Exercice 1. Soit X une v.a. gaussienne de densité : $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \forall x \in \mathbb{R}$. On note par $\Phi(x)$ sa fonction de répartition.

1. Calculer $\mathbb{E}(X+1)$ et $\mathbb{V}(-2X)$
2. On pose $Y = e^{|X|}$.
   (a) Déterminer la fonction de répartition de Y, notée $F_Y(x)$, en fonction de $\Phi(.)$.
   (b) Déterminer la densité de Y, notée $f_Y(x)$.
3. Soit Z la v.a réelle de densité

$$f_Z(x) = \begin{cases} \frac{1}{x}\varphi(\ln(x)) & \text{si } x > 0, \\ 0 & \text{sinon.} \end{cases}$$

   (a) Déterminer la fonction de répartition de Z, notée $F_Z(x)$, en fonction de $\Phi(.)$
   (b) Calculer $\mathbb{E}(Z)$

Solution.

1. X sui une loi normale centrée réduite. Donc en utilisant la linéarité de l'espérance et la propriété quadratique de la variance on trouve : $\mathbb{E}(X+1) = \mathbb{E}(X) + 1 = 1$ et $\mathbb{V}(-2X) = (-2)^2\mathbb{V}(X) = 4$
2. On a $X(\Omega) = \mathbb{R}$. Donc $(|X|)(\Omega) = [0, \infty[$ et $Y(\Omega) = [1, \infty[$. Par conséquent, pour tout $x < 1$, on a

$$\mathbb{P}(Y \leqslant x) = 0.$$

   Pour tout $x \geqslant 1$, on a

$$\begin{aligned} \mathbb{P}(Y \leqslant x) &= \mathbb{P}(e^{|X|} \leqslant x) = \mathbb{P}(|X| \leqslant \ln x) \\ &= \mathbb{P}(-\ln x \leqslant X \leqslant \ln x) \\ &= \mathbb{P}(X \leqslant \ln x) - \mathbb{P}(X \leqslant -\ln x). \end{aligned}$$

   Au final, en notant $\Phi(x)$ la fonction de répartition de X, la fonction de répartition de Y est

$$F_Y(x) = \begin{cases} \Phi(\ln x) - \Phi(-\ln x) & \text{si } x \geqslant 1, \\ 0 & \text{sinon.} \end{cases}$$

   comme $-\ln x \leqslant 0$ pour tout $x \geqslant 1$, alors

$$F_Y(x) = \begin{cases} 2\Phi(\ln x) - 1 & \text{si } x \geqslant 1, \\ 0 & \text{sinon.} \end{cases}$$

   Par dérivation, on en déduit la densité de Y :

$$f_Y(x) = F'_Y(x) = \begin{cases} \frac{2}{x}\varphi(\ln x) & \text{si } x \geqslant 1, \\ 0 & \text{sinon.} \end{cases}$$

3. On a $Z(\Omega) =]0, \infty[$, par conséquent, pour tout $x \leqslant 0$, on a

$$\mathbb{P}(Z \leqslant x) = 0.$$

Pour tout $x > 0$, on a :

$$\begin{aligned}
\mathbb{P}(Z \leqslant x) &= \int_{-\infty}^{x} f_Z(u)du = \int_{0}^{x} f_Z(u)du \\
&= \int_{0}^{x} \frac{1}{u}\varphi(u)du = \int_{0}^{x} \left(\frac{d\ \Phi(\ln u)}{du}\right) du \\
&= [\Phi(\ln u)]_0^x = \Phi(\ln x) - \lim_{x \to -\infty} \Phi(x) = \Phi(\ln x)
\end{aligned}$$

Au final, la fonction de répartition de Z est

$$F_Z(x) = \begin{cases} \Phi(\ln x) & \text{si } x > 0, \\ 0 & \text{sinon.} \end{cases}$$

4. Espérance de Z :

$$\mathbb{E}(Z) = \int_{-\infty}^{\infty} x f_Z(x)dx = \int_{0}^{\infty} x\frac{1}{x}\varphi(\ln x)dx = \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(\ln x)^2}{2}} dx$$

On posant $t = \ln x$ (donc $t \in \mathbb{R}$, $x = e^t$ et $dx = e^t dt$), on obtient :

$$\mathbb{E}(Z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} e^t dt = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t^2-2t+1-1)}{2}} dt = e^{\frac{1}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-1)^2}{2}} dt = e^{\frac{1}{2}}$$

car $h(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-1)^2}{2}}$ est la densité d'une v.a normale d'espérance 1 et d'écart type 1. Son intégrale sur $\mathbb{R}$ vaut donc 1.
On a donc $\mathbb{E}(Z) = \sqrt{e}$

Exercice 2. On rappelle que si (X, Y) est un couple de v.a de fonction de densité de probabilité jointe définie pour $|\rho| < 1$ (ρ étant le coefficient de corrélation) par

$$f_{(X,Y)}(x, y) = \frac{1}{2\,\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right\} \tag{1}$$

alors X est de loi $\mathcal{N}(\mu_X, \sigma_X^2)$ et Y de loi $\mathcal{N}(\mu_Y, \sigma_Y^2)$ avec $(\mu_X, \mu_Y) \in \mathbb{R}^2$ et $(\sigma_X, \sigma_Y) \in {\mathbb{R}_+^\star}^2$. Soit (X, Y) un couple de v.a. de fonction de densité de probabilité jointe définie pour $|\rho| < 1$ par

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right). \tag{2}$$

1. Déterminer (aucun calcul n'est nécessaire) la loi de X et la loi de Y

2. Déterminer la matrice de variance-covariance $\boldsymbol{\Sigma}$ du vecteur $\boldsymbol{V} = (X, Y)^t$

3. Quelle est la condition nécessaire est suffisante pour que le couple (X, Y) soit indépendant ?

4. Trouver la fonction de densité de probabilité conditionnelle de Y sachant X

5. Reconnaître la loi correspondante

6. Trouver la loi du vecteur $\boldsymbol{W} = (X - 2Y + 1, 2X - Y)^t$
7. En déduire la loi de $H = X - 2Y + 1$ et celle de $T = 2X - Y$
8. Quelle est la condition nécessaire est suffisante pour que le couple (H, T) soit indépendant ?

