Quizz Kahoot

Eva Lawrence, Aude Sportisse

UE 247 - Statistiques Inférentielles

1 Quizz numéro 1 : Rappels de Probabilités

. Question 1 : Parmi ces définitions, laquelle est celle d'une convergence en probabilité de la suite (X_n) vers X ?

1. $$\lim_{n\to\infty} \mathbb{P}(X_n = X) = 1$$
2. $$\forall \varepsilon > 0,\ \mathbb{P}(|X_n - X| > \varepsilon) = 0$$
3. $$\forall \varepsilon > 0,\ \mathbb{P}(|X_n - X| > \varepsilon) \underset{n\to\infty}{\to} 0$$
4. $$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 0.$$

Commentaire : La bonne réponse est la réponse 3.
Il faut introduire une limite pour que la réponse 2 soit correcte.
Les réponses 1 et 4 sont des "mauvaises définitions" de la convergence presque sûre. La limite doit être dans la proba pour la réponse 1 (i.e. c'est "l'évènement limite" que l'on regarde). Le membre de gauche de la question 4 est correct mais la proba doit valoir 1 au lieu de 0 pour définir la convergence presque sûre.

. Question 2 : Parmi ces implications de convergence, lesquelles sont justes.

1. Soit a une constante
$$X_n \xrightarrow{\mathbb{P}} a \quad \Rightarrow \quad X_n \xrightarrow{\mathcal{L}} a.$$
2. Soit a une constante
$$X_n \xrightarrow{\mathcal{L}} a \quad \Rightarrow \quad X_n \xrightarrow{\mathbb{P}} a \quad \Rightarrow \quad X_n \xrightarrow{\text{p.s.}} a.$$
3. Soit $p > 1$
$$X_n \xrightarrow{\mathrm{L}^p} X \quad \Rightarrow \quad X_n \xrightarrow{\mathbb{P}} X.$$
4. Soit f une fonction continue bornée
$$X_n \xrightarrow{\mathbb{P}} f(X) \quad \Rightarrow \quad X_n \xrightarrow{\mathcal{L}} f(X).$$

Commentaire : Toutes ces implications sont justes en dehors de la proposition 2. On retiendra dans le cas général

$$X_n \xrightarrow{\text{p.s.}} X \quad \Rightarrow \quad X_n \xrightarrow{\mathbb{P}} X \quad \Rightarrow \quad X_n \xrightarrow{\mathcal{L}} X$$

et pour $p > 1$

$$X_n \xrightarrow{\mathbf{L}^p} X \quad \Rightarrow \quad X_n \xrightarrow{\mathbb{P}} X.$$

Dans le cas où la variable aléatoire X est une constante déterministe a, on a alors

$$X_n \xrightarrow{\mathcal{L}} a \quad \Rightarrow \quad X_n \xrightarrow{\mathbb{P}} a.$$

. Question 3 : Soient X, Y deux variables aléatoires réelles et a,b deux scalaires. On veut développer $\text{Var}(aX + bY)$. Quelles sont les expressions correctes ?

1. X et Y sont i.i.d.
$$\text{Var}(aX + bY) = a\text{Var}(X) + b\text{Var}(Y)$$
2. X et Y sont i.i.d.
$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$$
3. X et Y sont centrées
$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$$
4. X et Y sont centrées
$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y).$$

Commentaire : Les réponses 2 et 4 sont justes.
Se rappeler que $\mathbf{Var}(aX + bY) = \mathbf{Cov}(aX + bY, aX + bY)$.
En développant avec la bilinéarité de la covariance on a alors

$$\mathbf{Var}(aX + bY) = a^2\mathbf{Var}(X) + b^2\mathbf{Var}(Y) + 2ab\mathbf{Cov}(X, Y).$$

Se rappeler enfin que si X et Y sont indépendantes alors $\mathbf{Cov}(X, Y) = 0$.
Le centrage des variables n'a pas d'impact sur la variance.

. Question 4 : parmi ces inégalités, lesquelles sont des inégalités de Markov ?

