Correction partiel de Probabilité

Colin 'Alaska' Coërchon

17 janvier 2023

1 Commentaires persos

*Après ajustement, j'ai eu la note de **17/20** à cette épreuve. Pour vous donner un ordre d'idée, j'avais tout fait sauf la dernière question de chaque exercice.*

***Il remet aux partiels quasiment les mêmes exercices depuis plus de 10 ans**. De toute évidence, je vous conseille donc d'en faire le maximum en regardant les partiels des années précédentes :)*

Pour toute question sur la correction (ou sur le ++), n'hésitez pas à me la poser sur Discord (`@colin_alaska`). Avec ~~beaucoup~~ un peu de chance, je vous répondrais !

*Pour ceux qui pensent continuer **en filière MA** l'an prochain, et qui pensent viser certains masters sélectifs, **essayez d'avoir une bonne note !** Ça ne peut que vous aider pour votre dossier, et puis il paraît que Ly Vath a une vraie importance dans la sélection des élèves pour des départs à l'étranger !*

2 Correction

Exercice 2.1 : Un peu de probabilités discrètes

On dit qu'une variable aléatoire X à valeurs dans $\mathbb{N}^*$ suit une loi géométrique de paramètre θ $(0 < \theta < 1)$ notée $\mathcal{G}(\theta)$ si la loi de X est définie par :

$$\forall k \in \mathbb{N}^*, \quad \mathbb{P}(X = k) = \theta(1-\theta)^{k-1}. \tag{1}$$

1. Déterminer $\mathbb{P}(X = 2)$ et calculer $\mathbb{P}(X < 3)$,
2. Démontrer que la fonction définie au (1) est bien une loi de probabilité.
3. Déterminer G_X, la fonction génératrice de X. En déduire l'espérance et la variance de X.
4. Pour $k \in \mathbb{N}$ et $m \in \mathbb{N}$, montrer que $\mathbb{P}(X > k + m \mid X > m) = \mathbb{P}(X > k)$
5. Soit Y et Z deux variables aléatoires indépendantes suivant la même loi $\mathcal{G}(\theta)$.
   (a) Déterminer la loi de $S = Y + Z$.
   (b) Déterminer la loi conditionnelle de Y sachant $\{S = s\}$.
   (c) Déterminer la loi de $U = \max\{Y, Z\}$.

Démonstration.

1. On a par définition : $\mathbb{P}(X=2) = \theta(1-\theta)^{2-1} = \theta(1-\theta)$.
   Et, comme $X(\Omega) = \mathbb{N}^*$:
$$\mathbb{P}(X<3) = \mathbb{P}(X=1) + \mathbb{P}(X=2) = \theta + \theta(1-\theta) = \theta(2-\theta)$$

2. • Comme $\theta \in]0,1[$, $\forall k \in \mathbb{N}^*, \quad \mathbb{P}(X=k) = \theta(1-\theta)^{k-1} \geq 0$.
   • Et deuxièmement :
$$\begin{aligned}\sum_{k\in\mathbb{N}^*} \mathbb{P}(X=k) &= \theta \sum_{k=1}^{+\infty} (1-\theta)^{k-1} \\ &= \theta \sum_{\ell=0}^{+\infty} (1-\theta)^{\ell} && \text{(changement de variable : } k=\ell+1) \\ &= \theta \times \frac{1}{1-(1-\theta)} && \text{(série géométrique)} \\ &= \frac{\theta}{\theta} = 1\end{aligned}$$
   Cela permet d'affirmer que la fonction définie au (1) est bien **une loi de probabilité**.

