Université de Montpellier - Faculté des Sciences

Année Universitaire 2024-2025

HAX506X - Probabilités

Corrigé de l'examen terminal - 14/01/2025

Rappels. Les résultats suivants peuvent être utilisés sans preuve.

1. *Critère de Bertrand.* La série à termes positifs
$$\sum_{n\geq 2} \frac{1}{n^\alpha (\ln n)^\beta}$$
converge si et seulement si $\alpha > 1$ ou ($\alpha = 1$ et $\beta > 1$).
2. La fonction de répartition Φ d'une loi normale $N \sim \mathcal{N}(0,1)$ vérifie
$$1 - \Phi(x) = \mathbb{P}(N \geq x) = \int_x^\infty \frac{\mathrm{e}^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\,\mathrm{d}t \underset{x\to\infty}{\sim} \frac{\mathrm{e}^{-\frac{x^2}{2}}}{x\sqrt{2\pi}}\,.$$
Quelques valeurs : $\Phi(-1.96) = 0.025\,, \Phi(-1.64) = 0.05\,, \Phi(1.64) = 0.95\,, \Phi(1.96) = 0.975$.

Exercice 1 - Minimum de lois uniformes

Soit $U_1, \ldots, U_n$ une suite de variables aléatoires indépendantes de loi uniforme sur $[0,1]$. On pose $Y_n = \min(U_1, \ldots, U_n)$.

1. Donner (sans détailler les calculs) la fonction de répartition de U_1.
La variable aléatoire U_1 suit une loi uniforme sur $[0,1]$ donc sa fonction de répartition est donnée par
$$F_{U_1}(t) = \begin{cases} 0 & \text{si } t < 0\,, \\ t & \text{si } t \in [0,1]\,, \\ 1 & \text{si } t > 1\,. \end{cases}$$
2. Montrer que pour $t \in [0,1]$,
$$\mathbb{P}(Y_n > t) = (1-t)^n\,.$$
Pour $t \in [0,1]$, le caractère i.i.d des variables aléatoires $U_1, \ldots, U_n$ donne
$$\mathbb{P}(Y_n > t) = \mathbb{P}(U_1 > t, \ldots, U_n > t) = \mathbb{P}(U_1 > t) \cdots \mathbb{P}(U_n > t) = \mathbb{P}(U_1 > t)^n\,,$$
ce qui conclut la preuve puisque $\mathbb{P}(U_1 > t) = 1 - F_{U_1}(t) = 1 - t$ si $t \in [0,1]$.
3. Montrer que $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$.
Soit $\epsilon > 0$. La question précédente donne
$$\mathbb{P}(|Y_n| \geq \epsilon) = \mathbb{P}(Y_n \geq \epsilon) \leq (1-\epsilon)^n \xrightarrow[n\to\infty]{} 0\,.$$
Donc $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

4. (a) Montrer que si X est une variable aléatoire positive, alors

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > t)\,\mathrm{d}t\,.$$

On écrit $\mathbb{P}(X > t) = \mathbb{E}[\mathbb{1}_{\{X>t\}}] = \mathbb{E}[\mathbb{1}_{]t,\infty[}(X)]$ et le théorème de Fubini-Tonelli pour des fonctions mesurables positives donne alors l'égalité

$$\int_0^\infty \mathbb{P}(X > t)\,\mathrm{d}t = \int_0^\infty \mathbb{E}[\mathbb{1}_{]t,\infty[}(X)] = \mathbb{E}\Big[\int_0^\infty \mathbb{1}_{]t,\infty[}(X)\,\mathrm{d}t\Big]\,.$$

L'intégrale dans l'espérance se réécrit

$$\int_0^\infty \mathbb{1}_{]t,\infty[}(X)\,\mathrm{d}t = \int_0^\infty \mathbb{1}_{[0,X[}(t)\,\mathrm{d}t = \int_0^X \mathrm{d}t = X\,,$$

ce qui donne le résultat.

(b) En déduire la valeur de $\mathbb{E}[Y_n]$.

Il suffit d'utiliser l'égalité précédente avec la variable aléatoire positive Y_n :

$$\mathbb{E}[Y_n] = \int_0^\infty \mathbb{P}(Y_n > t)\,\mathrm{d}t = \int_0^\infty (1-t)^n \mathbb{1}_{[0,1]}(t)\,\mathrm{d}t = \Big[\frac{-(1-t)^{n+1}}{n+1}\Big]_0^1 = \frac{1}{n+1}\,.$$

(c) En déduire que $Y_n \xrightarrow[n\to\infty]{L^1} 0$, puis que pour tout $p \geq 1$, $Y_n \xrightarrow[n\to\infty]{L^p} 0$.

