Consignes :
- Sont interdits : Documents, calculettes, téléphones, écouteurs, ordinateurs, tablettes.
- Il est interdit de composer avec un crayon.
- Votre feuille double d'examen doit porter, à l'emplacement réservé, vos nom, prénom, et signature.
- Cette zone réservée doit être cachée par collage.
- Vos feuilles intercalaires doivent être toutes numérotées.
- Le barème (sur 22) est donné à titre indicatif.

Exercice 1 *(8 pts)* Une urne contient une boule blanche et une boule noire indiscernables au toucher. On y prélève une boule, on note sa couleur, et on la remet dans l'urne en rajoutant α ($\alpha \neq 0$) autres boules de la couleur de la boule tirée (toutes ces boules sont indiscernables au toucher). On répète cette épreuve n fois ($n > 2$). Soit les variables aléatoires $(X_i)_{1 \leqslant i \leqslant n}$ définies par $X_i = 1$ si on obtient une boule blanche au i-ème tirage, et $X_i = 0$ sinon.

1. Donner la loi de X_1 et $\mathbb{E}[X_1]$
2. Déterminer la loi du couple (X_1, X_2)
3. En déduire la loi de X_2 puis $\mathbb{E}[X_2]$. Que peut-on remarquer ?

On définit pour $2 \leqslant m \leqslant n-1$, la variable aléatoire $Y_m = \sum_{i=1}^{m} X_i$.

4. Que représente la variable Y_m ? Donner $Y_m(\Omega)$
5. Déterminer la loi de Y_2
6. Sachant qu'au cours des m premiers tirages on a tiré k boules blanches, quel est le nombre de boules dans l'urne avant le $(m+1)$-ième tirage ?
7. Déterminer pour tout $k \in Y_m(\Omega)$, $\mathbb{P}(X_{m+1} = 1 | Y_m = k)$
8. Montrer que $\mathbb{P}(X_{m+1} = 1) = \frac{1+\alpha\mathbb{E}[Y_m]}{2+\alpha m}$. *Indication : On pourra utiliser la formule des probabilités totales*

Solution 1

1. X_1 vaut 1 lorsque la boule tirée est blanche et 0 sinon. Don X_1 suit une loi de Bernoulli de paramètre $\mathbb{P}(X_1 = 1) = \frac{1}{2}$ (boules équiprobables). On a donc $\mathbb{E}[X_1] = \frac{1}{2}$
2. Loi du couple (X_1, X_2) :

$X_1 \setminus X_2$	0	1
0	$\frac{1}{2}\left(\frac{\alpha+1}{\alpha+2}\right)$	$\frac{1}{2}\left(\frac{1}{\alpha+2}\right)$
1	$\frac{1}{2}\left(\frac{1}{\alpha+2}\right)$	$\frac{1}{2}\left(\frac{\alpha+1}{\alpha+2}\right)$

On a $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 0) = \frac{1}{2}$.

Pour déterminer la loi du couple on utilise la règle de Bayes : $\mathbb{P}(X_1, X_2) = \mathbb{P}(X_2|X_1)\mathbb{P}(X_1)$

— Sachant que l'on a obtenu une boule noire au premier tirage ($\{X_1 = 0\}$), il y a avant le deuxième tirage $\alpha + 2$ boules dans l'urne dont $\alpha + 1$ qui sont noires et une qui est blanche, donc $\mathbb{P}(X_2 = 0|X_1 = 0) = \frac{\alpha+1}{\alpha+2}$ et $\mathbb{P}(X_2 = 1|X_1 = 0) = \frac{1}{\alpha+2}$. On en déduit $\mathbb{P}(X_1 = 0, X_2 = 0) = \mathbb{P}(X_2 = 0|X_1 = 0)\mathbb{P}(X_1 = 0)$ et $\mathbb{P}(X_1 = 0, X_2 = 1) = \mathbb{P}(X_2 = 1|X_1 = 0)\mathbb{P}(X_1 = 0)$.

— Sachant que l'on a obtenu une boule blanche au premier tirage ($\{X_1 = 1\}$), il y a avant le deuxième tirage $\alpha + 2$ boules dans l'urne dont une qui est noire est $1 + \alpha$ qui sont blanches, donc $\mathbb{P}(X_2 = 0|X_1 = 1) = \frac{1}{\alpha+2}$ et $\mathbb{P}(X_1 = 1, X_2 = 0) = \mathbb{P}(X_2 = 0|X_1 = 1)\mathbb{P}(X_1 = 1)$. On en déduit $\mathbb{P}(X_1 = 1, X_2 = 0) = \mathbb{P}(X_2 = 0|X_1 = 1)\mathbb{P}(X_1 = 1)$ et $\mathbb{P}(X_1 = 1, X_2 = 1) = \mathbb{P}(X_2 = 1|X_1 = 1)\mathbb{P}(X_1 = 1)$.