Solution.

1. Il est clair que X et Y ont la même loi $\mathcal{N}(0, 1)$
2. La matrice de variance-covariance $\boldsymbol{\Sigma}$ du vecteur $\boldsymbol{V} = (X, Y)^t$ est donné par
$$\boldsymbol{\Sigma} = \text{Cov}((X, Y)^t) = \begin{pmatrix} \text{Var(X)} & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & \text{Var}(Y) \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$
car $\rho = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} = \text{Cov}(X, Y)$.
3. Soient $f_X(.)$ et $f_Y(.)$ les densités respectives de X et de Y. On a donc d'après la question 1. $f_X(t) = f_Y(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$ pour tout $t \in \mathbb{R}$. Il est donc clair que pour que le couple (X, Y) soit indépendant, i.e $f_{(X,Y)}(x, y) = f_X(x) f_Y(y)$ pour tout $(x, y) \in \mathbb{R}^2$, il faut et il suffit que $\rho = 0$. Le couple (X, Y) est en effet gaussien, s'il est décorrélé ($\rho = 0$) il est alors indépendant.
4. La fonction de densité de probabilité conditionnelle de Y sachant X s'obtient en utilisant le théorème de Bayes et est ainsi donnée par, pour tout $(x, y) \in \mathbb{R}^2$:
$$\begin{aligned}
f_{Y|X=x}(y) &= \frac{f_{(X,Y)}(x, y)}{f_X(x)} \\
&= \frac{\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right)}{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - x^2(1-\rho^2) - 2\rho xy + y^2}{2(1-\rho^2)}\right) \\
&= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left(-\frac{(y - \rho x)^2}{2(1-\rho^2)}\right) \\
&= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right) \text{ avec } \sigma = \sqrt{1-\rho^2} \text{ et } \mu = \rho x.
\end{aligned}$$
5. On reconnaît donc la densité de la loi $\mathcal{N}(\rho X, 1 - \rho^2)$
6. On peut écrire $\boldsymbol{W} = (X - 2Y + 1, 2X - Y)^t$ sous la forme $\boldsymbol{W} = \mathbf{A}\boldsymbol{V} + \boldsymbol{B}$, où $\mathbf{A} = \begin{pmatrix} 1 & -2 \\ 2 & -1 \end{pmatrix}$ et $\mathbf{B} = (1, 0)^t$. Comme $\boldsymbol{V}$ est un vecteur gaussien, $\boldsymbol{W}$ est alors aussi un vecteur gaussien. Son espérance est donnée par $\mathbb{E}[\boldsymbol{W}] = \boldsymbol{B}$ (le vecteur $\boldsymbol{V}$ étant centré) et sa matrice de

variance-covariance est donnée par :

$$\begin{aligned}
\boldsymbol{S} = \mathrm{Cov}(\boldsymbol{W}) &= \mathbf{A}\ \mathrm{Cov}(\boldsymbol{V})\mathbf{A}^t = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^t \\
&= \begin{pmatrix} 1 & -2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -2 & -1 \end{pmatrix} \\
&= \begin{pmatrix} 1-2\rho & \rho-2 \\ 2-\rho & 2\rho-1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -2 & -1 \end{pmatrix} \\
&= \begin{pmatrix} 1-2\rho-2\rho+4 & 2-4\rho-\rho+2 \\ 2-\rho-4\rho+2 & 4-2\rho-2\rho+1 \end{pmatrix} \\
&= \begin{pmatrix} 5-4\rho & 4-5\rho \\ 4-5\rho & 5-4\rho \end{pmatrix}.
\end{aligned}$$

Au finale on a $\boldsymbol{W} \sim \mathcal{N}(\boldsymbol{B}, \boldsymbol{S})$.

7. On a d'après la question précédente $\boldsymbol{W} = (H, T)^t$ est gaussien d'espérance $(\mathbb{E}[H], \mathbb{E}[T])^t = \boldsymbol{B}$ et de matrice de variance-covariance $\begin{pmatrix} \mathrm{Var}(H) & \mathrm{Cov}(H,T) \\ \mathrm{Cov}(H,T) & \mathrm{Var}(T) \end{pmatrix} = \boldsymbol{S}$. On en déduit donc que chacune de ses deux composantes H et T est gaussienne et que $\mathbb{E}[H] = 1$, $\mathrm{Var}(H) = 5-4\rho$, $\mathbb{E}[T] = 0$ et $\mathrm{Var}(T) = 5-4\rho$. Au finale : $H \sim \mathcal{N}(1, 5-4\rho)$ et $T \sim \mathcal{N}(0, 5-4\rho)$.

8. On a d'après la question 6. le couple (H,T) est gaussien ; il est donc indépendant s'il est décorrélé, i.e, $\mathrm{Cov}(H,T) = 5-4\rho = 0$. La CNS pour que le couple (H,T) soit indépendant est donc $\rho = \frac{4}{5}$.

Exercice 3. Soient X et Y deux v.a i.i.d suivant la loi normale $\mathcal{N}(0,1)$ et soient les v.a $U = X+Y$ et $V = X-Y$.

1. Montrer que (U,V) est un couple Gaussien.
2. Montrer que les v.a U et V sont indépendantes.

Solution.

1. — *Première méthode.* Comme X et Y sont deux variables aléatoires réelles indépendantes et identiquement distribuées suivant une loi normale, le vecteur (X,Y) est gaussien. Pour tout $(a,b) \in \mathbb{R}^2$, on a

$$aU + bV = (a+b)X + (a-b)Y.$$

   Comme (X,Y) est un vecteur gaussien, $aU + bV = (a+b)X + (a-b)Y$ suit une loi normale. Par conséquent, (U,V) est aussi un vecteur gaussien.