1. If $c > 0$, $\mathbb{E}\left[X^2\right] < \infty$ alors
$$\mathbb{P}\left(|X - \mathbb{E}\left[X\right]| \geqslant c\right) \leqslant \frac{\text{Var}(X)}{c^2}.$$
2. If $c > 0$, X est centrée et $\mathbb{E}\left[X^2\right] < \infty$ alors
$$\mathbb{P}\left(|X| \geqslant c\right) \leqslant \frac{\text{Var}(X)}{c^2}.$$
3. If $c > 0$, X est positive et $\mathbb{E}\left[X^2\right] < \infty$ alors
$$\mathbb{P}\left(X \geqslant c\right) \leqslant \frac{\mathbb{E}\left[X^2\right]}{c^2}.$$

4. If $c > 0$, X est positive et $\mathbb{E}[X] < \infty$ alors

$$\mathbb{P}(X \geqslant c) \leqslant \frac{\mathbb{E}[X]}{c}.$$

5. If $c > 0$, X est positive et $\mathbb{E}[|X|^p] < \infty$ alors

$$\mathbb{P}(|X| \geqslant c) \leqslant \frac{\mathbb{E}[X^p]}{c^p}.$$

Commentaire : La première inégalité est l'inégalité de Bienaymé-Chebychev. Toutes les autres sont des inégalités de Markov : pour $p = 2$ et X centrée dans la proposition 2, pour $p = 2$ et X positive dans la proposition 3, pour $p = 1$ et X positive dans la proposition 3, et pour $p \geqslant 1$ et X positive dans la proposition 5.

2 Quizz numéro 2 : Lois usuelles et caractérisation d'une loi

. Question 1 : Parmi ces fonctions, laquelle ne permet pas de définir la loi de la v.a.r. continue X de support $[a;b]$?

1. La fonction de densité f définie par
$$\int_a^b f(t)dt = 1.$$
2. La fonction de répartition F_X définie pour tout $x \in [a;b]$ par
$$F_X(x) = \mathbb{P}(X \leqslant x) = \int_a^x f(t)dt.$$
3. Les poids pour toute valeur $x \in [a;b]$
$$\mathbb{P}(X = x).$$
4. La fonction génératrice des moments
$$M_X(x) = \mathbb{E}[\exp(xX)] = \int_a^b f(t)\exp(xt)dt.$$

Commentaire : La mauvaise réponse est la proposition 3.
La variable aléatoire X est continue. Les poids pour toute valeur $x \in [a;b]$ $\mathbb{P}(X = x)$ sont donc tous nuls et n'apportent aucune information sur la variable aléatoire X.
Remarque, la proposition 1 manque de rigueur. La fonction de densité est définie à un ensemble de mesure nulle près, voir l'explication ci-dessous.

> **Remarque 4.15.** Il est clair d'après cette définition que si Y est une autre variable aléatoire ayant même loi que X (donc mêmes probabilités d'appartenance aux intervalles), elle a aussi la densité f. D'autre part, il n'y a pas unicité de la densité d'une variable aléatoire. Par exemple $g_1 = \mathbf{1}_{[0,1]}$ et $g_2 = \mathbf{1}_{]0,1[}$ sont deux densités de probabilité qui donnent les mêmes intégrales : $\int_a^b g_1(t)\,dt = \int_a^b g_2(t)\,dt$ pour toute paire de réels a et b. Ces deux fonctions peuvent chacune être prise comme densité de la loi uniforme sur $[0,1]$.
>
> L'exemple ci-dessus montre qu'il ne suffit pas de vérifier que deux variables aléatoires ont des densités qui diffèrent en un point pour en déduire qu'elles n'ont pas même loi. Nous admettrons le lemme suivant qui donne une condition suffisante pratique pour que deux variables à densité n'aient pas même loi.

. Question 2 : Parmi ces propositions, donner la moyenne et la variance de la loi uniforme discrète $X \sim \mathcal{U}(1;k)$.

