Université Paris I, Panthéon - Sorbonne

Première Année Licence M.A.S.H.S. 2021 – 2022

Probabilités

Examen final, Mai 2022

Examen de 2h00. Tout document ou calculatrice est interdit.

Exercice 1 (Sur 10 points)

On note X_0 une variable aléatoire définie sur un espace de probabilités $(\Omega, \mathcal{A}, \mathbb{P})$ qui suit une loi de Bernoulli de paramètre 1/2. On note également X_1 une variable aléatoire définie également sur $(\Omega, \mathcal{A}, \mathbb{P})$, qui ne prend que les valeurs 0 et 1 et telle que:

$$\mathbb{P}\big(X_1 = 1 \mid X_0 = 1\big) = p \quad \text{et} \quad \mathbb{P}\big(X_1 = 0 \mid X_0 = 0\big) = q,$$

où p et q sont deux réels fixés dans $]0,1[$.

1. Déterminer par le calcul $\mathbb{E}[X_0]$ puis $\mathbb{E}[X_0^k]$ pour $k \in \mathbf{N}^*$ **(1pt)**.
2. Déterminer en justifiant $\mathbb{P}\big(X_1 = 0 \mid X_0 = 1\big)$ et $\mathbb{P}\big(X_1 = 1 \mid X_0 = 0\big)$ **(1pt)**.
3. Démontrer que X_1 suit une loi de Bernoulli de paramètre $\frac{1}{2}(p+1-q)$ **(1pt)**.
4. A quelles conditions sur p et q les variables X_0 et X_1 sont-elles indépendantes? **(1.5pts)**.
5. Plus généralement, on considère la suite de variables aléatoires $(X_n)_{n\in\mathbf{N}}$ définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans $\{0,1\}$ et telles que pour tout $n \in \mathbf{N}$:

$$\mathbb{P}\big(X_{n+1} = 1 \mid X_n = 1\big) = p \quad \text{et} \quad \mathbb{P}\big(X_{n+1} = 0 \mid X_n = 0\big) = q.$$

Si on note p_n le paramètre de la loi de Bernoulli suivie par X_n, montrer que pour tout $n \in \mathbf{N}^*$, $p_{n+1} = p_n(p+q-1) + (1-q)$ pour $n \in \mathbf{N}^*$ **(1pt)**. Montrer par récurrence que

$$p_n = \frac{q-p}{2(2-p-q)}(p+q-1)^n + \frac{1-q}{2-p-q} \qquad \text{pour tout } n \in \mathbf{N} \textbf{ (3pts)}.$$

En déduire que (X_n) converge en loi vers une loi que l'on précisera **(1.5pts)**.

Proof. 1. On a $\mathbb{E}[X] = 1/2 * 1 + 1/2 * 0 = 1/2$. De même, $\mathbb{E}[X^k] = 1/2 * 1^k + 1/2 * 0^k = 1/2$ pour tout $k \in \mathbf{N}^*$.

2. On a par la formule des probabilités totales

$$\begin{aligned} &\mathbb{P}\big(X_1 = 0 \cap X_0 = 1\big) + \mathbb{P}\big(X_1 = 1 \cap X_0 = 1\big) = \mathbb{P}\big(X_0 = 1\big) \\ &\qquad \Longrightarrow \quad \mathbb{P}\big(X_1 = 0 \mid X_0 = 1\big)\,\mathbb{P}\big(X_0 = 1\big) + \mathbb{P}\big(X_1 = 1 \mid X_0 = 1\big)\,\mathbb{P}\big(X_0 = 1\big) = \mathbb{P}\big(X_0 = 1\big). \end{aligned}$$

D'où $\mathbb{P}\big(X_1 = 0 \mid X_0 = 1\big) + \mathbb{P}\big(X_1 = 1 \mid X_0 = 1\big) = 1$, soit $\mathbb{P}\big(X_1 = 0 \mid X_0 = 1\big) = 1 - p$. De même $\mathbb{P}\big(X_1 = 1 \mid X_0 = 0\big) = 1 - q$

3. X_1 ne prend pour valeurs que 0 et 1: elle suit donc une loi de Bernoulli. Son paramètre est donné par $\mathbb{P}(X_1 = 1)$, et par la formule des probabilités totales:

$$\begin{aligned} \mathbb{P}\big(X_1 = 1\big) &= \mathbb{P}\big(X_1 = 1 \cap X_0 = 1\big) + \mathbb{P}\big(X_1 = 1 \cap X_0 = 0\big) \\ &= \mathbb{P}\big(X_1 = 1 \mid X_0 = 1\big)\mathbb{P}\big(X_0 = 1\big) + \mathbb{P}\big(X_1 = 1 \mid X_0 = 0\big)\mathbb{P}\big(X_0 = 0\big) \\ &= \frac{1}{2}(p + 1 - q). \end{aligned}$$