   — *Deuxième méthode.* On peut écrire $\mathbf{L} = (X+Y, X-Y)$ sous la forme

$$\mathbf{L}^t = \mathbf{A}\mathbf{M}^t,$$

   où $\mathbf{M} = (X,Y)$ et

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Comme $\mathbf{M} = (X, Y)$ est un vecteur gaussien, $\mathbf{L} = (X+Y, X-Y)$ est aussi un vecteur gaussien.

2. Comme (U, V) est un vecteur gaussien, U et V sont indépendantes si, et seulement si, $\mathbb{C}(U, V) = 0$. Or, en utilisant le fait que X et Y suivent la loi normale centrée réduite, on a

$$\begin{aligned} \mathbb{C}(U, V) &= \mathbb{E}(UV) - \mathbb{E}(U)\mathbb{E}(V) = \mathbb{E}(X^2 - Y^2) - 0 \times 0 \\ &= \mathbb{E}(X^2) - \mathbb{E}(Y^2) = 1 - 1 = 0. \end{aligned}$$

Donc U et V sont indépendantes.

Exercice 4. Soit $V = (X, Y)^t$ un vecteur gaussien centré tel que $\mathbb{E}\left(X^2\right) = 4$ et $\mathbb{E}\left(Y^2\right) = 1$, et les v.a $2X + Y$ et $X - 3Y$ sont indépendantes.

1. Déterminer la matrice de covariance de V.
2. Montrer que le vecteur $W = (X + Y, 2X - Y)^t$ est gaussien.
3. Déterminer sa matrice de covariance.

Solution.

1. Comme (X, Y) un vecteur gaussien centré tel que $\mathbb{E}\left(X^2\right) = 4$ et $\mathbb{E}\left(Y^2\right) = 1$, on a

$$\mathbb{V}(X) = \mathbb{E}\left(X^2\right) - \left(\mathbb{E}\left(X\right)\right)^2 = \mathbb{E}\left(X^2\right) - 0^2 = 4$$

et

$$\mathbb{V}(Y) = \mathbb{E}\left(Y^2\right) - \left(\mathbb{E}\left(Y\right)\right)^2 = \mathbb{E}\left(Y^2\right) - 0^2 = 1.$$

De plus, comme $2X + Y$ et $X - 3Y$ sont indépendantes, on a

$$\mathbb{C}(2X + Y, X - 3Y) = 0.$$

Or,

$$\begin{aligned} \mathbb{C}(2X + Y, X - 3Y) &= \mathbb{C}(2X, X) + \mathbb{C}(2X, -3Y) + \mathbb{C}(Y, X) + \mathbb{C}(Y, -3Y) \\ &= 2\mathbb{C}(X, X) + (-2 \times 3)\mathbb{C}(X, Y) + \mathbb{C}(Y, X) - 3\mathbb{C}(Y, Y) \\ &= 2\mathbb{V}(X) - 6\mathbb{C}(X, Y) + \mathbb{C}(X, Y) - 3\mathbb{V}(Y) \\ &= 2 \times 4 - 5\mathbb{C}(X, Y) - 3 \times 1 \\ &= 5(1 - \mathbb{C}(X, Y)). \end{aligned}$$

Donc $\mathbb{C}(2X + Y, X - 3Y) = 0$ entraîne

$$\mathbb{C}(X, Y) = 1.$$

La matrice de covariance de (X, Y) est

$$\mathbf{V} = \begin{pmatrix} \mathbb{V}(X) & \mathbb{C}(X, Y) \\ \mathbb{C}(Y, X) & \mathbb{V}(Y) \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}.$$

2. — *Première méthode.* Pour tout $(a,b) \in \mathbb{R}^2$, on a

$$a(X+Y) + b(2X-Y) = (a+2b)X + (a-b)Y.$$

Comme (X,Y) est un vecteur gaussien, $a(X+Y)+b(2X-Y) = (a+2b)X+(a-b)Y$ suit une loi normale. Par conséquent, $(X+Y, 2X-Y)$ est un vecteur gaussien. En utilisant le résultat de la question 1-, il vient

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\mathbb{C}(X,Y) = 4+1+2\times 1 = 7.$$

De même

$$\mathbb{V}(2X-Y) = 4\mathbb{V}(X) + \mathbb{V}(Y) - 4\mathbb{C}(X,Y) = 16+1-4\times 1 = 13.$$

On a également

$$\begin{aligned}
\mathbb{C}(X+Y, 2X-Y) &= \mathbb{E}\left((X+Y)(2X-Y)\right) - \mathbb{E}\left(X+Y\right)\mathbb{E}\left(2X-Y\right) \\
&= \mathbb{E}(2X^2 - XY + 2XY - Y^2) \\
&= 2\mathbb{E}\left(X^2\right) + \mathbb{E}\left(XY\right) - \mathbb{E}\left(Y^2\right) \\
&= 8+1-1 = 8.
\end{aligned}$$

En combinant ces résultats, on obtient la matrice de covariance de $(X+Y, 2X-Y)$:

$$\mathbf{W} = \begin{pmatrix} 7 & 8 \\ 8 & 13 \end{pmatrix}.$$

— *Deuxième méthode.* On peut écrire $\mathbf{L} = (X+Y, 2X-Y)$ sous la forme

$$\mathbf{L}^t = \mathbf{A}\mathbf{M}^t,$$

où $\mathbf{M} = (X,Y)$ et

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}.$$

Comme $\mathbf{M} = (X,Y)$ est un vecteur gaussien, $\mathbf{L} = (X+Y, 2X-Y)$ est aussi un vecteur gaussien. La matrice de covariance de $\mathbf{L}$ est donnée par la formule

$$\mathbf{W} = \mathbf{A}\mathbf{V}\mathbf{A}^t,$$

où $\mathbf{V}$ désigne la matrice de covariance du vecteur $\mathbf{M} = (X,Y)$ obtenue au résultat de la question 1-. Il vient

$$\mathbf{W} = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 7 & 8 \\ 8 & 13 \end{pmatrix}.$$

Exercice 5. Soient X_1 et X_2 deux v.a i.i.d suivant chacune la loi normale $\mathcal{N}(0,1)$. Soit $\mathbf{Y} = (Y_1, Y_2)^t$ le vecteur aléatoire tel que

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{B},$$

où $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}$, $\mathbf{X} = (X_1, X_2)^t$ et $\mathbf{B} = (2,3)^t$.

