaque X_i est la durée d'un échange, dont la fonction génératrice des moments est de la forme $g_i(t) = (1 - \alpha t)^{-2}$ avec $\alpha > 0$.

a) Montrer que $\mathbb{E}[X_i] = 2\alpha$ et $\mathrm{Var}(X_i) = 2\alpha^2$.

En développant à l'ordre 2, on trouve

$$g_i(t) = (1 - \alpha t)^{-2} = (1 + \alpha t + \alpha^2 t^2 + \cdots)^2 = 1 + 2\alpha t + 3\alpha^2 t^2 + \cdots$$

d'où $\mu = 2\alpha$ et $\mu_2 = 6\alpha^2$, on en tire $\mathrm{Var}(X) = \mu_2 - \mu^2 = 6\alpha^2 - 4\alpha^2 = 2\alpha^2$.

b) En précisant les hypothèses effectuées, obtenir la fonction génératrice des moments de X et donner une approximation normale de cette variable aléatoire lorsque k est grand.

Si on suppose les X_i indépendantes (ce qui est raisonnable), on peut utiliser le fait que la fonction génératrice d'une somme est les produit des fonctions génératrices :

$$g_X(t) = g_1(t) \cdots g_k(t) = (1 - \alpha t)^{-2k}.$$

Il s'agit d'une variable aléatoire d'espérance $2k\alpha$ et de variance $2k\alpha^2$, somme de variables i.i.d. et donc approximativement normale d'après le théorème central limite. On peut donc l'approximer par une variable

$$Y \sim \mathcal{N}(2k\alpha,\, 2k\alpha^2).$$

c) En prenant $k = 18$, $\alpha = \frac{1}{18}$, combien de temps avant la fin des transports en commun doit-on démarrer le match pour être sûr à 95 % que le public ne rate pas le dernier métro ?

Avec les valeurs fournies, la durée X du match suit approximativement une loi $\mathcal{N}(2, \frac{1}{9})$. On cherche la valeur d pour laquelle

$$0{,}95 = \mathbb{P}[X \leqslant d] = \mathbb{P}[Z \leqslant 3(d-2)]$$

où $Z := 3(X-2) \sim \mathcal{N}(0,1)$. En consultant la fonction de répartion de la loi normale centrée réduite fournie en annexe, on voit qu'on doit donc prévoir environ

$$3(d-2) \approx 1{,}65 \implies d \approx 2 + \frac{1{,}65}{3} = 2{,}55 \text{ heures}$$

soit 2 h 33 pour le match afin d'être sûr à 95 % de ne pas dépasser la durée prévue.

Exercice 4

On dispose de 2 machines identiques fonctionnant indépendamment et pouvant tomber en panne au cours d'une journée avec la probabilité p.

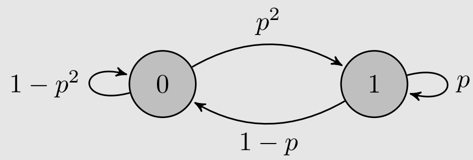
a) On suppose que, si une machine tombe en panne un jour, le réparateur vient s'en occuper pendant la nuit (mais qu'il n'a le temps de réparer qu'une seule machine par nuit).

Modéliser le nombre de machines en panne au début de la n-ième journée à l'aide d'une chaîne de Markov à deux états dont vous préciserez les probabilités de transitions.

Si un matin les deux machines fonctionnent, il y a probabilité $(1-p)^2$ qu'aucune machine ne tombe en panne, et $2p(1-p)$ que l'une ou l'autre tombe en panne pendant la journée ; dans ces deux cas, les deux machines seront fonctionnelles le lendemain matin. C'est seulement lorsque les deux machines tombent en panne le même jour, ce qui se produit avec probabilité p^2, que le réparateur repart le matin en laissant une machine toujours en panne.

De même, si le matin une machine est en panne : si la seconde a une défaillance pendant la journée (probabilité p), le réparateur ne pourra en réparer qu'une pendant la nuit. Par contre, si la seconde machine lui laisse un peu de répit (probabilité $1-p$), les deux machines seront de nouveau fonctionnelles le lendemain matin.

Sous forme de chaîne de Markov, chaque état correspondant au nombre de machines défectueuses au début de la journée, cela donne donc :



b) En déterminant l'état stationnaire de cette chaîne, calculer l'espérance du nombre de machines en panne en début de journée en régime établi.

Soit X le nombre de machines défectueuses le matin en régime établi, avec probabilités respectives p_0 et p_1. Le vecteur d'état est alors un point fixe du système :

$$\begin{bmatrix} 1-p^2 & 1-p \\ p^2 & p \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \end{bmatrix} = \begin{bmatrix} p_0 \\ p_1 \end{bmatrix} \qquad \Longleftrightarrow \qquad \begin{bmatrix} -p^2 & 1-p \\ p^2 & p-1 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