4. On a X_1 et X_0 indépendantes si $\mathbb{P}\big(X_1 = 1 \mid X_0 = 1\big) = \mathbb{P}\big(X_1 = 1 \mid X_0 = 0\big) = \mathbb{P}\big(X_1 = 1\big)$. Cela implique que $\frac{1}{2}\big(p + 1 - q\big) = p = 1 - q$, ces 2 égalités se réduisant à la seule équation $p = 1 - q$.
5. On reprend le même raisonnement qu'en 3.: pour tout $n \in \mathbf{N}$, X_{n+1} ne prend pour valeurs que 0 et 1: elle suit donc une loi de Bernoulli. Son paramètre est donné par $\mathbb{P}(X_{n+1} = 1)$, et par la formule des probabilités totales:

$$\begin{aligned} \mathbb{P}\big(X_{n+1} = 1\big) &= \mathbb{P}\big(X_{n+1} = 1 \cap X_n = 1\big) + \mathbb{P}\big(X_{n+1} = 1 \cap X_n = 0\big) \\ &= \mathbb{P}\big(X_{n+1} = 1 \mid X_n = 1\big)\mathbb{P}\big(X_n = 1\big) + \mathbb{P}\big(X_{n+1} = 1 \mid X_n = 0\big)\mathbb{P}\big(X_n = 0\big) \\ &= p\,p_n + (1-q)\,(1-p_n) = (p+q-1)\,p_n + 1 - q. \end{aligned}$$

Montrons la propriété P_n : $p_n = \frac{q-p}{2(2-p-q)}\,(p+q-1)^n + \frac{1-q}{2-p-q}$ par récurrence.
Pour $n = 0$, on a $p_0 = \frac{q-p}{2(2-p-q)} + \frac{1-q}{2-p-q} = \frac{q-p+2-2q}{2(2-p-q)} = 1/2$ ce qui est vrai donc P_0 est vraie.
Supposons P_n vraie. Alors:

$$\begin{aligned} p_{n+1} &= (p+q-1)\,p_n + 1 - q \\ &= \frac{q-p}{2(2-p-q)}\,(p+q-1)^{n+1} + \frac{(1-q)(p+q-1)}{2-p-q} + 1 - q \\ &= \frac{q-p}{2(2-p-q)}\,(p+q-1)^{n+1} + (1-q)\Big(\frac{p+q-1}{2-p-q} + \frac{2-p-q}{2-p-q}\Big) \\ &= \frac{q-p}{2(2-p-q)}\,(p+q-1)^{n+1} + (1-q)\,\frac{1}{2-p-q}. \end{aligned}$$

Donc P_{n+1} est vraie et ainsi P_n est vraie pour tout $n \in \mathbf{N}$.
De ceci on en déduit que $p_n \underset{n\to\infty}{\longrightarrow} \dfrac{1-q}{2-p-q}$ et ainsi $X_n \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{B}\Big(\frac{1-q}{2-p-q}\Big)$.

□

Exercice 2 (Sur 4 points)

En 2010, le magazine Causette avait mené une enquête sur les liens entre les orientations politiques des femmes politiques et leur couleur de cheveux. En arrondissant les chiffres et en schématisant, on obtenait que 50% des politiciennes de droite (D) et d'extrême droite (ED) étaient blondes quand seules 15% des politiciennes de gauche (G) l'étaient. Et à l'époque, la droite et la gauche représentaient chacune 40% des politiciennes, l'extrême droite 20%. Une politicienne inconnue arrivait en 2010 vers vous, elle n'était pas blonde. Quelle est la probabilité qu'elle était de gauche?

Proof. Soit B l'événement "Etre blonde". D'après l'énoncé, $\mathbb{P}(B \,|\, D) = \mathbb{P}(B \,|\, ED) = 0.5$ et $\mathbb{P}(B \,|\, G) = 0.15$. De plus $\mathbb{P}(D) = \mathbb{P}(G) = 0.4$ alors que $\mathbb{P}(ED) = 0.2$.
On cherche $\mathbb{P}(G \,|\, \overline{B})$. Mais, en utilisant la formule des probabilités totales:

$$\mathbb{P}(G \,|\, \overline{B}) = \frac{\mathbb{P}(G \cap \overline{B})}{\mathbb{P}(\overline{B})} = \frac{\mathbb{P}(G) - \mathbb{P}(G \cap B)}{1 - \mathbb{P}(B)} = \frac{\mathbb{P}(G) - \mathbb{P}(B \,|\, G)\mathbb{P}(G)}{1 - \mathbb{P}(B)}.$$