1. $$\mathbb{E}[X] = \frac{k+1}{2} \qquad \text{Var}(X) = \frac{k^2}{12}$$
2. $$\mathbb{E}[X] = \frac{k+1}{2} \qquad \text{Var}(X) = \frac{k^2-1}{12}$$
3. $$\mathbb{E}[X] = \frac{k}{2} \qquad \text{Var}(X) = \frac{k^2-1}{12}$$
4. $$\mathbb{E}[X] = \frac{k}{2} \qquad \text{Var}(X) = \frac{k^2}{12}$$

Indication : se rappeler de ce résultat classique se démontrant par récurrence

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

Commentaire : On peut éliminer rapidement les réponses 3 et 4 en se rappelant que l'espérance d'une uniforme discrète vaut la valeur moyenne des deux bornes. On utilise ensuite l'indication pour calculer $\mathbb{E}\left[X^2\right]$. On a alors

$$\begin{aligned}\mathbf{Var}\,(X) &= \sum_{l=1}^{k} \frac{1}{k} l^2 - \left(\frac{k+1}{2}\right)^2 \\ &= \frac{1}{k}\frac{k(k+1)(2k+1)}{6} - \left(\frac{k+1}{2}\right)^2 \\ &= \frac{k^2-1}{12}.\end{aligned}$$

. Question 3 : Soient X_1, X_2 deux variables aléatoires exponentielles de paramètres λ_1 et λ_2 indépendantes. Quelle information est correcte ?

1. $X_1 + X_2 \sim \mathcal{E}(\lambda_1 + \lambda_2)$.
2. Si $\lambda_1 = \lambda_2$, alors $X_1 + X_2 \sim \Gamma(2, \lambda_1)$.
3. $X_1 + X_2 \sim \mathcal{E}(\lambda_1 \lambda_2)$.
4. Si $\lambda_1 = \lambda_2$, $\frac{1}{2}\mathbb{E}\left[X_1 X_2\right] = \frac{1}{2\lambda_1^2}$.

Commentaire : Les propositions 2 et 4 sont correctes. Les variables X_1 et X_2 sont indépendantes, on a donc l'espérance du produit égal au produit des espérances ce qui permet de justifier la proposition 4. Cf. question 4 ci-dessous pour la justification de la proposition 2.
On peut invalider les autres propositions en calculant l'espérance de la somme des aléatoires étudiées. En effet, si $X \sim \mathcal{E}(\lambda)$ alors $\mathbb{E}[X] = \frac{1}{\lambda}$. On a donc $\mathbb{E}[X_1 + X_2] = \frac{\lambda_1+\lambda_2}{\lambda_1\lambda_2}$, ce qui permet d'éliminer les propositions 1 et 3.

. Question 4 : Expliquer en une phrase comment démontrer la propriété : Si $\lambda_1 = \lambda_2$, alors $X_1 + X_2 \sim \Gamma(2, \lambda_1)$.

Commentaire : On a la correspondance suivante entre la loi Γ et la loi $\mathcal{E}(\lambda)$

$$\mathcal{E}(\lambda) = \Gamma(1, \lambda)$$

et la propriété suivantes sur les lois Γ

$$\Gamma(a, \lambda) + \Gamma(b, \lambda) = \Gamma(a+b, \lambda).$$

. Question 5 : Soit $X \sim \mathcal{B}(p)$ et $Y \sim \mathcal{B}(n, p)$. Quelle affirmation est fausse ?

1. $\frac{1}{n}\mathbb{E}\left[Y\right] = \mathbb{E}\left[X\right] = p$.

2. $F_X(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1-p & \text{si } x \in [0;1[, \\ 1 & \text{si } x \geqslant 1. \end{cases}$

3. $\mathbb{E}\left[\exp(tX)\right] = M_X(t) = \left(M_Y(t)\right)^{\frac{1}{n}}$.

4. $\mathbb{P}\left(Y = k\right) = \mathbb{P}\left(X = 0\right)^k \mathbb{P}\left(X = 1\right)^{n-k}$.

Commentaire : La proposition 4 est fausse. Remarque, il faut échanger les puissances des deux termes du produit pour obtenir une expression correcte $\mathbb{P}\left(Y = k\right) = \mathbb{P}\left(X = 1\right)^k \left(X = 0\right)^{n-k}$**.**

Les propositions 1 et 3 se démontrent en utilisant le fait qu'une variable $Y \sim \mathcal{B}(n,p)$ **est équivalente à** n **tirages indépendants d'une variable** $X \sim \mathcal{B}(p)$**.**

Pour la proposition 2, la variable X **prend la valeur 1 avec probabilité** p **et la valeur 0 avec la probabilité** $1-p$**. Il est alors clair que la fonction de répartition vaut** $1-p$ **lorsque** X **prend une valeur dans l'intervalle** $[0;1[$**.**

3 Quizz numéro 3 : TCL et estimation

. Question 1 : Soit $(X_i)_{i=1,\dots,n}$, $n \geq 3$ un échantillon i.i.d., t.q. $\mathrm{Var}(X_1) = \sigma^2$ et $\mathbb{E}(X_1) = \mu$. Etant donné que σ^2 est connu, on propose deux estimateur pour μ:

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

et

$$Z_n = \frac{1}{2}(X_n + X_{n-1}).$$

En calculant le bias et le risque quadratique, lequel choisissez-vous?