3. Formellement, par définition de la fonction génératrice, on a :
$$\forall t \in \mathbb{R}, \quad G_X(t) \triangleq \mathbb{E}\left[t^X\right] = \sum_{k\in\mathbb{N}^*} t^k\, \mathbb{P}(X=k) = t\,\theta \sum_{k\in\mathbb{N}^*} (t(1-\theta))^{k-1}$$
   Cette série converge uniquement pour $t \in]-\frac{1}{1-\theta}, \frac{1}{1-\theta}[\triangleq I_\theta$, car $\theta \in]0,1[$.
   Ainsi, on définit la fonction génératrice G_X par :
$$\begin{aligned}\forall t \in I_\theta, \quad G_X(t) &= t\,\theta \sum_{k=1}^{+\infty} (t(1-\theta))^{k-1} \\ &= t\,\theta \sum_{\ell=0}^{+\infty} (t(1-\theta))^{\ell} && \text{(changement de variable : } k=\ell+1) \\ &= t \times \theta \times \frac{1}{1-t(1-\theta)}\end{aligned}$$
   Et donc :
$$\forall t \in I_\theta, \quad \boxed{G_X(t) = \frac{t\,\theta}{t\,\theta - t + 1}}$$
   Sur son intervalle de définition, nos connaissances sur les séries entières nous permettent d'assurer que G_X est $\mathcal{C}^\infty$ sur I_θ. On en déduit alors G'_X et G''_X sur I_θ :
$$G'_X(t) = \frac{\theta}{(t\,\theta - t + 1)^2} \qquad G''_X(t) = \frac{2\theta(1-\theta)}{(t\,\theta - t + 1)^3}$$
   Et finalement :
$$\boxed{\mathbb{E}[X] = G'_X(1) = \frac{1}{\theta}}$$
   Et pour la variance :
$$\mathrm{Var}(X) = G''_X(1) + G'_X(1) - \left(G'_X(1)\right)^2 = \frac{2(1-\theta)}{\theta^2} + \frac{1}{\theta} - \left(\frac{1}{\theta}\right)^2$$
$$\Longrightarrow \quad \boxed{\mathrm{Var}(X) = \frac{1-\theta}{\theta^2}}$$

4. Soit $k \in \mathbb{N}$ et $m \in \mathbb{N}$,

$$\mathbb{P}(X > k+m \,|\, X > m) = \frac{\mathbb{P}(X > k+m \,, X > m)}{\mathbb{P}(X > m)}$$

Si $k+m > m$, on a $\{X > k+m\} \bigcap \{X > m\} = \{X > k+m\}$ car le premier ensemble est inclus dans le deuxième. On en déduit donc la suite du calcul :

$$\mathbb{P}(X > k+m \,|\, X > m) = \frac{\mathbb{P}(X > k+m)}{\mathbb{P}(X > m)} = \frac{\displaystyle\sum_{\ell=k+m+1}^{+\infty} \mathbb{P}(X=k)}{\displaystyle\sum_{\ell=m+1}^{+\infty} \mathbb{P}(X=k)} \underset{\ell=\tilde{\ell}+1}{=} \frac{\displaystyle\sum_{\ell=k+m}^{+\infty} (1-\theta)^\ell}{\displaystyle\sum_{\ell=m}^{+\infty} (1-\theta)^\ell}$$

Et, *en connaissant bien nos séries géométriques*, on trouve donc :

$$\mathbb{P}(X > k+m \,|\, X > m) = \frac{\dfrac{(1-\theta)^{k+m}}{1-(1-\theta)}}{\dfrac{(1-\theta)^m}{1-(1-\theta)}} = (1-\theta)^k$$

De plus, on remarque que :

$$\mathbb{P}(X > k) = \sum_{\ell=k+1}^{+\infty} \theta(1-\theta)^{\ell-1} \underset{\ell=\tilde{\ell}+1}{=} \theta \sum_{\ell=k}^{+\infty} \theta(1-\theta)^\ell = \theta \times \frac{(1-\theta)^k}{1-(1-\theta)} = (1-\theta)^k$$

Finalement, on a donc bien :

$$\forall (k,m) \in {\mathbb{N}^*}^2, \qquad \boxed{\mathbb{P}(X > k+m \,|\, X > m) = \mathbb{P}(X > k)}$$

Note : On dit alors que X est *sans mémoire.* Et pour le petit point culture, la loi géométrique est en réalité **l'unique loi de probabilité discrète à perte de mémoire**.