La question précédente implique que

$$\mathbb{E}[|Y_n|] = \mathbb{E}[Y_n] = \frac{1}{n+1} \xrightarrow[n\to\infty]{} 0\,,$$

ce qui assure la convergence de $(Y_n)_{n\geq 1}$ vers 0 dans L^1.

Soit alors $p > 1$. La variable aléatoire Y_n étant à valeurs dans $[0,1]$, elle vérifie $|Y_n|^p \leq |Y_n|$, donc par croissance de l'espérance, il vient

$$\mathbb{E}[|Y_n|^p] \leq \mathbb{E}[Y_n] \xrightarrow[n\to\infty]{} 0\,,$$

Ainsi, $(Y_n)_{n\geq 1}$ converge vers 0 dans L^p.

5. (a) Déterminer la fonction de répartition de nY_n.

La variable aléatoire nY_n est à valeurs dans $[0,n]$, donc F_{nY_n} est nulle avant 0 et vaut 1 après n. Si $t \in [0,n]$, alors $t/n \in [0,1]$ et la question, 2. donne

$$F_{nY_n}(t) = \mathbb{P}(nY_n \leq t) = \mathbb{P}(Y_n \leq t/n) = 1 - \mathbb{P}(Y_n > t/n) = 1 - (1-t/n)^n\,.$$

Bref, on a montré que $F_{nY_n}(t) = \big(1 - (1-t/n)^n\big)\mathbb{1}_{[0,n]}(t)$.

(b) Montrer que $(nY_n)_{n\geq 1}$ converge en loi vers une loi exponentielle de paramètre à déterminer.

Pour $t \in \mathbb{R}$, on a la convergence

$$F_{nY_n}(t) = \big(1 - (1-t/n)^n\big)\mathbb{1}_{[0,n]}(t) \xrightarrow[n\to\infty]{} (1 - \mathrm{e}^{-t})\mathbb{1}_{[0,\infty[}(t)\,,$$

La limite correspond à la fonction de répartition d'une loi exponentielle de paramètre 1, donc

$$nY_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{E}(1)\,.$$

EXERCICE 2 - Lois normales

Soit $X_1, \dots, X_n$ une suite de variables aléatoires indépendantes de loi normale.

1. On suppose dans cette question que pour tout $n \geq 1$, $X_n \sim \mathcal{N}(\mu, \sigma^2)$ avec $\sigma > 0$.

   (a) Déterminer la limite presque sûre de

   $$\frac{X_1^2 + \cdots + X_n^2}{n}.$$

   Les variables aléatoires $X_1^2, \dots, X_n^2$ sont indépendantes et identiquement distribuées puisque les X_k le sont. Par ailleurs, comme les X_k suivent une loi $\mathcal{N}(\mu, \sigma^2)$, les variables aléatoires X_k^2 vérifient

   $$\mathbb{E}[X_1^2] = \mathrm{Var}(X_1^2) + \mathbb{E}[X_1]^2 = \sigma^2 + \mu^2 .$$

   La loi forte des grands nombres assure alors que

   $$\frac{X_1^2 + \cdots + X_n^2}{n} \xrightarrow[n\to\infty]{p.s.} \sigma^2 + \mu^2 .$$

   (b) En déduire la limite presque sûre de

   $$\frac{X_1 + \cdots + X_n}{X_1^2 + \cdots + X_n^2}.$$

   La loi forte des grands nombres appliquées aux variables aléatoires i.i.d. et intégrables $X_1, \dots, X_n$ donne

   $$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{p.s.} \mu .$$

   En effectuant le quotient des deux convergences précédentes on obtient

   $$\frac{X_1 + \cdots + X_n}{X_1^2 + \cdots + X_n^2} \xrightarrow[n\to\infty]{p.s.} \frac{\mu}{\sigma^2 + \mu^2} .$$

2. On suppose maintenant que pour tout $n \geq 2$, $X_n \sim \mathcal{N}(0, 1/\ln(n))$.

   (a) Montrer, en utilisant la définition de la convergence en loi, que $X_n \xrightarrow[n\to\infty]{\mathcal{L}} \delta_0$.