1. $\bar{X}_n$ car il est sans biais, et Z_n a un bias.
2. Ils sont tous les deux sans biais, on choisit $\bar{X}_n$ car son risque quadratique est inférieur, quel que soit la valeur de $n \geq 3$.
3. Les deux sont équivalents: ils sont tous les deux sans bias et leur risque quadratique est égal à $\frac{\sigma^2}{n}$.
4. On choisit $\bar{Z}_n$ car il est sans biais et plus facile à calculer.

Commentaire : Ces estimateurs sont tous les deux sans biais, en fait Z_n est la somme des deux derniers termes divisée par 2, c'est comme si on fait la moyenne empirique mais seulement sur les 2 dernières observations (X_{n-1}) et X_n. Il suffit pour s'en convaincre de calculer $\mathbb{E}[\bar{X}_n]$ et $\mathbb{E}[Z_n]$ en utilisant la linéarité de l'espérance et le fait que les variables $(X_i)_{i=1,\dots,n}$ ont la même loi. Pour le risque quadratique, c'est assez logique que Z_n ait une variance supérieur, car on ne prend en compte que les deux derniers observations... Pour s'en convaincre, on calcule,

$$\mathbb{E}[(\bar{X}_n - \mu)^2] = \mathrm{Var}(\bar{X}_n) = \sigma^2/n$$

et

$$\mathbb{E}[(Z_n - \mu)^2] = \mathrm{Var}(Z_n) = \frac{1}{4}\left(\mathrm{Var}(X_n) + \mathrm{Var}(X_{n-1})\right) = \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{\sigma^2}{2},$$

où on a utilisé l'indépendance de X_n et X_{n-1} dans la deuxième égalité.

. Question 2 : Soit $\bar{X}_n$ tq

$$\sqrt{n}\bar{X}_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

On veut connaître la convergence asymptotique de $\exp(\bar{X}_n)$. Laquelle de ces propositions est vraie ?

1. $\sqrt{n}(e^{\bar{X}_n} - e) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 e^2)$
2. $\sqrt{n}(e^{\bar{X}_n} - e) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$
3. $\sqrt{n}e^{\bar{X}_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, e^{\sigma^2})$
4. $\sqrt{n}(e^{\bar{X}_n} - 1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$

Commentaire : Le plus prudent, c'est d'écrire la Delta Méthode sur un bout de papier. On applique la Delta Méthode à la fonction continue et dérivable en 0 $g(x) = e^x$, avec $g'(0) = e^0 = 1$. Cela donne, $\sqrt{n}(e^{\bar{X}_n} - e^0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1 \times \sigma^2)$ La réponse est la 4.

. Question 3 : $(X_i)_{i=1,\dots,n}$ et $(Y_i)_{i=1,\dots,n}$ sont deux échantillons i.i.d., tq $\mathrm{Var}(X_1) = \sigma_X^2$, $\mathbb{E}[X_1] = \mu_X$ et $\mathrm{Var}(Y_1) = \sigma_Y^2$, $\mathbb{E}[Y_1] = \mu_Y$.

Quelles propositions sont justes ?