5. (a) On suppose que Y et Z sont deux variables aléatoires indépendantes suivant la même loi $\mathcal{G}(\theta)$. On note : $S = Y + Z$. On peut alors déterminer la loi de S.
Soit $n \in \mathbb{N}^*$, on a :

$$\begin{aligned}
\mathbb{P}(S=n) &= \mathbb{P}(Y+Z=n) = \mathbb{P}(\{Y+Z=n\}) \\
&= \mathbb{P}\left(\bigcup_{k=1}^{n-1} (\{Y=k\} \cap \{Z=n-k\}) \right) \\
&= \sum_{k=1}^{n-1} \mathbb{P}(Y=k \,, Z=n-k) && \text{(par union disjointe)} \\
&= \sum_{k=1}^{n-1} \mathbb{P}(Y=k)\, \mathbb{P}(Z=n-k) && \text{(Y et Z indépendantes)} \\
&= \sum_{k=1}^{n-1} \theta(1-\theta)^{k-1}\, \theta(1-\theta)^{n-k-1} \\
&= \sum_{k=1}^{n-1} \theta^2(1-\theta)^{n-2} \\
\Longrightarrow \quad & \boxed{\mathbb{P}(S=n) = (n-1)\,\theta^2(1-\theta)^{n-2}}
\end{aligned}$$

(b) Soit $s \in \mathbb{N}^*$, $\forall k \in [\![1, s-1]\!]$, on a :

$$\begin{aligned}
\mathbb{P}(Y = k \mid S = s) &= \frac{\mathbb{P}(Y = k\,, Y + Z = s)}{\mathbb{P}(S = s)} \\
&= \frac{\mathbb{P}(Y = k\,, Z = s - k)}{\mathbb{P}(S = s)} \\
&= \frac{\mathbb{P}(Y = k)\,\mathbb{P}(Z = s - k)}{\mathbb{P}(S = s)} && \text{(Y et Z indépendantes)} \\
&= \frac{\theta(1-\theta)^{k-1}\,\theta(1-\theta)^{s-k-1}}{(s-1)\,\theta^2(1-\theta)^{s-2}} \\
&= \frac{1}{s-1}
\end{aligned}$$

Ainsi, si l'on note $(Y \,|\, S = s)$ la loi de Y sachant $\{S = s\}$ définie sur $[\![1, s-1]\!]$, on vient de montrer qu'elle suit une loi uniforme discrète. On a donc :

$$\boxed{(Y \,|\, S = s) \sim \mathbb{U}_{[\![1,s-1]\!]}}$$

(c) On note $U = \max\{Y, Z\}$. Pour déterminer la loi de U, il suffit de remarquer que :

$$\forall n \in \mathbb{N}^*, \quad (U \leq n \iff Y \leq n \textbf{ et } Z \leq n)$$

On peut alors écrire que :

$$\begin{aligned}
\mathbb{P}(U \leq n) &= \mathbb{P}(\{Y \leq n\} \cap \{Z \leq n\}) \\
&= \mathbb{P}(Y \leq n)\,\mathbb{P}(Z \leq n) && \text{(Y et Z indépendantes)} \\
&= \theta \sum_{k=1}^{n} (1-\theta)^{k-1} \;\times\; \theta \sum_{k=1}^{n} (1-\theta)^{k-1} \\
&= \theta^2 \left(\sum_{\ell=0}^{n-1} (1-\theta)^{\ell} \right)^2 && \text{(changement de variable : } k = \ell + 1) \\
&= \theta^2 \left(\frac{1 - (1-\theta)^n}{1 - (1-\theta)} \right)^2 \\
&= (1 - (1-\theta)^n)^2
\end{aligned}$$

Et donc :

$$\boxed{\mathbb{P}(U \leq n) = 1 - 2(1-\theta)^n + (1-\theta)^{2n}}$$

Pour finir, comme U suit une loi discrète sur $\mathbb{N}^*$, on remarque que :

$$\forall n \in \mathbb{N}^*, \quad \mathbb{P}(U = n) = \mathbb{P}(U \leq n) - \mathbb{P}(U \leq n-1)$$