   En déduire, sans calcul, que $X_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

   Soit $h : \mathbb{R} \to \mathbb{R}$ une fonction continue bornée. La formule de transfert et le changement de variable $y = \sqrt{\ln(n)}x$ donnent

   $$\mathbb{E}[h(X_n)] = \int_{\mathbb{R}} h(x) \frac{\sqrt{\ln(n)}}{\sqrt{2\pi}} \,\mathrm{e}^{-\frac{x^2 \ln(n)}{2}} \,\mathrm{d}x = \int_{\mathbb{R}} h(y/\sqrt{\ln(n)}) \frac{1}{\sqrt{2\pi}} \,\mathrm{e}^{-\frac{y^2}{2}} \,\mathrm{d}y .$$

   La fonction $y \mapsto h_n(y) = h(y/\sqrt{\ln(n)}) \,\mathrm{e}^{-\frac{y^2}{2}}$ est intégrable sur $\mathbb{R}$ et vérifie

   $$h_n(y) \xrightarrow[n\to\infty]{} h(0) \,\mathrm{e}^{-\frac{y^2}{2}} \quad \text{et} \quad h_n(y) \leq \|h\|_\infty \,\mathrm{e}^{-\frac{y^2}{2}} .$$

Le théorème de convergence dominée assure alors que

$$\int_{\mathbb{R}} h_n(y)\,\mathrm{d}y \xrightarrow[n\to\infty]{} \int_{\mathbb{R}} h(0)\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{y^2}{2}}\,\mathrm{d}y = h(0) = \int_{\mathbb{R}} h(y)\,\mathrm{d}\delta_0(y)\,.$$

Donc $X_n \xrightarrow[n\to\infty]{\mathcal{L}} \delta_0$.

La convergence en loi vers une constante implique la convergence en probabilité donc $X_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

(b) À partir des résultats donnés en rappel dans l'en-tête du sujet, montrer que

$$\mathbb{P}(X_n \geq 1) \underset{n\to\infty}{\sim} \frac{1}{\sqrt{2\pi n \ln(n)}}\,.$$

La variable aléatoire $\sqrt{\ln(n)}\ X_n$ suit une loi normale $\mathcal{N}(0,1)$. Ainsi,

$$\begin{aligned}\mathbb{P}(X_n \geq 1) &= \mathbb{P}\Big(\sqrt{\ln(n)}\ X_n \geq \sqrt{\ln(n)}\Big)\\ &= 1-\Phi(\sqrt{\ln(n)}) \underset{n\to\infty}{\sim} \frac{\mathrm{e}^{-\frac{\sqrt{\ln(n)}^2}{2}}}{\sqrt{2\pi\ln(n)}} = \frac{1}{\sqrt{2\pi n\ln(n)}}\,.\end{aligned}$$

(c) En déduire que $\mathbb{P}(\limsup_{n\to\infty}\{X_n \geq 1\}) = 1$.

Les événements $A_n = \{X_n \geq 1\}$, $n \geq 2$, sont indépendants et la série $\sum_{n\geq 2}\mathbb{P}(A_n)$ diverge d'après le critère de Bertrand. Le lemme de Borel-Cantelli donne alors

$$\mathbb{P}(\limsup_{n\to\infty}\{X_n \geq 1\}) = 1\,.$$

(d) En déduire que la suite $(X_n)_{n\geq 1}$ ne converge pas vers 0 presque sûrement.

D'après la question précédente, la suite $(X_n)_{n\geq 2}$ prend des valeurs supérieures à 1 une infinité de fois presque sûrement, donc elle ne peut pas converger vers 0 presque sûrement.

Exercice 3 - Pharmacovigilance

Une entreprise pharmaceutique souhaite contrôler ses médicaments. Chaque médicament a une probabilité $p \in]0,1[$ d'être mal dosé. Un agent de contrôle teste n médicaments qui sortent de la chaîne de production. On note X_k la variable aléatoire valant 1 si le k-ème médicament testé est mal dosé, et 0 sinon. Enfin, on pose $S_n = X_1 + \cdots + X_n$.

