

ÉPREUVE 2 : UE PROBABILITÉS ET STATISTIQUE
DURÉE 1 HEURE 30

Tous les documents sont interdits. La calculatrice et le téléphone portable sont également interdits et doivent rester dans votre sac. Chaque réponse devra être clairement justifiée pour être validée.

Exercice 1 :

On se fixe un nombre entier n strictement positif et deux paramètres réels positifs p_x et p_y tels que $p_x + p_y \leq 1$. La loi trinomiale (n, p_x, p_y) est la loi du couple (X, Y) définie pour tout $(i, j) \in \mathbb{N}^2$ par

$$\mathbb{P}(X = i, Y = j) = \begin{cases} \frac{n!}{i!j!(n-i-j)!} p_x^i p_y^j (1-p_x-p_y)^{n-i-j} & \text{si } i+j \leq n \\ 0 & \text{sinon.} \end{cases}$$

On rappelle la formule du binôme : $(a+b)^n = \sum_{k=0}^n C_n^k a^k b^{n-k}$.

1. Montrer que $\sum_i \sum_j \mathbb{P}(X = i, Y = j) = 1$.

On a

$$\begin{aligned} \sum_{i,j} \mathbb{P}(X = i, Y = j) &= \sum_{(i,j): i+j \leq n} \frac{n!}{i!j!(n-i-j)!} p_x^i p_y^j (1-p_x-p_y)^{n-i-j} \\ &= \sum_{i=0}^n \sum_{j=0}^{n-i} \frac{n!}{i!j!(n-i-j)!} p_x^i p_y^j (1-p_x-p_y)^{n-i-j} \\ &= \sum_{i=0}^n \frac{n!}{i!(n-i)!} p_x^i \sum_{j=0}^{n-i} \frac{(n-i)!}{j!(n-i-j)!} p_y^j (1-p_x-p_y)^{n-i-j} \\ &= \sum_{i=0}^n \frac{n!}{i!(n-i)!} p_x^i (p_y + 1 - p_x - p_y)^{n-i} \\ &= \sum_{i=0}^n C_n^i p_x^i (1-p_x)^{n-i} \\ &= (p_x + 1 - p_x)^n = 1. \end{aligned}$$

2. Déterminer les lois marginales de X et de Y . Identifier les lois.

On a

$$\begin{aligned} P(X = i) &= \sum_j \mathbb{P}(X = i, Y = j) \\ &= \sum_{j=0}^{n-i} \frac{n!}{i!j!(n-i-j)!} p_x^i p_y^j (1-p_x-p_y)^{n-i-j} \\ &= \frac{n!}{i!(n-i)!} p_x^i \sum_{j=0}^{n-i} \frac{(n-i)!}{j!(n-i-j)!} p_y^j (1-p_x-p_y)^{n-i-j} \end{aligned}$$

$$\begin{aligned}
&= \frac{n!}{i!(n-i)!} p_x^i (p_y + 1 - p_x - p_y)^{n-i} \\
&= C_n^i p_x^i (1 - p_x)^{n-i}.
\end{aligned}$$

Donc X est de loi binomiale de paramètres (n, p_x) . Par symétrie, en remplaçant X par Y et p_x par p_y , on en déduit que Y est de loi binomiale de paramètres (n, p_y) .

3. Les variables X et Y sont-elles indépendantes ?

Clairement

$$\mathbb{P}(X = i)\mathbb{P}(Y = j) \neq \mathbb{P}(X = i, Y = j).$$

Donc X et Y ne sont pas indépendantes.

4. Calculer $\mathbb{P}(X + Y = k)$ pour $k \in \mathbb{N}$. Identifier la loi de $X + Y$.

Remarquons pour commencer que $\mathbb{P}(X + Y = k) = 0$ pour $k > n$. Pour $k \in \{0, \dots, n\}$, on a

$$\begin{aligned}
\mathbb{P}(X + Y = k) &= \sum_{i=0}^k \mathbb{P}(X = i, Y = k - i) \\
&= \sum_{i=0}^k \frac{n!}{i!(k-i)!(n-i-k+i)!} p_x^i p_y^{k-i} (1 - p_x - p_y)^{n-i-k+i} \\
&= \frac{n!}{k!(n-k)!} (1 - p_x - p_y)^{n-k} \sum_{i=0}^k \frac{k!}{i!(k-i)!} p_x^i p_y^{k-i} \\
&= C_n^k (p_x + p_y)^k (1 - p_x - p_y)^{n-k}.
\end{aligned}$$

Donc $X + Y$ est de loi binomiale de paramètres $(n, p_x + p_y)$.