Mais d'après la formule des probabilités totales,

$$\mathbb{P}(B) = \mathbb{P}(G \cap B) + \mathbb{P}(D \cap B) + \mathbb{P}(ED \cap B) = \mathbb{P}(B \,|\, G)\mathbb{P}(G) + \mathbb{P}(B \,|\, D)\mathbb{P}(D) + \mathbb{P}(B \,|\, ED)\mathbb{P}(ED).$$

On en déduit que $\mathbb{P}(B) = 0.15*0.4 + 0.5*0.4 + 0.5*0.2 = 0.36$ et par conséquent, $\mathbb{P}(G \,|\, \overline{B}) = (0.4 - 0.15*0.4)/(1 - 0.36) = 0.34/0.64 = 17/32$. □

Exercice 3 (Sur 14 points)

On s'intéresse au niveau de crue maximum (en mètres) atteint par une rivière chaque année. On modélise ce niveau par une variable aléatoire X de fonction de répartition

$$F_X(x) = \begin{cases} 1 - (1+x)^{-\alpha} & \text{si } 0 \le x \\ 0 & \text{sinon.} \end{cases}$$

où $\alpha > 2$ est un paramètre inconnu.

1. Montrer que X est variable aléatoire continue et déterminer sa densité de probabilité f_X **(1.5pts)**.
2. Démontrer que $\mathbb{E}[X] = \dfrac{1}{\alpha - 1}$ puis que $\mathrm{var}(X) = \dfrac{\alpha}{(\alpha-1)^2(\alpha-2)}$ **(2.5pts)**.
3. Afin de construire une digue de taille suffisante, on cherche à connaître le niveau $N > 0$ tel qu'il n'y ait exactement qu'une chance sur 100 d'avoir une crue supérieure à N. Comment appelle-t-on N pour la distribution de X α **(0.5pts)**? Déterminer N en fonction de α **(1pt)**.

4. Par la suite on veut déterminer un intervalle de confiance sur α pour en déduire un intervalle de confiance sur N. On a un historique des 400 dernières années de crue. Pour en tenir compte, on considère 400 variables aléatoires $X_i,\ i=1,\dots,400$ indépendantes et identiquement distribuées suivant la loi de densité f_X. On note $\overline{X}$ la moyenne empirique des $X_i,\ i=1,\dots,400$. Démontrer que
$$\mathbb{P}\Big(\overline{X}\geq \frac{100}{\alpha-1}\Big)\leq 0.01 \quad \textbf{(2.5pts)}.$$
En déduire que $\mathbb{P}\big(N\geq 100^{\frac{\overline{X}}{100+\overline{X}}}-1\big)\geq 0.99$ **(1pt)**. Qu'en déduire **(0.5pts)**?

5. On note $\overline{\sigma^2}=\dfrac{1}{400}\displaystyle\sum_{i=1}^{400}\big(X_i-\overline{X}\big)^2$. Donner la loi approximative de $\overline{X}$ en fonction de $\overline{\sigma^2}$ **(0.5pts)**.
En notant q un quantile (on précisera lequel?) de la loi $\mathcal{N}(0,1)$, en déduire un intervalle de confiance de $(\alpha-1)^{-1}$ de niveau 99% en fonction de $\overline{X}$ et de $\overline{\sigma^2}$ **(0.5pts)**, puis un intervalle de confiance de N de niveau 99% **(1pt)**.

6. On note par classes les différentes valeurs observées:

Hauteur maximale de crue en mètres	Effectifs
[0 , 1[	356
[1 , 2[	20
[2 , 3[	16
[3 , 4[	4
[4 , 5[	4

Tracer le polygône des fréquences cumulées de ces données **(0.5pts)**. En déduire une première approximation de N **(1pt)**. Déterminer une approximation de la moyenne empirique **(0.5pts)** et en déduire une seconde approximation de N **(0.5pts)**.

Proof. 1. La fonction F_X est continue en 0 (limite à gauche et valeur à droite $=0$), et donc sur $\mathbf{R}$. De plus, F_X est dérivable sur $\mathbf{R}^*$: X est donc une variable aléatoire continue de densité $f_X(x)=F_X'(x)$ pour $x\neq 0$, soit
$$f_X(x)=\begin{cases}\alpha\,(1+x)^{-\alpha-1} & \text{si } 0<x\\ 0 & \text{sinon.}\end{cases}$$