Et alors :

$$\begin{aligned}
\mathbb{P}(U = n) &= (1-\theta)^{2n} - 2(1-\theta)^n - \left[(1-\theta)^{2(n-1)} - 2(1-\theta)^{n-1}\right] \\
&= (1-\theta)^{2n} - (1-\theta)^{2n-2} + 2(1-\theta)^{n-1} - 2(1-\theta)^n \\
&= (1-\theta)^{2n-2}[(1-\theta)^2 - 1] + 2(1-\theta)^{n-1}[1 - (1-\theta)] \\
&= (1-\theta)^{2n-2}(\theta^2 - 2\theta) + 2\,\theta(1-\theta)^{n-1} \\
&= \theta(1-\theta)^{2n-2}(\theta - 2) + 2\,\theta(1-\theta)^{n-1}
\end{aligned}$$

Finalement, on a :

$$\forall n \in \mathbb{N}^*, \quad \boxed{\mathbb{P}(U = n) = 2\,\theta(1-\theta)^{n-1} - \theta(2-\theta)(1-\theta)^{2n-2}}$$

$\square$

Exercice 2.2 : Un classique de Ly Vath

1. On considère l'application définie sur $\mathbb{R}$ par

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

Montrer que f est une densité de probabilité.

2. Soient X et Y deux variables aléatoires réelles indépendantes telles que X admette f pour densité et Y suive la loi uniforme sur $[-1, 1]$.
   (a) Quelle est la loi de la variable $Z = -Y$?
   (b) Déterminer une densité h de la variable $X - Y$.

Démonstration.

1. • $\forall x \in [-1, 1]$, on a : $|x| \leq 1$, et donc $f(x) \geq 0$. Sinon, $f(x) = 0 \geq 0$
   On en déduit donc que : $\forall x \in \mathbb{R}$, $f(x) \geq 0$.
   • On a :

$$\begin{aligned} \int_{\mathbb{R}} f(x)\,\mathrm{d}x &= \int_{-1}^{1} (1 - |x|)\,\mathrm{d}x \\ &= \int_{-1}^{0} (1 - |x|)\,\mathrm{d}x + \int_{0}^{1} (1 - |x|)\,\mathrm{d}x \\ &= \int_{-1}^{0} (1 + x)\,\mathrm{d}x + \int_{0}^{1} (1 - x)\,\mathrm{d}x \\ &= \left[x + \frac{x^2}{2}\right]_{-1}^{0} + \left[x - \frac{x^2}{2}\right]_{0}^{1} \\ &= -\left(-1 + \frac{1}{2}\right) + \left(1 - \frac{1}{2}\right) = 1 \end{aligned}$$

   On peut donc affirmer que f est bien une densité de probabilité sur $\mathbb{R}$.

2. (a) On suppose que $Y \sim \mathcal{U}(-1, 1)$, on sait donc que $\forall x \in \mathbb{R}$, $f_Y(x) = \frac{1}{2}\mathbb{1}_{[-1,1]}(x)$.
   On note alors $Z = -Y$. Pour déterminons la loi de Z, j'ai choisi de passer par les fonctions de répartitions :

$$\begin{aligned} \forall x \in \mathbb{R}, \quad F_Z(x) \triangleq \mathbb{P}(Z \leq x) &= \mathbb{P}(-Y \leq x) \\ &= \mathbb{P}(Y \geq -x) \\ &= 1 - \mathbb{P}(Y \leq -x) \qquad \text{(la mesure d'un point est nulle)} \\ &= 1 - F_Y(-x) \end{aligned}$$

   En dérivant, on a alors :

$$f_Z(x) = {F_Z}'(x) = -[-{F_Y}'(-x)] = f_Y(-x) = \frac{1}{2}\mathbb{1}_{[-1,1]}(-x)$$