1. Déterminer la loi de $\mathbf{Y}$.
2. Déterminer la loi de $\mathbf{Z} = (Y_1 + Y_2 + 1, 3Y_1 - Y_2)^t$.
3. Déterminer la loi de $Y_1 + Y_2 + 1$, et la loi de $3Y_1 - Y_2$.

Solution.

1. Comme $\mathbf{X} = (X_1, X_2)$ est un vecteur gaussien et $\mathbf{Y}$ est de la forme $\mathbf{Y}^t = \mathbf{A}\mathbf{X}^t + \mathbf{M}^t$, alors $\mathbf{Y}$ est aussi un vecteur gaussien. Comme $\mathbf{X} = (X_1, X_2)$ est centré, on a

$$\mathbb{E}\left(\mathbf{Y}^t\right) = \mathbb{E}\left(\mathbf{A}\mathbf{X}^t + \mathbf{M}^t\right) = \mathbf{A}\mathbb{E}\left(\mathbf{X}^t\right) + \mathbf{M}^t = 0 + \mathbf{M}^t = \mathbf{M}^t.$$

La matrice de covariance de $\mathbf{X} = (X_1, X_2)$ est $\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Par conséquent, la matrice de covariance de $\mathbf{Y}$ est

$$\mathbf{V} = \mathbf{A}\mathbf{I}\mathbf{A}^t = \begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}.$$

On en déduit que $\mathbf{Y} = (Y_1, Y_2)$ est un vecteur gaussien de moyenne $\mathbf{M} = (2, 3)$ et de matrice de covariance $\mathbf{V} = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}$.

2. En posant $\mathbf{Z} = (Y_1 + Y_2 + 1, 3Y_1 - Y_2)$, on peut écrire

$$\mathbf{Z} = \mathbf{B}\mathbf{Y}^t + \mathbf{N}^t,$$

où

$$\mathbf{B} = \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix},$$

$\mathbf{Y} = (Y_1, Y_2)$ et $\mathbf{N} = (1, 0)$. En utilisant le résultat de la question 1-, il vient

$$\begin{aligned} \mathbb{E}\left(\mathbf{Z}^t\right) &= \mathbb{E}\left(\mathbf{B}\mathbf{Y}^t + \mathbf{N}^t\right) = \mathbf{B}\mathbb{E}\left(\mathbf{Y}^t\right) + \mathbf{N}^t = \mathbf{B}\mathbf{M}^t + \mathbf{N}^t \\ &= \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 6 \\ 3 \end{pmatrix}. \end{aligned}$$

La matrice de covariance de $\mathbf{Y} = (Y_1, Y_2)$ est $\mathbf{V} = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}$. Par conséquent, la matrice de covariance de $\mathbf{Z}$ est

$$\mathbf{W} = \mathbf{B}\mathbf{V}\mathbf{B}^t = \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 9 & 3 \\ 3 & 17 \end{pmatrix}.$$

On en déduit que $\mathbf{Z} = (Y_1 + Y_2 + 1, 3Y_1 - Y_2)$ est un vecteur gaussien de moyenne $\mathbf{N} = (6, 3)$ et de matrice de covariance $\mathbf{W} = \begin{pmatrix} 9 & 3 \\ 3 & 17 \end{pmatrix}$.

3. Le résultat de la question 2- nous assure que $Y_1 + Y_2 + 1$ suit la loi normale de moyenne 6 et de variance 9, et $3Y_1 - Y_2$ suit la loi normale de moyenne 3 et de variance 17.

Exercice 6. Une densité de probabilité (ou loi de probabilité dans le cas discret) d'une v.a réelle X, notée $f(x;\theta)$ appartient à la *famille exponentielle* si $f(x;\theta)$ peut s'écrire sous la forme

$$f(x;\theta) = \exp[\eta(\theta)T(x) + a(\theta) + b(x)]\cdot \tag{3}$$

où les fonctions η, T, a et b sont à valeurs dans $\mathbb{R}$. La statistique $T(x)$ est appelée statistique exhaustive naturelle et le paramètre $\eta = \eta(\theta)$ est appelé paramètre naturel. Soit X une variable aléatoire de densité $f(x;\theta)$ appartenant à la famille exponentielle.
Déterminer une statistique exhaustive pour le paramètre θ pour un n-échantillon i.i.d $(X_1,\dots,X_n)$ de v.a de densité $f(x;\theta)$.

Solution. La densité jointe pour les observations *i.i.d* $\mathbf{x} = (x_1,\dots,x_n)$ est donnée par :

$$\begin{aligned} f(\mathbf{x};\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\ &= \prod_{i=1}^{n} \exp[\eta(\theta)T(x) + a(\theta) + b(x)] \\ &= \exp\left[\sum_{i=1}^{n} (\eta(\theta)T(x_i) + a(\theta) + b(x_i))\right] \\ &= \underbrace{\exp\left[\sum_{i=1}^{n} \eta(\theta)T(x_i) + n\,a(\theta)\right]}_{u(T(x_1,\dots,x_n);\theta)} \underbrace{\exp\left[\sum_{i=1}^{n} b(x_i)\right]}_{v(x_1,\dots,x_n)} \end{aligned}$$

A partir de ce résultat, et d'après le théorème de factorisation (critère de Fisher-Neyman), on a : $\sum_{i=1}^{n} T(x_i)$ est une statistique exhaustive pour θ.

