

# TD d'analyse de données (ECOL401\_BE)

## Partie 2 : Probabilité

licence de sciences de la vie — deuxième année

février-mars 2024

## Exercice 2.1

Au cours d'une expérience sur le comportement des animaux, des rats doivent choisir entre quatre portes d'apparence identique. L'une est bonne et les trois autres mauvaises : chaque fois que le rat choisit une mauvaise porte, il reçoit une décharge électrique et est ramené au départ, cela jusqu'à ce qu'il choisisse la bonne porte. On note  $X$  le nombre d'essais.

1. On suppose que le rat n'a pas de mémoire et qu'il choisit à chaque essai de façon équiprobable entre chaque porte. Déterminer la loi de  $X$  et son espérance.
2. On suppose que le rat a une mémoire parfaite : à chaque nouvel essai, il choisit de façon équiprobable entre les portes qu'il n'a pas encore essayées. Déterminer la loi de  $X$  et son espérance.

*Indication.* Pour chaque essai  $i \geq 1$ , on pourra considérer la variable  $Y_i$  prenant la valeur 1 si le rat choisit la bonne porte et la valeur 0 sinon.

## Exercice 2.1 — correction (question 1)

Le rat n'a pas de mémoire, donc les variables  $Y_i$  sont deux à deux *indépendantes*.

Pour chaque entier  $k \geq 1$ , on calcule

$$\begin{aligned} P(X = k) &= P(Y_1 = 0, Y_2 = 0, \dots, Y_{k-1} = 0, Y_k = 1) \\ &= P(Y_1 = 0)P(Y_2 = 0) \cdots P(Y_{k-1} = 0)P(Y_k = 1) \\ &= \left(\frac{3}{4}\right)^{k-1} \frac{1}{4} = (1-p)^{k-1}p \end{aligned}$$

Ainsi la variable  $X$  suit une loi géométrique de paramètre  $1/4$  dont l'espérance vaut 4.

## Exercice 2.1 — correction (question 2)

On calcule d'abord

$$P(X = 1) = P(Y_1 = 1) = \frac{1}{4}.$$

Puis on trouve

$$\begin{aligned} P(X = 2) &= P(Y_1 = 0, Y_2 = 1) \\ &= P(Y_2 = 1 \mid Y_1 = 0)P(Y_1 = 0) \\ &= \frac{1}{3} \times \frac{3}{4} = \frac{1}{4}. \end{aligned}$$

## Exercice 2.1 — correction (question 2)

De la même manière, on obtient

$$\begin{aligned}P(X = 3) &= P(Y_1 = 0, Y_2 = 0, Y_3 = 1) \\&= P(Y_3 = 1 \mid Y_1 = 0, Y_2 = 0)P(Y_1 = 0, Y_2 = 0) \\&= P(Y_3 = 1 \mid Y_1 = 0, Y_2 = 0)P(Y_2 = 0 \mid Y_1 = 0)P(Y_1 = 0) \\&= \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} = \frac{1}{4}\end{aligned}$$

et

$$\begin{aligned}P(X = 4) &= P(Y_1 = 0, Y_2 = 0, Y_3 = 0, Y_4 = 1) \\&= P(Y_4 = 1 \mid Y_1 = 0, Y_2 = 0, Y_3 = 0)P(Y_1 = 0, Y_2 = 0, Y_3 = 0) \\&= P(Y_4 = 1 \mid Y_1 = 0, Y_2 = 0, Y_3 = 0)P(Y_3 = 0 \mid Y_1 = 0, Y_2 = 0) \\&\quad \times P(Y_2 = 0 \mid Y_1 = 0)P(Y_1 = 0) \\&= 1 \times \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} = \frac{1}{4}.\end{aligned}$$

## Exercice 2.1 — correction (question 2)

Pour terminer, on a

$$P(X > 4) = 1 - P(X \leq 4) = 1 - 4 \times \frac{1}{4} = 0.$$

> En conclusion, la variable  $X$  suit une loi uniforme sur l'ensemble  $\{1, 2, 3, 4\}$  et son espérance vaut 2,5, c'est-à-dire

$$X \sim \frac{\delta_1 + \delta_2 + \delta_3 + \delta_4}{4} \quad \text{et} \quad E[X] = 2,5.$$

## Exercice 2.2

Dans une espèce animale, un gène comprend deux allèles notés  $A$  (dominant) et  $a$  (récessif). On croise des hétérozygotes de génotype  $Aa$  et on considère des portées de 10 descendants. Soit  $X$  le nombre d'individus de phénotype  $A$ . Calculer la distribution de probabilité de cette variable aléatoire réelle  $X$ .