   Or, comme on a $\mathbb{1}_{[-1,1]}(-x) = \mathbb{1}_{[-1,1]}(x)$ puisque l'intervalle est centré autour de 0. On en déduit donc que $f_Z = f_Y$. Et donc,

$$\boxed{Z \sim \mathcal{U}(-1, 1)}$$

(b) On pose $T = X - Y$ de densité h. On remarque alors que $T = X + Z$.
Le cours nous permet alors d'affirmer que la somme de X et Z a pour densité la fonction $h = f_X * f_Z$, produit de convolution des densités de X et Z.
On a donc, $\forall x \in \mathbb{R}$:

$$\begin{aligned} h(x) = (f_X * f_Z)(x) &= \int_{\mathbb{R}} f_X(t)\, f_Z(x-t)\,\mathrm{d}t \\ &= \int_{\mathbb{R}} (1-|t|)\mathbb{1}_{[-1,1]}(t) \times \frac{1}{2}\mathbb{1}_{[-1,1]}(x-t)\,\mathrm{d}t \\ &= \frac{1}{2}\int_{\mathbb{R}} (1-|t|)\mathbb{1}_{[-1,1]}(t) \times \mathbb{1}_{[-1+x,1+x]}(t)\,\mathrm{d}t \\ &= \frac{1}{2}\int_{\mathbb{R}} (1-|t|)\,\mathbb{1}_{[-1,1]\cap[-1+x,1+x]}(t)\,\mathrm{d}t \end{aligned}$$

On suppose pour commencer que $0 \leq x \leq 1$:

$$\begin{aligned} h(x) &= \frac{1}{2}\int_{\mathbb{R}} (1-|t|)\mathbb{1}_{[-1+x,1]}(t)\,\mathrm{d}t \qquad (\text{ car } [-1,1]\cap[-1+x,1+x] = [-1+x,1]) \\ &= \frac{1}{2}\int_{-1+x}^{1} (1-|t|)\,\mathrm{d}t \\ &= \frac{1}{2}\left(\int_{x-1}^{0} (1-|t|)\,\mathrm{d}t + \int_{0}^{1} (1-|t|)\,\mathrm{d}t\right) \\ &= \frac{1}{2}\left(\int_{x-1}^{0} (1+t)\,\mathrm{d}t + \int_{0}^{1} (1-t)\,\mathrm{d}t\right) \\ &= \frac{1}{2}\left[-\left((x-1) + \frac{(x-1)^2}{2}\right) + \left(1 - \frac{1}{2}\right)\right] \\ &= \frac{1}{2}\left[-x+1-\left(\frac{x^2}{2} - x + \frac{1}{2}\right) + \frac{1}{2}\right] \\ &= \frac{1}{2}\left(1 - \frac{x^2}{2}\right) \end{aligned}$$

Si $1 < x \leq 2$:

$$\begin{aligned} h(x) &= \frac{1}{2}\int_{\mathbb{R}} (1-|t|)\mathbb{1}_{[-1+x,1]}(t)\,\mathrm{d}t \\ &= \frac{1}{2}\int_{-1+x}^{1} (1-|t|)\,\mathrm{d}t \\ &= \frac{1}{2}\int_{x-1}^{1} (1-t)\,\mathrm{d}t \\ &= \frac{1}{2}\left[\left(1-\frac{1}{2}\right) - \left((x-1) - \frac{(x-1)^2}{2}\right)\right] \\ &= \frac{1}{2}\left[\frac{1}{2} - x + 1 + \frac{(x-1)^2}{2}\right] \\ &= 1 - x + \frac{x^2}{4} \end{aligned}$$

De la même manière, avec $-2 \leq x < -1$ et $-1 \leq x < 0$, en se ramenant sur l'intervalle

$[-1, 1+x]$, on arrive à :

$$h(x) = \begin{cases} 1 + x + \dfrac{x^2}{4} & \text{si } x \in [-2, -1[\\ \dfrac{1}{2}\left(1 - \dfrac{x^2}{2}\right) & \text{si } x \in [-1, 0[\end{cases}$$

Et, avec $x \in \mathbb{R} \backslash [-2, 2]$, on a $\mathbb{1}_{[-1,1]} \times \mathbb{1}_{[-1+x,1+x]} = \mathbb{1}_{\emptyset} = 0$. Donc $h(x) = 0$.