Exercice 7. Soit X une v.a suivant une loi Poisson de paramètre $\theta > 0$, i.e, $\forall x \in \mathbb{N}$:

$$p_X(x;\theta) = \mathbb{P}(X = x) = e^{-\theta}\frac{\theta^x}{x!}\cdot$$

On considère un n-échantillon $(X_1,\dots,X_n)$ généré suivant une loi de Poisson de paramètre θ, $p_X(x;\theta)$

1. Déterminer une statistique exhaustive pour θ.
2. Calculer l'estimateur du maximum de vraisemblance (EMV) de θ.
3. En déduire son expression en fonction de la statistique exhaustive calculée dans la question précédente.

Solution. La loi jointe pour l'échantillon i.i.d $(x_1,\dots,x_n)$ est donnée par :

$$\begin{aligned} p(x_1,\dots,x_n;\theta) &= \prod_{i=1}^{n} p(x_i;\theta) \\ &= \prod_{i=1}^{n} [e^{-\theta}\frac{\theta^{x_i}}{x_i!}] \\ &= \theta^{\sum_{i=1}^{n} x_i} e^{-n\theta} \frac{1}{\prod_{i=1}^{n} x_i!} \\ &= \underbrace{\theta^{\sum_{i=1}^{n} x_i} e^{-n\theta}}_{u(T(x_1,\dots,x_n);\theta)} \underbrace{\frac{1}{\prod_{i=1}^{n} x_i!}}_{v(x_1,\dots,x_n)} \end{aligned}$$

A partir de ce résultat, et d'après le théorème de factorisation (critère de Fisher-Neyman), on a : $\sum_{i=1}^{n} X_i$ est une statistique exhaustive pour θ. La vraisemblance s'écrit :

$$\begin{aligned}
L(\theta; x_1, \ldots, x_n) &= p(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} p(x_i; \theta) && (4)\\
&= \prod_{i=1}^{n} e^{-\theta} \frac{\theta^{x_i}}{x_i!} && (5)\\
&= e^{-\theta n} \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} && (6)
\end{aligned}$$

La log-vraisemblance est donc donnée par

$$\begin{aligned}
\ln L(\theta; x_1, \ldots, x_n) &= \ln e^{-\theta n} \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} && (7)\\
&= \ln e^{-\theta n} + \ln \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} && (8)\\
&= -\theta n + \sum_{i=1}^{n} \ln \frac{\theta^{x_i}}{x_i!} && (9)\\
&= -\theta n + \ln \theta \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i!) && (10)\\
& && (11)
\end{aligned}$$

La dérivée première s'annule quand :

$$\begin{aligned}
\frac{\partial \ln L(\theta; x_1, \ldots, x_n)}{\partial \theta} &= 0 && (12)\\
-n + \frac{\sum_{i=1}^{n} x_i}{\theta} &= 0 && (13)
\end{aligned}$$

donc

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

La dérivée seconde s'écrit :

$$\frac{\partial^2 \ln L(\theta; x_1, \ldots, x_n)}{\partial \theta^2} = -\frac{\sum_{i=1}^{n} x_i}{\theta^2} \leq 0$$

car $x_i \in \mathbb{N}$. L'estimateur est donné par :

$$\Lambda_n = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

Remarque : on peut remarquer que l'estimateur s'exprime à l'aide de la statistique exhaustive pour λ : $\Lambda = \frac{T}{n}$

Exercice 8. *Risque quadratique et décomposition biais-variance* Soit θ le paramètre d'une loi de probabilité et soit $\widehat{\Theta}_n$ un estimateur de ce paramètre que l'on cherche à construire à partir d'un n-échantillon de v.a. i.i.d selon cette loi.

1. On note par $b(\widehat{\Theta}_n, \theta)$ le biais de $\widehat{\Theta}_n$ comme estimateur de θ. Donner l'expression de $b(\widehat{\Theta}_n, \theta)$.

2. On appelle risque quadratique de $\widehat{\Theta}_n$ comme estimateur de θ la quantité définie par : $\ell(\widehat{\Theta}_n, \theta) = \mathbb{E}\big[(\widehat{\Theta}_n - \theta)^2\big]$. Montrer que ce risque peut être décomposé selon la "décomposition biais-variance" suivante :

$$\ell(\widehat{\Theta}_n, \theta) \quad = \quad \mathrm{Var}(\widehat{\Theta}_n) + (b(\widehat{\Theta}_n, \theta))^2.$$

On cherche donc à trouver l'estimateur qui minimise ce risque.

3. Quel est, parmi tous les estimateurs sans biais de θ, celui que l'on doit choisir ?

4. Qu'appelle-t-on un tel estimateur ?

On considère un n-échantillon $(X_1, \cdots, X_n)$ i.i.d selon la loi de Bernoulli $\mathcal{B}(\theta)$ où $\theta = \mathbb{P}(X_i = 1)$ est le paramètre à estimer. On considère les deux estimateurs suivants pour θ : $\widehat{\Theta}_1 = X_1$ et $\widehat{\Theta}_n = \frac{\sum_{i=1}^n X_i}{n}$.

1. Quel est, au sens du risque quadratique, le meilleur estimateur parmi les deux ?

2. En déduire un estimateur convergent de θ

Solution.

1. Un estimateur $\widehat{\Theta}_n$ est dit sans biais pour θ lorsque son espérance mathématique est égale à la valeur du paramètre θ : $\mathbb{E}[\widehat{\Theta}_n] = \theta$. Le biais est donc donné par : $b(\widehat{\Theta}_n, \theta) = \mathbb{E}[\widehat{\Theta}_n] - \theta$

2. On a

$$\begin{aligned}
\mathbb{E}\big[(\widehat{\Theta}_n - \theta)^2\big] &= \mathbb{E}\left[\widehat{\Theta}_n^2 - 2\widehat{\Theta}_n\theta + \theta^2\right] \\
&= \mathbb{E}[\widehat{\Theta}_n^2] - 2\mathbb{E}[\widehat{\Theta}_n]\theta + \theta^2 \\
&= \mathbb{E}[\widehat{\Theta}_n^2] - \mathbb{E}[\widehat{\Theta}_n]^2 + \mathbb{E}[\widehat{\Theta}_n]^2 - 2\mathbb{E}[\widehat{\Theta}_n]\theta + \theta^2 \\
&= \mathbb{E}[\widehat{\Theta}_n^2] - \mathbb{E}[\widehat{\Theta}_n]^2 + \Big(\mathbb{E}[\widehat{\Theta}_n] - \theta\Big)^2 \\
&= \mathrm{Var}(\widehat{\Theta}_n) + (b(\widehat{\Theta}_n, \theta))^2
\end{aligned}$$

donc $\ell(\widehat{\Theta}_n, \theta) = \mathrm{Var}(\widehat{\Theta}_n) + (b(\widehat{\Theta}_n, \theta))^2$.