5. Calculer $\mathbb{P}(X = i | Y = j)$ quand $i + j \leq n$. Identifier la loi de X conditionnellement à $\{Y = j\}$.

Pour $i + j \leq n$, par application de la formule de Bayes

$$\begin{aligned}
\mathbb{P}(X = i | Y = j) &= \frac{\mathbb{P}(X = i, Y = j)}{\mathbb{P}(Y = j)} \\
&= \frac{(n-j)!}{i!(n-i-j)!} p_x^i \frac{(1 - p_x - p_y)^{n-i-j}}{(1 - p_y)^{n-j}} \\
&= C_{n-j}^i \left(\frac{p_x}{1 - p_y} \right)^i \left(\frac{1 - p_x - p_y}{1 - p_y} \right)^{n-i-j} \\
&= C_{n-j}^i \left(\frac{p_x}{1 - p_y} \right)^i \left(1 - \frac{p_x}{1 - p_y} \right)^{n-i-j}.
\end{aligned}$$

Donc la loi conditionnelle de X sachant $\{Y = j\}$ est une loi binomiale de paramètres $(n - j, \frac{p_x}{1 - p_y})$.

6. Soit $G_{(X,Y)}$ la fonction génératrice du couple (X, Y) définie sur $[0, 1]^2$ par

$$G_{(X,Y)}(s, t) = \mathbb{E}(s^X t^Y).$$

Calculer cette fonction.

On a

$$G_{(X,Y)}(s, t) = \mathbb{E}(s^X t^Y)$$

$$\begin{aligned}
&= \sum_i \sum_j s^i t^j \mathbb{P}(X = i, Y = j) \\
&= \sum_{i=0}^n \sum_{j=0}^{n-i} s^i t^j \frac{n!}{i! j! (n-i-j)!} p_x^i p_y^j (1-p_x-p_y)^{n-i-j} \\
&= \sum_{i=0}^n \frac{n!}{i! (n-i)!} (s p_x)^i \sum_{j=0}^{n-i} \frac{(n-i)!}{j! (n-i-j)!} (t p_y)^j (1-p_x-p_y)^{n-i-j} \\
&= \sum_{i=0}^n C_n^i (s p_x)^i (t p_y + 1 - p_x - p_y)^{n-i} \\
&= (s p_x + t p_y + 1 - p_x - p_y)^n \\
&= (1 - (1-s)p_x - (1-t)p_y)^n.
\end{aligned}$$

7. Par un choix judicieux de t , retrouver la loi de X .

En prenant $t = 1$, on a

$$G_{(X,Y)}(s, 1) = \mathbb{E}(s^X) = (1 - (1-s)p_x)^n,$$

qui est exactement la fonction génératrice d'une loi binomiale de paramètres (n, p_x) .

8. Par un autre choix judicieux de t , retrouver la loi de $X + Y$.

En prenant $t = s$, on a

$$G_{(X,Y)}(s, s) = \mathbb{E}(s^{X+Y}) = (1 - (1-s)(p_x + p_y))^n,$$

qui est exactement la fonction génératrice d'une loi binomiale de paramètres $(n, p_x + p_y)$.

Exercice 2 :

A l'approche des élections, un institut de sondage contacte successivement des individus. Les appels sont supposés indépendants et chaque individu répond qu'il va voter pour le candidat A avec probabilité p_A ou pour le candidat B avec probabilité $p_B = 1 - p_A$. Le but est d'estimer p_A .

1. Soit N_A le nombre de réponses en faveur du candidat A collectées en n appels. Est-ce que $\frac{N_A}{n}$ converge ? si oui de quelle convergence s'agit-il ? et vers quoi $\frac{N_A}{n}$ converge ? On énoncera clairement le résultat du cours utilisé (hypothèses et conclusion).

Soit

$$X_i = \begin{cases} 1 & \text{si le } i\text{ème individu vote } A, \text{ avec probabilité } p_A \\ 0 & \text{si le } i\text{ème individu vote } B, \text{ avec probabilité } p_B = 1 - p_A. \end{cases}$$

Alors $N_A = \sum_{i=1}^n X_i$. Les variables aléatoires X_i sont indépendantes, de même loi, la loi bernoulli de paramètre p_A , et donc aussi de carré intégrable. On peut donc appliquer la loi faible des grand nombres qui donne la convergence en probabilité suivante :

$$\frac{N_A}{n} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mathbb{E}(X_1) = p_A.$$

2. Après avoir énoncé l'inégalité de Tchebychev, retrouver le résultat de la question précédente.

Inégalité de Tchebychev : $\forall a > 0$:

$$\mathbb{P}(|X - \mathbb{E}X| > a) \leq \frac{\text{Var}(X)}{a^2}.$$

On veut montrer que $\frac{N_A}{n} \xrightarrow{\mathbb{P}} p_A$. Clairement $\forall a > 0$:

$$\mathbb{P}\left(\left|\frac{N_A}{n} - p_A\right| > a\right) = \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - p_A\right| > a\right) \leq \frac{\text{Var}(\sum_{i=1}^n X_i/n)}{a^2} = \frac{\text{Var}(X_1)}{na^2} \leq \frac{\mathbb{E}(X^2)}{na^2} \rightarrow 0,$$

la dernière égalité provenant du fait que les variables aléatoires X_i sont indépendantes et de même loi et donc que $\text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{Var}(X_i) = n\text{Var}(X_1)$.