Ainsi, on a donc :

$$\boxed{\forall x \in \mathbb{R}, \quad h(x) = \begin{cases} 1 + x + \dfrac{x^2}{4} & \text{si } x \in [-2, -1[\\ \dfrac{1}{2}\left(1 - \dfrac{x^2}{2}\right) & \text{si } x \in [-1, 1] \\ 1 - x + \dfrac{x^2}{4} & \text{si } x \in]1, 2] \\ 0 & \text{sinon} \end{cases}}$$

Note *: Promis c'est la bonne réponse, je me suis bien fait chier à vérifier que $\int_{\mathbb{R}} h(x)\,\mathrm{d}x$ vaille bien 1, et c'est bien le cas !*

$\square$

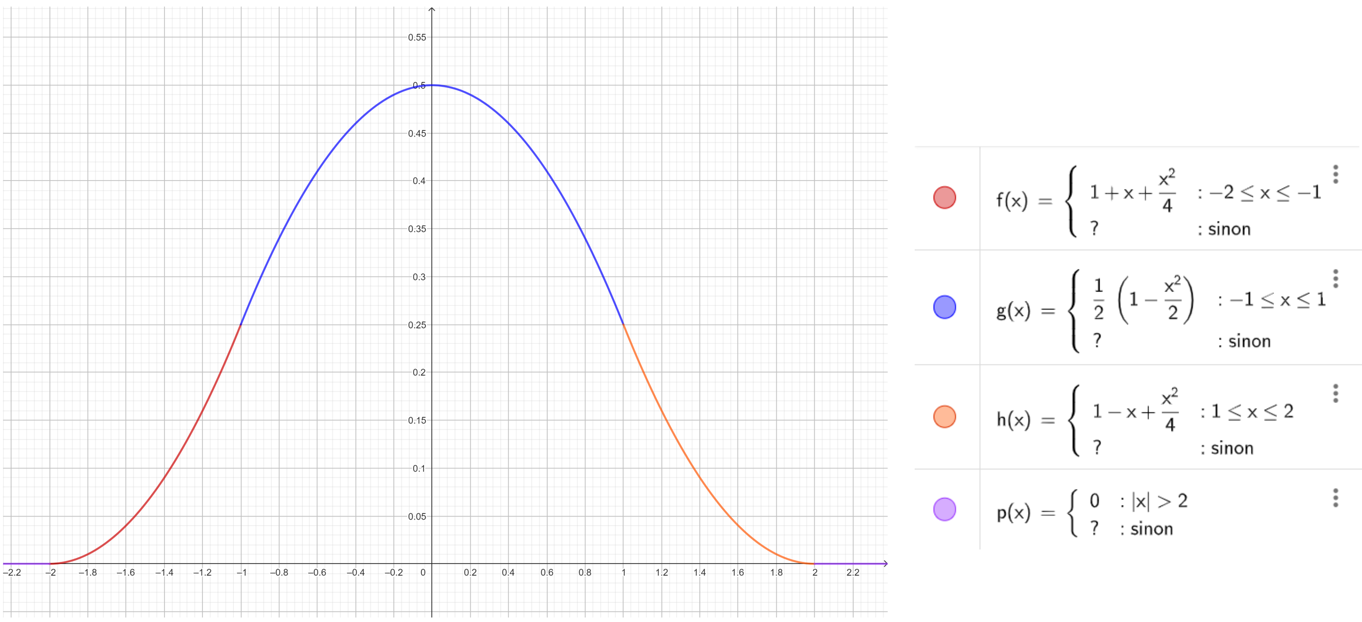

Figure 1 – Courbe représentative de h

Exercice 2.3 : Sur la loi du Khi-2

Soit $X_1, \dots, X_n$ des variables aléatoires réelles indépendantes de même loi, d'espérance m, de variance σ^2 finie. On pose

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \Sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2 \quad V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

On suppose que la loi de X_i est la loi gaussienne $\mathcal{N}(m, \sigma^2)$.

Rappel : Soit une suite $(Y_i, i \geq 1)$ de variables aléatoires indépendantes et de loi $\mathcal{N}(0,1)$, on rappelle que $\sum_{i=1}^{n} Y_i^2$ suit la loi $\chi^2(n)$ (à n degrés de liberté).

1. Quelle est la loi de $\overline{X}_n$?
2. Quelle est la loi de $\frac{n\Sigma_n^2}{\sigma^2}$?
3. Montrer que $\overline{X}_n$ et V_n sont indépendantes.