3. A partir de la décomposition biais-variance du risque quadratique, on voit que le meilleur estimateur sans biais à choisir est celui dont la variance est minimale (qui atteint la borne de Cramér-Rao).

4. Un estimateur sans-biais à variance minimale est un estimateur efficace.

5. On a les v.a $(X_i)_{1 \leqslant i \leqslant n}$ sont i.i.d selon la loi $\mathcal{B}(\theta)$ donc pour tout $1 \leqslant i \leqslant n$, $\mathbb{E}[X_i] = \theta$ et $\mathrm{Var}(X_i) = \theta(1-\theta)$. L'espérance de l'estimateur $\widehat{\Theta}_1$ est donc $\mathbb{E}[\widehat{\Theta}_1] = \mathbb{E}[X_1] = \theta$. Il est donc sans biais. Celle de $\widehat{\Theta}_n$ est $\mathbb{E}[\widehat{\Theta}_n] = \mathbb{E}\left[\frac{\sum_{i=1}^n X_i}{n}\right] = \theta$. Il est aussi sans biais.
Par contre, la variance de $\widehat{\Theta}_1$ est $\mathrm{Var}(\widehat{\Theta}_1) = \mathrm{Var}(X_1) = \theta(1-\theta)$ et celle de $\widehat{\Theta}_n = \mathrm{Var}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{\theta(1-\theta)}{n}$, qui est n fois plus petite que la variance de l'estimateur $\widehat{\Theta}_1$. La valeur du risque quadratique est aussi n fois plus petite pour l'estimateur $\widehat{\Theta}_n$ qui est donc le meilleur estimateur parmi les deux.

6. On a d'après la question précédent $\widehat{\Theta}_n$ est un estimateur sans biais pour θ et $\lim\limits_{n\to\infty} \mathrm{Var}\left(\widehat{\Theta}_n\right) = \lim\limits_{n\to\infty} \frac{\theta(1-\theta)}{n} = 0$. On en déduit qu'il est un estimateur convergent de θ.

Exercice 9. *Estimateurs et Borne Inférieure de Cramér-Rao (CRLB)* Soit X une v.a. gaussienne univariée de densité :

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

On considère un n-échantillon i.i.d $(X_1,\ldots,X_n)$ de v.a de densité $f(.;\mu,\sigma^2)$

1. Calculer la borne inférieure de Cramér-Rao pour un estimateur sans biais de l'espérance μ (variance σ^2 connue)
2. Calculer l'estimateur du maximum de vraisemblance de l'espérance μ. Que peut-on remarquer ?
3. Montrer qu'il est sans biais
4. En déduire qu'il est efficace
5. En déduire qu'il est convergent
6. Calculer la la borne inférieure de Cramér-Rao pour un estimateur sans biais de la variance σ^2 (espérance μ connu)
7. Calculer l'estimateur du maximum de vraisemblance de la variance σ^2 qu'on notera S^2.
8. On suppose que μ est connu. Montrer que que S^2 est efficace et consistent pour σ^2.
9. Maintenant mn suppose que μ est inconnu. En remplaçant dans l'expression de S^2 l'espérance μ par celle de son estimateur calcuée en .2, que peut-on dire sur le biais et l'efficacité de S^2 ?
10. Montrer que la variance de l'estimateur S^2 est $\frac{2\sigma^4}{n-1}$.
11. En déduire l'efficacité de cet estimateur.
12. Que se passe-t-il si on considère $\frac{n}{n-1}S^2$ comme estimateur de la variance au lieu de S^2 ?

Solution.

1. On a

$$\begin{aligned}
\ln f(x;\mu) &= \ln \frac{1}{\sigma\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} && (14)\\
&= \ln\left[\frac{1}{\sigma\sqrt{2\,\pi}}\right] - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2. && (15)
\end{aligned}$$

et

$$\begin{aligned}
\frac{\partial \ln f(x;\mu)}{\partial\mu} &= \frac{1}{\sigma^2}\left(\frac{x-\mu}{\sigma}\right) && (16)\\
\left(\frac{\partial \ln f(x;\mu)}{\partial\mu}\right)^2 &= \frac{1}{\sigma^4}\left(\frac{x-\mu}{\sigma}\right)^2 && (17)
\end{aligned}$$

Ainsi

$$\mathbb{E}\left[\left(\frac{\partial \ln f(x;\mu)}{\partial\mu}\right)^2\right] = \frac{1}{\sigma^4}\mathbb{E}\left[(x-\mu)^2\right] = \frac{1}{\sigma^4}\sigma^2 = \frac{1}{\sigma^2} \qquad (18)$$

La CRLB pour l'estimateur de μ est donc

$$\text{CRLB} = \frac{1}{n\mathbb{E}\left[\left(\frac{\partial \ln f(x;\mu)}{\partial\mu}\right)^2\right]} = \frac{\sigma^2}{n} \qquad (19)$$