1. $\sqrt{n}(\bar{X}_n - \mu_X) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_X^2)$ et $\sqrt{n}(\bar{Y}_n^2 - \mu_Y^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\sigma_Y^2\mu_Y^2)$
2. Si $\forall i \in \{1, \dots, n\}, X_i \perp\!\!\!\perp Y_i$,
$$\sqrt{n}(\bar{X}_n + \bar{Y}_n - (\mu_X + \mu_Y)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_X^2 + \sigma_Y^2)$$
3. Si $\forall i \in \{1, \dots, n\}$, les couples (X_i, Y_i) sont indépendants et $X_i \perp\!\!\!\perp Y_i$,
$$\sqrt{n}(\bar{X}_n + \bar{Y}_n - (\mu_X + \mu_Y)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_X^2 + \sigma_Y^2)$$
4. Si $\forall i \in \{1, \dots, n\}$ les couples (X_i, Y_i) sont indépendants,
$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i + Y_i) - (\mu_X + \mu_Y)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_X^2 + \sigma_Y^2)$$

Commentaire : Il faut regarder les questions une par une. Pour la première, la convergence en loi de $\bar{X}_n$ est évidemment vraie, pour la convergence en loi de $\bar{Y}_n^2$, il suffira d'appliquer la Delta Méthode avec $g(x) = x^2$ et $g'(\mu_Y) = 2\mu_Y$. On a donc bien $\sqrt{n}(\bar{Y}_n^2 - \mu_Y^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\sigma_Y^2\mu_Y^2)$. Pour les questions suivantes, notez que $\bar{X}_n + \bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i + Y_i)$. Pour appliquer le TCL il nous faut vraiment l'indépendance de (X_i, Y_i) avec (X_j, Y_j) si $i \neq j$. Donc il faut que les couples (X_i, Y_i) soient indépendants (et la réponse 2 est fausse). Pour obtenir la variance asymptotique $\sigma_X^2 + \sigma_Y^2$, on a aussi besoin que $X_i \perp\!\!\!\perp Y_i$, puisque

$$\mathrm{Var}(X_i + Y_i) = \mathrm{Var}(X_i) + \mathrm{Var}(Y_i) = \sigma_X^2 + \sigma_Y^2$$

si $X_i \perp\!\!\!\perp Y_i$. Les réponses vraies sont la 1 et la 3.

. Question 4 : On a
$$\sqrt{n}(\bar{X}_n - \mu_x) \xrightarrow{\mathcal{L}} X$$
et
$$\sqrt{n}(\bar{Y}_n - \mu_Y) \xrightarrow{\mathcal{L}} Y.$$

Dans quel(s) cas a-t-on
$$\sqrt{n}(\bar{X}_n - \mu_x) + \sqrt{n}(\bar{Y}_n - \mu_Y) \xrightarrow{\mathcal{L}} X + Y\ ?$$

1. C'est vrai tout le temps.
2. Si $X = 0$.
3. Si on a également $\sqrt{n}(\bar{X}_n - \mu_x) \xrightarrow{\mathbb{P}} X$ et $\sqrt{n}(\bar{Y}_n - \mu_Y) \xrightarrow{\mathbb{P}} Y$.
4. Si on a également $\sqrt{n}(\bar{X}_n - \mu_x) \xrightarrow{\mathbb{P}} X$ ou $\sqrt{n}(\bar{Y}_n - \mu_Y) \xrightarrow{\mathbb{P}} Y$.

Commentaire : Ce n'est pas vrai tout le temps, il y a un contre exemple dans le polycopié en dessous du théorème de Slutsky 1.5, la convergence en loi n'est pas stable par addition ! (Ni par multiplication.) Si $X = 0$, on a la convergence en loi

de $\sqrt{n}(\bar{X}_n - \mu_x)$ vers une constante 0, donc la convergence en proba vers 0, et on peut appliquer le théorème de Slutsky, qui nous donne bien

$$\sqrt{n}(\bar{X}_n - \mu_x) + \sqrt{n}(\bar{Y}_n - \mu_Y) \xrightarrow{\mathcal{L}} X + Y.$$

Si on a $\sqrt{n}(\bar{X}_n - \mu_x) \xrightarrow{\mathbb{P}} X$ et $\sqrt{n}(\bar{Y}_n - \mu_Y) \xrightarrow{\mathbb{P}} Y$, on a

$$\sqrt{n}(\bar{X}_n - \mu_x) + \sqrt{n}(\bar{Y}_n - \mu_Y) \xrightarrow{\mathbb{P}} X + Y,$$

et donc la convergence en loi. Par contre, cela ne suffira pas que $\sqrt{n}(\bar{X}_n - \mu_x) \xrightarrow{\mathbb{P}} X$ ou $\sqrt{n}(\bar{Y}_n - \mu_Y) \xrightarrow{\mathbb{P}} Y$. Les réponses vraies sont les 2 et 3.