Démonstration.

1. On a :
$$\forall i \in [\![1, n]\!], \quad X_i \sim \mathcal{N}(m, \sigma^2)$$
On sait alors, d'après le TCL ou plus simplement avec un passage par les fonctions caractéristiques, que le moyenne empirique $\overline{X}_n$ suit une loi normale de paramètre m et $\frac{\sigma^2}{n}$. On a donc :
$$\boxed{\overline{X}_n \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right)}$$

2. On a :
$$\frac{n\Sigma_n^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - m)^2 = \sum_{i=1}^{n}\left(\frac{X_i - m}{\sigma}\right)^2$$
Or, comme $X_i \sim \mathcal{N}(m, \sigma^2)$, on en déduit que :
$$\forall i \in [\![1, n]\!], \quad \frac{X_i - m}{\sigma} \sim \mathcal{N}(0, 1)$$
Et donc, comme les $(X_i)_{i \in [\![1,n]\!]}$ sont indépendantes, et par définition de la loi du χ^2, on en déduit que :
$$\boxed{\frac{n\Sigma_n^2}{\sigma^2} \sim \chi^2(n)}$$

3. *En toute honnêteté, je n'arrive pas à faire cette question. J'ai essayé l'idée principale selon laquelle la non-corrélation suffit car on travaille avec des variables normales, mais mon calcul arrive difficilement à quelque chose...*
*J'ai l'impression qu'il faut utiliser des gros théorèmes comme le **Théorème de Cochran** qui permettent d'arriver directement au résultat, ou de tout simplement y aller au bluff en expliquant que "évidemment, quand on développe V_n, la non corrélation saute aux yeux !"*

   Bref, si quelqu'un a une réponse claire à cette question, je suis preneur ! Je me sens un peu bête là...

   Je vous propose donc la réponse de ChatGPT sur cette question :

Pour démontrer que $\overline{X}_n$ et V_n sont indépendantes, analysons d'abord ces deux variables :

- $\overline{X}_n$ est la moyenne des X_i, où chaque X_i suit une loi normale $\mathcal{N}(m, \sigma^2)$. Ainsi, $\overline{X}_n$ suit une loi normale $\mathcal{N}(m, \frac{\sigma^2}{n})$.

- V_n est la variance échantillonnale, définie par $V_n = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$.

Considérons la forme développée de V_n :

$$\begin{aligned} V_n &= \frac{1}{n-1}\sum_{i=1}^n (X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2) \\ &= \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n\overline{X}_n^2\right). \end{aligned}$$

Cette forme montre que V_n peut être exprimée comme une combinaison linéaire des X_i, qui est indépendante de $\overline{X}_n$. Dans le contexte des variables aléatoires normales, **l'indépendance est équivalente à la non-correlation**. Ainsi, si nous pouvons montrer que la covariance entre $\overline{X}_n$ et V_n est nulle, cela impliquerait leur indépendance.

En conclusion, $\overline{X}_n$ et V_n sont indépendantes car V_n est une combinaison linéaire des X_i qui ne dépend pas directement de $\overline{X}_n$, et dans le cas de la loi normale, l'absence de corrélation implique l'indépendance.

□

Exercice 2.4 : Un exemple de VA à densité non corrélées ET non indépendantes

Soit X une variable gaussienne centrée réduite et ε une variable vérifiant $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = \frac{1}{2}$ indépendantes.

1. Quelle est la loi de $Y = \varepsilon X$?
2. Donner la matrice de covariance du vecteur (X, Y).
3. Le vecteur (X, Y) est-il gaussien ? Les variables X et Y sont-elles indépendantes ?

Démonstration.