2. on a $\mathbb{E}[\bar{X}] = \mathbb{E}[\frac{1}{n}\sum_{i=1}^n X_i] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \mu$. L'estimateur $\bar{X}$ est donc sans biais.
3. Ensuite, on a $\mathrm{Var}(\bar{X}) = \mathrm{Var}(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$
   Puisque la variance de $\bar{X}$ est $\frac{\sigma^2}{n}$, il est donc à variance minimale pour μ quand la v.a. X est a densité normale $\mathcal{N}(\mu, \sigma^2)$
4. Soit $\theta = \sigma^2$, on a

$$\begin{aligned} \ln f(x;\theta) &= \ln \frac{1}{\sqrt{\theta 2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}\frac{(x-\mu)^2}{\theta}} && (20)\\ &= -\frac{1}{2}\ln 2\pi\theta - \frac{1}{2}\frac{(x-\mu)^2}{\theta}. && (21)\end{aligned}$$

   et

$$\begin{aligned} \frac{\partial \ln f(x;\theta)}{\partial\theta} &= -\frac{1}{2\theta} + \frac{(x-\mu)^2}{2\theta^2} && (22)\\ \frac{\partial^2 \ln f(x;\theta)}{\partial^2\theta} &= \frac{1}{2\theta^2} - \frac{(x-\mu)^2}{\theta^3} && (23)\\ & && (24)\end{aligned}$$

   Ainsi

$$\mathbb{E}\left[\frac{\partial^2 \ln f(x;\theta)}{\partial^2\theta}\right] = \frac{1}{2\theta^2} - \frac{\mathbb{E}\left[(x-\mu)^2\right]}{\theta^3} = \frac{1}{2\theta^2} - \frac{\theta}{\theta^3} = -\frac{1}{2\theta^2} = -\frac{1}{2\sigma^4} \quad (25)$$

   La CRLB pour l'estimateur de σ^2 est donc donnée par [1]

$$\mathrm{CRLB} = -\frac{1}{n\mathbb{E}\left[\frac{\partial^2 \ln f(x;\mu)}{\partial^2\theta}\right]} = \frac{2\sigma^4}{n} \quad (26)$$

5. On a $\mathrm{Var}(S^2) = \frac{2\sigma^4}{n}$ donc l'efficacité de S^2 est donnée par :

$$e(S^2) = \frac{\mathrm{CRLB}}{\mathrm{Var}(S^2)} = \frac{2\sigma^4/n}{2\sigma^4/(n-1)} = \frac{n-1}{n}$$

   S^2 n'est donc pas un estimateur efficace pour σ^2 (car $e(S^2) < 1$). Cependant, S^2 est *asymptotiquement* efficace car

$$\lim_{n\to+\infty} e(S^2) = 1$$

The likelehood to be maximized is given as the joint probability density function for sample $(x_1, \ldots, x_n)$ of n independent identically distributed normal random variables

$$f(x_1,\ldots,x_n;\mu,\sigma^2) = \prod_{i=1}^n f(x_i;\mu,\sigma^2) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}. \quad (27)$$

Maximizing this likelihood is equivalent to maximzing the following log-likehood function

$$\begin{aligned} \mathcal{L}(\mu,\sigma^2) &= \log f(x_1,\ldots,x_n;\mu,\sigma^2) = \log \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}\\ &= \sum_{i=1}^n \log\frac{1}{\sigma\sqrt{2\,\pi}} + \sum_{i=1}^n \log \mathrm{e}^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}\\ &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2. && (28)\end{aligned}$$

1. ici on a utilisé la deuxième forme de la CRLB

Lets first start by maximzing (28) with respect to μ. We have to set its partial derivative w.r.t μ equat to zero. By taking into account the fact that in (28), only the quantity $-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$ depends on μ, we can therefore write

$$\frac{\partial\mathcal{L}(\mu,\sigma^2)}{\partial\mu} = \frac{\partial - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}{\partial\mu} = 0, \tag{29}$$

that is

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}2(x_i-\mu)\frac{\partial(x_i-\mu)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0. \tag{30}$$

Finally we have $\sum_{i=1}^{n}(x_i-\mu) = \sum_{i=1}^{n}x_i - n\mu = 0$ which results in the ML estimate $\hat{\mu}$ of μ

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

that is the sample mean!.

We can see that the ML estimator of μ given by

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n}X_i$$

is unbiased. Indeed, we have

$$\mathbb{E}[\widehat{\mu}] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}n\mu = \mu$$

where we have used the fact that $\mathbb{E}[X_i] = \mu$ which is the true mean.

To estimate σ^2, we similarly differentiate the log-likelihood (28) with respect to σ and equate to zero. Since in (28) only the quantity $-\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$ depends on σ, we have

$$\frac{\partial\mathcal{L}(\mu,\sigma^2)}{\partial\sigma} = \frac{\partial\left(-\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)}{\partial\sigma} = 0, \tag{31}$$

that is

$$-\frac{n}{2\sigma^2} - \frac{-1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \tag{32}$$

whiche gives

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = n. \tag{33}$$

Finally we therefore have

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \tag{34}$$

and by replacing μ by its ML estimate we get the ML estimate $\hat{\sigma}^2$ for σ^2

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 \tag{35}$$

which is the empirical variance!

However, in contrast to the ML estimator for the mean which is unbiased, the one for the variance, as we can see it is biased.

To calculate the expectation of the ML estimator $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\widehat{\mu})^2$, by using the fact that

$$\left(\sum_{i=1}^{n}X_i\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}X_iX_j$$

we first write

$$
\begin{aligned}
\widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\widehat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i-\frac{1}{n}\sum_{j=1}^{n}X_j\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2+\left(\frac{1}{n}\sum_{j=1}^{n}X_j\right)^2-\frac{2}{n}X_i\sum_{j=1}X_j\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2+\frac{1}{n^2}\sum_{j=1}^{n}\sum_{h=1}^{n}X_jX_h-\frac{2}{n}\sum_{j=1}X_iX_j\right),
\end{aligned}
\tag{36}
$$

the expectation of $\widehat{\sigma}^2$ is then given by

$$
\mathbb{E}[\widehat{\sigma}^2] = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{E}[X_i^2]+\frac{1}{n^2}\sum_{j=1}^{n}\sum_{h=1}^{n}\mathbb{E}[X_jX_h]-\frac{2}{n}\sum_{j=1}\mathbb{E}[X_iX_j]\right).
$$