3. À partir d'un résultat fondamental du cours (que vous énoncerez), montrer que $\frac{N_A}{n}$ correctement renormalisé converge en loi et préciser vers quoi.

On peut appliquer directement le théorème central limite car les variables aléatoires X_i sont indépendantes, de même loi, la loi bernoulli de paramètre p_A , et donc aussi de carré intégrable. Par conséquent

$$\sqrt{n} \frac{\frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_1)}{\sqrt{\text{Var}(X_1)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

ce qui donne

$$\sqrt{n} \frac{\frac{N_A}{n} - p_A}{\sqrt{p_A(1 - p_A)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Exercice 3 :

Un professeur a construit l'histogramme suivant concernant les notes en mathématiques obtenues par ses élèves dans une classe.

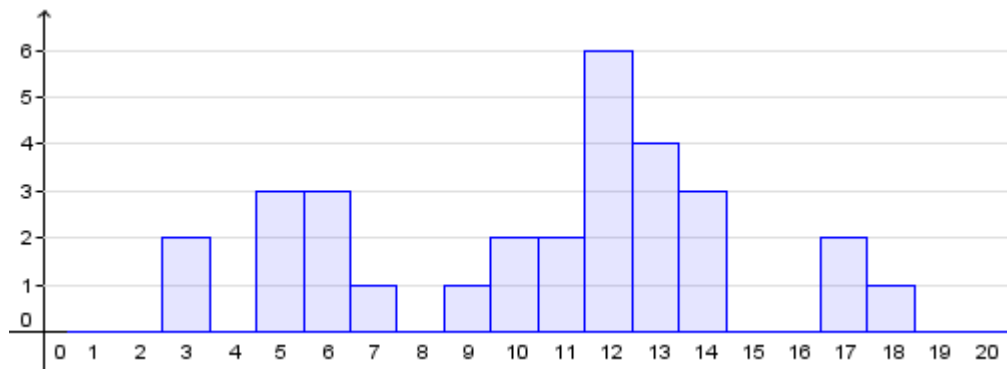


FIGURE 1 – Notes des étudiants en abscisse, nombre d'étudiants en ordonnée.

1. Construire un tableau avec les différentes modalités et les fréquences absolues correspondantes. Donner l'effectif total de la classe.

Modalités	3	5	6	7	9	10	11	12	13	14	17	18
Fréquences absolues	2	3	3	1	1	2	2	6	4	3	2	1

Effectif = $2+3+3+1+1+2+2+6+4+3+2+1=30$.

2. Calculer la moyenne empirique.

Moyenne empirique = $\frac{2*3+3*5+3*6+1*7+1*9+2*10+2*11+6*12+4*13+3*14+2*17+1*18}{30} = \frac{315}{30} = 10.5$.

3. Calculer la médiane.

30 est un chiffre pair, donc si x_1, \dots, x_{30} sont les notes des étudiants et $x_1^* \leq \dots \leq x_{30}^*$ sont les notes ordonnées, la médiane vaut $(x_{15}^* + x_{16}^*)/2 = 12$.

4. Calculer le premier quartile.

$n/4 = 7.5 \notin \mathbb{N}$ donc le 1er quartile vaut $q_{30,1/4} = x_{[7.5]+1}^* = x_8 = 6$.

5. Calculer le troisième quartile.

$3n/4 = 22.5 \notin \mathbb{N}$ donc le 3ème quartile vaut $q_{30,3/4} = x_{[22.5]+1}^* = x_{23} = 13$.

6. Calculer l'écart interquartile.

L'écart interquartile vaut $q_{30,3/4} - q_{30,1/4} = 13 - 6 = 7$.

7. Construire le boxplot associé.

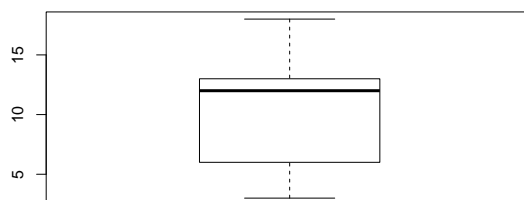


FIGURE 2 – Répartition des notes