3. La loi marginale de X_2 s'obtient en marginalisant la loi jointe du couple par rapport à la variable X_1 et est définie par : $\mathbb{P}(X_2 = j) = \sum_{i=0}^{1} \mathbb{P}(X_1 = i, X_2 = j)$ pour $j \in \{0, 1\}$. On en déduit d'après la loi du couple que $\mathbb{P}(X_2 = 1) = \mathbb{P}(X_2 = 0) = \frac{1}{2}$. Il en suit que $\mathbb{E}[X_2] = \frac{1}{2}$. X_2 est donc de même loi $\mathcal{B}(\frac{1}{2})$ que X_1.
4. La variable $Y_m = \sum_{i=1}^{m} X_i$ représente le nombre de boules blanches tirées au cours des m premiers tirages. On a clairement $Y_m(\Omega) = [\![0, m]\!]$
5. La loi de Y_2 est définie par $\mathbb{P}(Y_2 = k) = \mathbb{P}(X_1 + X_2 = k)$ pour tout $k \in [\![0, 2]\!]$. D'après le tableau de la loi du couple (X_1, X_2) (question 2.) la loi de Y_2 est donc donnée par le tableau suivant :

k	0	1	2
$\mathbb{P}(Y_2 = k)$	$\frac{1}{2}\left(\frac{\alpha+1}{\alpha+2}\right)$	$\frac{1}{\alpha+2}$	$\frac{1}{2}\left(\frac{\alpha+1}{\alpha+2}\right)$

6. Si au cours des m premiers tirages on a tiré k boules blanches, on a alors tiré $m - k$ boules noires. On a donc rajouté dans l'urne αk boules blanches et $\alpha(m - k)$ boules noires. Il y a donc $2 + \alpha k + \alpha(m - k) = 2 + \alpha m$ boules dans l'urne avant le $(m + 1)$-ième tirage.
7. L'évènement $\{Y_m = k\}$ signifie qu'au cours des m premiers tirages on a tiré k boules blanches. On a donc avant avant le $(m + 1)$-ième tirage $1 + \alpha k$ boules blanches dans l'urne. D'après la question précédente on donc pour tout $k \in [\![0, m]\!]$, $\mathbb{P}(X_{m+1} = 1 | Y_m = k) = \frac{1+\alpha k}{2+\alpha m}$.
8. En utilisant la formule des probabilités totales sur le système complet d'évènements $\{(Y_m = k)\}_{k=0}^{m}$ on a

$$\begin{aligned}
\mathbb{P}(X_{m+1} = 1) &= \sum_{k=0}^{m} \mathbb{P}(X_{m+1} = 1 | Y_m = k)\mathbb{P}(Y_m = k) \\
&= \sum_{k=0}^{m} \frac{1 + \alpha k}{2 + \alpha m} \mathbb{P}(Y_m = k) \\
&= \frac{1}{2 + \alpha m} \left(\sum_{k=0}^{m} \mathbb{P}(Y_m = k) + \alpha \sum_{k=0}^{m} k\mathbb{P}(Y_m = k) \right) \\
&= \frac{1}{2 + \alpha m} \left(1 + \alpha \mathbb{E}[Y_m]\right).
\end{aligned}$$

Exercice 2 *(8 pts)* 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

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

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

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

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 2

1. Il est clair que X et Y ont la même loi $\mathcal{N}(0,1)$
2. La matrice de variance-covariance $\mathbf{\Sigma}$ du vecteur $\boldsymbol{V} = (X,Y)^t$ est donné par

$$\mathbf{\Sigma} = \mathrm{Cov}((X,Y)^t) = \begin{pmatrix} \mathrm{Var(X)} & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}(Y) \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

car $\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \mathrm{Cov}(X,Y)$.

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 :

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

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 *(6 pts)* 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 dispose d'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 $\widehat{\Theta}_n = \frac{\sum_{i=1}^n X_i}{n}$.

5. Quel est, au sens du risque quadratique, le meilleur estimateur parmi les deux ?
6. En déduire un estimateur convergent de θ

Solution 3

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