Since the variables are mutually independent, we have, for $i \neq j$, $\mathbb{E}[X_iX_j] = \mathbb{E}[X_i][X_j] = \mu^2$. When $i = j$, we have $\mathbb{E}[X_iX_i] = \mathbb{E}[X_i^2] = \sigma^2+\mu^2$ from the variance formula. The previous equation is therefore wrerritten as

$$
\begin{aligned}
\mathbb{E}[\widehat{\sigma}^2] &= \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2+\mu^2+\frac{1}{n^2}\left(n\left((n-1)\mu^2+\left(\sigma^2+\mu^2\right)\right)\right)-\frac{2}{n}\left((n-1)\mu^2+\left(\sigma^2+\mu^2\right)\right)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2+\mu^2+\frac{1}{n^2}\left(n^2\mu^2+n\sigma^2\right)-\frac{2}{n}\left(n\mu^2+\sigma^2\right)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2+\mu^2+\mu^2+\frac{1}{n}\sigma^2-2\mu^2-\frac{2}{n}\sigma^2\right) \\
&= \frac{1}{n}(n\sigma^2-\sigma^2) \\
&= \frac{n-1}{n}\sigma^2.
\end{aligned}
\tag{37}
$$

The ML estimator $\mathbb{E}[\widehat{\sigma}^2]$ for σ^2 is therefore unbiased since $\mathbb{E}[\widehat{\sigma}^2] \neq \sigma^2$. However it is asymptotically unbiased : $\lim_{n\to\infty}\mathbb{E}[\widehat{\sigma}^2] = \sigma^2$.

We can correct the bias by taking $\sigma^{\star 2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$ as an unbiased estimator of the variance, rather than the ML estimator $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$. It is indeed easy to check that $\mathbb{E}[\sigma^{\star 2}] = \sigma^2$.

⋆

Exercice 10. *Régression et maximum de vraisemblance vs moindres carrés* Considérons le modèle de régression linéaire simple

$$
Y_i = \beta_0+\beta_1X_i+E_i, \tag{38}
$$

les erreurs E_i sont supposés centrées (d'espérance nulle) et de variance σ^2 connue. On cherche à estimer les cœfficients de régression (β_0,β_1) à partir d'un échantillon *i.i.d.* $\{(x_1,y_1),\ldots,(x_n,y_n)\}$. L'objectif est de montrer que, dans le cas Gaussien, l'estimateur du maximum de vraisemblance est celui des moindres carrés. Pour ce faire

1. Donner l'expression de la fonction *somme des carrés des résidus (RSS)* minimisée par la méthode des moindres carrés. On notera cette fonction $RSS(\beta_0, \beta_1)$
2. On montre que sous le modèle (38) la densité conditionnelle des observations y_i est Gaussienne
$$f(y_i|x_i; \beta_0; \beta_1)$$
d'espérance conditionnelle $\mu = \mathbb{E}[Y_i|X_i]$ et de variance σ^2.

   Calculer cette espérance à partir de (38)
3. Décrire la fonction de *log-vraisemblance conditionnelle*[2] $L(\beta_0, \beta_1; y_1, \ldots, y_n|x_1, \ldots, x_n)$. On notera cette fonction $\ln L(\beta_0, \beta_1)$, où L représente la vraisemblance conditionnelle
4. En déduire que l'estimateur du maximum de vraisemblance est identique à celui des moindres carrés.

Solution.

Exercice 11. *Régression linéaire et moindres carrés* On rappelle le modèle de régression linéaire simple sous sa forme matricielle :

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{E} \tag{39}$$

où $\boldsymbol{Y}$ est le vecteur aléatoires des observations, $\mathbf{X}$ la matrice de design (matrice des prédicteurs), $\boldsymbol{\beta}$ le vecteur des cœfficients de régression et $\boldsymbol{E}$ le vecteur aléatoire représentant le bruit supposé centré (d'espérance nulle) et de matrice de covariance $\sigma^2\mathbf{I}$, $\mathbf{I}$ étant la matrice identité.

1. Donner l'expression de l'estimateur des moindres carrés (EMC) $\hat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$
2. Montrer qu'il est sans biais
3. Calculer sa matrice de covariance

Solution. On a
$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}$$

et

$$\begin{aligned}
\mathbb{E}[\hat{\boldsymbol{\beta}}] &= \mathbb{E}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{E}[\boldsymbol{Y}] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{E}[\mathbf{X}\boldsymbol{\beta} + \boldsymbol{E}] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \\
&= \boldsymbol{\beta} \qquad (40)
\end{aligned}$$

2. Le terme "conditionnelle" ici vient du fait que, comme on est dans un contexte de régression où la variable expliquée Y qui est conditionné par la variable explicative X, on a donc une densité conditionnelle $f(y|x; \beta_0, \beta_1)$ plutôt qu'une densité (non conditionnelle) $f(y; \boldsymbol{\beta}_0, \boldsymbol{\beta}_1)$ comme en estimation de densité classique.

L'EMC est donc sans biais.

On a également

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y} \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T[\mathbf{X}\boldsymbol{\beta} + \boldsymbol{E}] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{E} \\
&= \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{E} \qquad (41)
\end{aligned}$$

donc

$$\begin{aligned}
\mathrm{Cov}[\hat{\boldsymbol{\beta}}] &= \mathbb{E}\left[(\hat{\boldsymbol{\beta}} - \mathbb{E}[\hat{\boldsymbol{\beta}}])(\hat{\boldsymbol{\beta}} - \mathbb{E}[\hat{\boldsymbol{\beta}}])^T\right] \\
&= \mathbb{E}\left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\right] \\
&= \mathbb{E}\left[((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{E})((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{E})^T\right] \\
&= \mathbb{E}\left[((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{E}\boldsymbol{E}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\right] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{E}\left[\boldsymbol{E}\boldsymbol{E}^T\right]\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T[\sigma^2\mathbf{I}]\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \qquad (42)
\end{aligned}$$

On peut montrer que cette variance est la variance minimale. L'EMC étant sans biais. Il est donc efficace.

l'EMC est le meilleur estimateur linéaire en bsY et sans biais (BLUE)