1. J'ai choisi la **méthode de la fonction muette**.
   Soit g une fonction bornée et mesurable. Pour déterminer la loi de Y, nous calculons l'espérance de $g(Y)$ en utilisant l'indépendance de X et ε :

$$\begin{aligned} \mathbb{E}[g(Y)] &= \mathbb{E}[g(\varepsilon X)] \\ &= \mathbb{E}\left[g(X)\mathbb{1}_{\{\varepsilon=1\}} + g(-X)\mathbb{1}_{\{\varepsilon=-1\}}\right] \\ &= \mathbb{E}[g(X)]\mathbb{E}\left[\mathbb{1}_{\{\varepsilon=1\}}\right] + \mathbb{E}[g(-X)]\mathbb{E}\left[\mathbb{1}_{\{\varepsilon=-1\}}\right] \\ &= \mathbb{E}[g(X)]\mathbb{P}(\varepsilon = 1) + \mathbb{E}[g(-X)]\mathbb{P}(\varepsilon = -1) \\ &= \frac{1}{2}\mathbb{E}[g(X)] + \frac{1}{2}\mathbb{E}[g(-X)] \end{aligned}$$

Puisque X est symétrique par rapport à zéro, $g(-X)$ a la même distribution que $g(X)$. En effet :

$$\begin{aligned}\mathbb{E}[g(-X)] &= \int_{-\infty}^{+\infty} g(-x) f_X(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{+\infty} g(u) f_X(-u)\,\mathrm{d}u && \text{(changement de signe } x = -u\text{)} \\ &= \int_{-\infty}^{+\infty} g(u) f_X(u)\,\mathrm{d}u && \text{(par symétrie de la densité } f_X\text{)} \\ &= \mathbb{E}[g(X)]\end{aligned}$$

où f_X est la densité de X, correspondant à la distribution normale centrée réduite.

Par conséquent, nous avons :

$$\mathbb{E}[g(Y)] = \frac{1}{2}\mathbb{E}[g(X)] + \frac{1}{2}\mathbb{E}[g(X)] = \mathbb{E}[g(X)],$$

Ce qui démontre que Y a la même distribution que X. On a donc :

$$\boxed{Y \sim \mathcal{N}(0,1)}$$

2. La matrice de covariance de (X, Y) est donnée par :

$$\Sigma = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Var}(Y) \end{pmatrix}.$$

Puisque X et ε sont indépendants, la covariance entre X et Y est :

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] \\ &= \mathbb{E}[X \times \varepsilon X] - \mathbb{E}[X]\,\mathbb{E}[\varepsilon X] \\ &= \mathbb{E}[X^2 \times \varepsilon] - \mathbb{E}[X]^2\mathbb{E}[\varepsilon] \\ &= \mathbb{E}[X^2]\,\mathbb{E}[\varepsilon] - \mathbb{E}[X]^2\mathbb{E}[\varepsilon] \\ &= 0 && \text{(car } \mathbb{E}[\varepsilon] = 0\text{)}\end{aligned}$$

Par conséquent, la matrice de covariance est :

$$\boxed{\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.}$$

3. On vient de prouver dans la question 2 que **les variables X et Y sont non corrélées**. Seulement, on remarque que :

$$\mathbb{P}(X + Y = 0) = \mathbb{P}(X + \varepsilon X = 0) = \mathbb{P}(\varepsilon = -1) = \frac{1}{2}$$

Si (X, Y) était un vecteur gaussien, alors toute combinaison linéaire de X et Y suivrait une loi normale. Mais on vient de montrer que $\mathbb{P}(X + Y = 0) = \frac{1}{2} \neq 0$, ce qui prouve que $X + Y$ ne suit pas une loi normale.

On aboutit à une contradiction. Donc (X, Y) **n'est pas un vecteur gaussien**.

Comme X et Y sont non corrélées, l'indépendance de X et Y entraînerait que (X, Y) soit un vecteur gaussien. On peut donc conclure que :

$$\boxed{X \text{ et } Y \textbf{ ne sont pas indépendantes.}}$$

Note 1 *: Les idées de correction de cet exo viennent tout droit de la correction d'Alphoenix du partiel de 2021-2022 (lien), un grand merci à lui !*

Note 2 *: La question 3 provient de ce document : lien (page 5). Il donne une correction très synthétique de tout l'exo, avec notamment une correction plus simple par les fonctions de répartition pour la question 1. Je vous conseille d'aller y jeter un coup d'oeil !*

$\square$