2. On a grâce à une intégration par parties:
$$\mathbb{E}[X]=\int_0^\infty \alpha\,\frac{x}{(1+x)^{\alpha+1}}\,dx=\Big[-\frac{x}{(x+1)^\alpha}\Big]_0^\infty+\Big[\frac{1}{1-\alpha}\,\frac{1}{(x+1)^{\alpha-1}}\Big]_0^\infty=\frac{1}{\alpha-1}.$$
De même,
$$\mathbb{E}[X^2]=\int_0^\infty \alpha\,\frac{x^2}{(1+x)^{\alpha+1}}\,dx=\Big[-\frac{x^2}{(x+1)^\alpha}\Big]_0^\infty+2\int_0^\infty\frac{x}{(x+1)^\alpha}\,dx=\frac{2}{1-\alpha}\Big(\Big[\frac{x}{(x+1)^{\alpha-1}}\Big]_0^\infty-\int_0^\infty\frac{dx}{(x+1)^{\alpha-1}}\Big).$$
soit $\mathbb{E}[X^2]=\dfrac{2}{(\alpha-1)(\alpha-2)}$. En conséquence, $\mathrm{var}(X)=\dfrac{2}{(\alpha-1)(\alpha-2)}-\dfrac{1}{(\alpha-1)^2}=\dfrac{\alpha}{(\alpha-1)^2(\alpha-2)}$.

3. On cherche N tel que $\mathbb{P}(X>N)=0.01$, soit $\mathbb{P}(X\leq N)=0.99$: N est le quantile à 99% de la loi de X. On a donc $1-F_X(N)=0.01$ soit $(1+N)^{-\alpha}=0.01$, d'où $(1+N)^\alpha=100$ donc $N=100^{1/\alpha}-1$.

4. Comme les X_i sont positives, alors $\overline{X}$ est positive et on peut appliquer l'inégalité de Markov, soit:
$$\mathbb{P}\big(\overline{X}\geq\varepsilon\big)\leq\frac{\mathbb{E}[\overline{X}]}{\varepsilon}\quad\text{pour tout } \varepsilon>0.$$
Or $\mathbb{E}[\overline{X}]=\mathbb{E}[X]=(\alpha-1)^{-1}$ car les X_i sont des vaiid. En choisissant $\varepsilon=100\,\mathbb{E}[\overline{X}]=100\,(\alpha-1)^{-1}$, on obtient le résultat.
On a $\overline{X}\geq\dfrac{100}{\alpha-1}\iff\alpha\geq\dfrac{100+\overline{X}}{\overline{X}}\iff\dfrac{1}{\alpha}\leq\dfrac{\overline{X}}{100+\overline{X}}\iff N=100^{\frac{1}{\alpha}}-1\leq 100^{\frac{\overline{X}}{100+\overline{X}}}-1$, d'où le résultat.
On a ainsi un intervalle de confiance à 99% sur N: $\Big[100^{\frac{\overline{X}}{100+\overline{X}}}-1,\infty\Big[$.

5. Comme les X_i sont des vaiid, d'après le second TLC, pour n grand (et ici $n=400$ donc grand) on a $\overline{X}\overset{\mathcal{L}}{\sim}\mathcal{N}\big(\frac{1}{\alpha-1},\frac{\overline{\sigma^2}}{400}\big)$.
On a donc d'après le cours: $\mathbb{P}\Big(\overline{X}-q\,\frac{\overline{\sigma}}{20}\leq\frac{1}{\alpha-1}\leq\overline{X}+q\,\frac{\overline{\sigma}}{20}\Big)\simeq 0.99$, où q est le quantile à 99.5% de la loi $\mathcal{N}(0,1)$.
Comme précédemment, on a:
$$\overline{X}-q\,\frac{\overline{\sigma}}{20}\leq\frac{1}{\alpha-1}\leq\overline{X}+q\,\frac{\overline{\sigma}}{20}\iff\frac{1}{\overline{X}+q\,\frac{\overline{\sigma}}{20}}+1\leq\alpha\leq\frac{1}{\overline{X}-q\,\frac{\overline{\sigma}}{20}}+1\iff\frac{\overline{X}-q\,\frac{\overline{\sigma}}{20}}{1+\overline{X}-q\,\frac{\overline{\sigma}}{20}}\leq\frac{1}{\alpha}\leq\frac{\overline{X}+q\,\frac{\overline{\sigma}}{20}}{1+\overline{X}+q\,\frac{\overline{\sigma}}{20}}$$

Ainsi, un intervalle 99% pour N est: $\left[100^{\frac{\overline{X}-q\,\frac{\overline{\sigma}}{20}}{1+\overline{X}-q\,\frac{\overline{\sigma}}{20}}}-1\,,\,100^{\frac{\overline{X}+q\,\frac{\overline{\sigma}}{20}}{1+\overline{X}+q\,\frac{\overline{\sigma}}{20}}}-1\right]$.

6. On trace le polygône des fréquences cumulées et comme N est le quantile d'ordre 99%, on obtient sur le dessin, ou par le calcul $N \simeq 4$.

   On obtient $\overline{X} \simeq 0.7$ et comme une approximation de N est $100^{\frac{\overline{X}}{1+\overline{X}}}-1$, on obtient $N \simeq 100^{7/17}-1 \simeq 5.66$.

$\square$