1. Donner la loi de S_n, puis son espérance et sa variance.

La variable aléatoire S_n correspond à la somme de n variables aléatoires indépendantes de loi de Bernoulli de paramètre p, donc S_n suit une loi binomiale de paramètres n et p. Par suite,

$$\mathbb{E}[S_n] = np \quad \text{et} \quad \mathrm{Var}(S_n) = np(1-p)\,.$$

2. Montrer que $(S_n - np)/\sqrt{n}$ converge en loi vers une loi normale de paramètres à préciser.
   Les variables aléatoires $X_1, \ldots, X_n$ étant indépendantes, identiquement distribuées et de carré intégrable, le théorème central limite assure que
$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)\,,$$
   ce qui se réécrit ici
$$\frac{S_n - np}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, p(1-p))\,.$$

3. En déduire, en justifiant, la valeur de q telle que
$$\mathbb{P}\Big(\frac{S_n}{n} - q\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq p \leq \frac{S_n}{n} + q\frac{\sqrt{p(1-p)}}{\sqrt{n}}\Big) \xrightarrow[n\to\infty]{} 0.95\,.$$
   La convergence en loi précédente se réécrit en termes de fonction de répartition via
$$\mathbb{P}\Big(\frac{S_n - np}{\sqrt{np(1-p)}} \leq x\Big) \xrightarrow[n\to\infty]{} \mathbb{P}(N \leq x) = \Phi(x)\,, \quad x \in \mathbb{R}\,,$$
   avec $N \sim \mathcal{N}(0,1)$. On obtient alors
$$\mathbb{P}\Big(-q \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq q\Big) \xrightarrow[n\to\infty]{} \mathbb{P}(-q \leq N \leq q) = \Phi(q) - \Phi(-q) = 2\Phi(q) - 1\,,$$
   par symétrie de la densité de la loi normale centrée réduite. On souhaite avoir $2\Phi(q) - 1 = 0.95$, donc $\Phi(q) = 0.975$. Le rappel donne alors $q = 1.96$. Ainsi, on a établi la convergence
$$\mathbb{P}\Big(-1.96 \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq 1.96\Big) \xrightarrow[n\to\infty]{} 0.95\,,$$
   ce qui se réécrit
$$\mathbb{P}\Big(\frac{S_n}{n} - 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq p \leq \frac{S_n}{n} - 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\Big) \xrightarrow[n\to\infty]{} 0.95\,.$$

4. Donner un intervalle I_C ne dépendant que de n et de S_n, tel qu'asymptotiquement le paramètre inconnu p appartienne à l'intervalle I_C avec une probabilité d'au moins 95%. On précisera le centre de l'intervalle et sa longueur.
   On vient de montrer qu'asymptotiquement, le paramètre inconnu p appartient à l'intervalle
$$\Big[\frac{S_n}{n} - 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}, \frac{S_n}{n} - 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\Big]$$
   avec une probabilité de 0.95. La majoration $p(1-p) \leq 1/4$ pour tout $p \in]0,1[$ implique que
$$\Big[\frac{S_n}{n} - 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}, \frac{S_n}{n} - 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\Big] \subset \Big[\frac{S_n}{n} - \frac{0.98}{\sqrt{n}}, \frac{S_n}{n} - \frac{0.98}{\sqrt{n}}\Big]\,.$$

Donc la probabilité que p soit dans le deuxième intervalle est asymptotiquement d'au moins 95%. L'intervalle recherché est donc

$$I_C = \Big[\frac{S_n}{n} - \frac{0.98}{\sqrt{n}}, \frac{S_n}{n} - \frac{0.98}{\sqrt{n}}\Big],$$

qui est centré en S_n/n et de longueur $2 \times 0.98/\sqrt{n} = 1.96/\sqrt{n}$.

5. Comment évolue la longueur de intervalle I_C quand n augmente ? Interpréter le résultat.

   La longueur de l'intervalle I_C est $2 \times 0.98/\sqrt{n} = 1.96/\sqrt{n}$ qui diminue avec n. Dire que n augmente signifie qu'on a davantage de données, donc l'intervalle de confiance est plus précis.

6. On souhaite maintenant proposer un intervalle I'_C qui contienne asymptotiquement p avec un probabilité d'au moins 90%. Sans refaire tous les calculs, dire si l'intervalle I'_C sera plus grand ou plus petit que l'intervalle I_C. On justifiera la réponse.

   Si on choisit un probabilité de 0.90, alors seule la valeur de q déterminée en question 3. est modifiée : on cherche cette fois q telle que $2\Phi(q) - 1 = 0.9$. On obtient une valeur de q plus petite, donc un intervalle I'_C plus petit.