donc une solution de l'équation

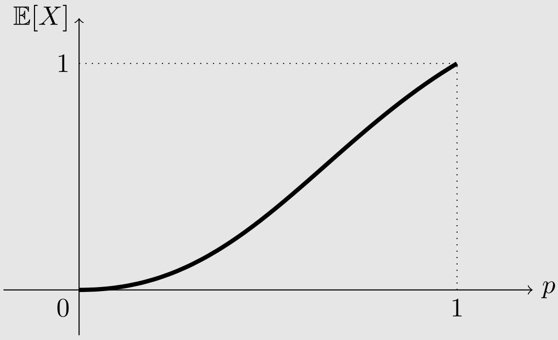
$$p^2\, p_0 = (1-p)\, p_1.$$

En ajoutant la normalisation $p_0 + p_1 = 1$, on trouve comme unique solution

$$p_0 = \frac{1-p}{p^2 + (1-p)}, \qquad p_1 = \frac{p^2}{p^2 + (1-p)}.$$

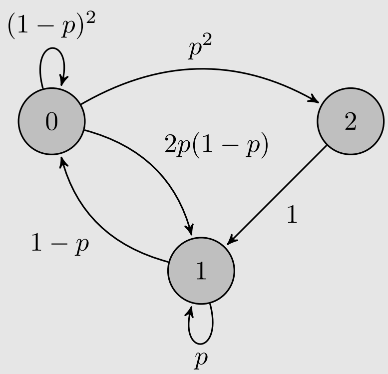
On a donc

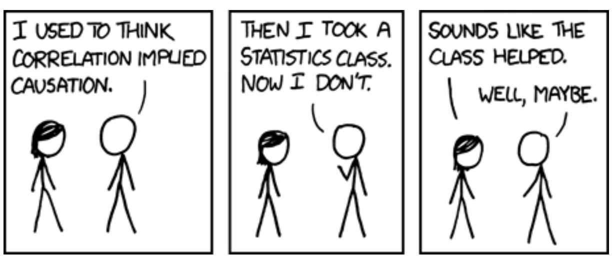
$$\mathbb{E}[X] = 0 \cdot p_0 + 1 \cdot p_1 = \frac{p^2}{p^2 + (1-p)}.$$



c) Reprendre la question a) en supposant qu'une machine tombant en panne une journée n'est réparée que la nuit du lendemain, le réparateur ne pouvant toujours réparer qu'une seule machine par nuit.

Cette fois on doit utiliser une machine à 3 états :





Statistiques

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, \mathrm{d}x, \qquad \mathrm{Var}(X) = \sigma_X^2 = \mathrm{Cov}(X, X)$$

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\, \mathbb{E}[Y], \qquad \mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X\, \sigma_Y}$$

Fonction génératrice

$$g_X(t) = \widehat{f_X}\left(-\frac{t}{2\pi i} \right) = \mathbb{E}[e^{\,tX}] = \int_{-\infty}^{+\infty} f_X(x)\, e^{tx}\, \mathrm{d}x = \sum_{n=0}^{\infty} \frac{\mu_n}{n!} t^n = 1 + \mu_1\, t + \frac{\mu_2}{2} t^2 + \frac{\mu_3}{6} t^3 + \cdots$$

avec $\mu_n = \mathbb{E}[X^n]$ les moments de la variable aléatoire X et f_X sa densité

Quelques lois fréquemment utilisées

- loi binomiale $\mathcal{B}(n, p)$:

$$\mathbb{P}[X = k] \;=\; \binom{n}{k} p^k (1-p)^{n-k} \quad (k = 0, 1, \ldots, n)$$

espérance $\quad np, \qquad$ variance $\quad np(1-p)$

fonction génératrice $\quad (1 - p + pe^t)^n$

- loi géométrique $\mathcal{G}(p)$:

$$\mathbb{P}[X = k] \;=\; (1-p)^{k-1} p \quad (k \geqslant 1)$$

espérance $\quad \dfrac{1}{p}, \qquad$ variance $\quad \dfrac{1-p}{p^2}$

fonction génératrice $\quad \dfrac{p}{e^{-t} + p - 1}$

- loi de Poisson $\mathcal{P}(\lambda)$:

$$\mathbb{P}[X=k] = \frac{\lambda^k}{k!}e^{-\lambda} \quad k \geqslant 0$$

$$\text{espérance} \quad \lambda, \qquad \text{variance} \quad \lambda$$

$$\text{fonction génératrice} \quad e^{\lambda(e^t-1)}$$

- loi uniforme $\mathcal{U}([0,1])$:

$$\text{densité} \quad f(x) = 1 \quad (0 \leqslant x \leqslant 1)$$

$$\text{espérance} \quad \frac{1}{2}, \qquad \text{variance} \quad \frac{1}{12}$$

$$\text{fonction génératrice} \quad \frac{e^t-1}{t}$$

- loi exponentielle $\mathcal{E}(\lambda)$:

$$\text{densité} \qquad f(x) = \lambda e^{-\lambda x} \quad (x \geqslant 0)$$

$$\text{espérance} \quad \frac{1}{\lambda}, \qquad \text{variance} \quad \frac{1}{\lambda^2}$$

$$\text{fonction génératrice} \quad \frac{\lambda}{\lambda - t}$$

- loi normale $\mathcal{N}(\mu, \sigma^2)$:

$$\text{densité} \quad f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\text{espérance} \quad \mu, \qquad \text{variance} \quad \sigma^2$$

$$\text{fonction génératrice} \quad e^{\mu t}\, e^{\frac{1}{2}\sigma^2 t^2}$$

Fonction de répartition d'une loi normale centrée réduite