## Exercice 2.2 — correction

La variable  $X$  suit une loi binomiale de paramètre  $(n, p)$  avec  $n = 10$  et  $p = 0,75$ . Sa distribution de probabilité est donc donnée par l'égalité

$$P(X = k) = C_{10}^k \times 0,75^k \times 0,25^{10-k}, \quad k \in \{0, \dots, 10\}.$$

On calcule alors les nombres  $P(X = 0), \dots, P(X = 10)$  qui valent respectivement

$$10^{-6}, \quad 3 \cdot 10^{-5}, \quad 0,0004, \quad 0,0031, \quad 0,0162, \quad 0,0584, \\ 0,1460, \quad 0,2503, \quad 0,2816, \quad 0,1877, \quad 0,0563.$$

## Exercice 2.3

La proportion normale de polynucléaires basophiles est de 0,5 pour cent leucocytes. Calculer la probabilité de n'en trouver aucun parmi 200 leucocytes observés.

## Exercice 2.3 — correction

Le nombre de polynucléaires basophile parmi 200 leucocytes est une variable aléatoire réelle  $X$  suivant une loi binomiale de paramètre  $(n, p)$  avec  $n = 200$  et  $p = 0,005$ .

La probabilité de trouver aucun polynucléaire basophile vaut donc

$$P(X = 0) = C_{200}^0 \times 0,005^0 \times (1 - 0,005)^{200} = 0,995^{200} = 0,367.$$

## Exercice 2.4

Il y a une probabilité égale à 0,75 qu'un animal réagisse positivement à l'injection d'un médicament. Ce médicament est injecté à huit animaux choisis au hasard. Calculer les probabilités des évènements suivants :

1. aucun animal ne réagit positivement ;
2. exactement deux animaux réagissent positivement ;
3. au plus deux animaux réagissent positivement ;
4. au moins deux animaux réagissent positivement.

## Exercice 2.4 — correction

On note  $X$  le nombre d'animaux parmi les huit ayant réagi positivement. Cette variable aléatoire  $X$  suit une loi binomiale de paramètre  $(n, p)$  avec  $n = 8$  et  $p = 0,75$ .

On note  $q := 1 - p$ .

## Exercice 2.4 — correction

1. La probabilité qu'aucun animal ne réagisse positivement vaut

$$P(X = 0) = C_n^0 p^0 q^n = 1,526 \times 10^{-5}.$$

2. La probabilité qu'exactement deux animaux réagissant positivement vaut

$$P(X = 2) = C_n^2 p^2 q^6 = 3,845 \times 10^{-3}.$$

3. La probabilité qu'au plus deux animaux réagissent positivement vaut

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 4,227 \times 10^{-3}.$$

4. La probabilité qu'au moins deux animaux réagissent positivement vaut

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) \\ &= 1 - P(X = 0) - P(X = 1) = 0,9996. \end{aligned}$$

## Exercice 2.5

Dans une population très nombreuse, des études ont montré qu'il y avait 2 % d'individus du type  $\gamma$ . On choisit au hasard un échantillon de 30 individus. Calculer

1. la probabilité de n'obtenir aucun individu du type  $\gamma$ ,
2. la probabilité d'obtenir exactement un individu du type  $\gamma$ ,
3. la probabilité d'obtenir trois ou moins d'individus du type  $\gamma$ ,
4. la probabilité d'obtenir trois ou plus d'individus du type  $\gamma$ .

## Exercice 2.5 — correction

Le nombre d'individu de type  $\gamma$  est une variable aléatoire réelle suivant une loi binomiale de paramètre  $(30, 0,02)$ .

1. La probabilité de n'obtenir aucun individu du type  $\gamma$  vaut

$$P(X = 0) = 0,545.$$

2. La probabilité d'obtenir exactement un individu du type  $\gamma$  vaut

$$P(X = 1) = 0,334.$$

3. La probabilité d'obtenir trois ou moins d'individus du type  $\gamma$  vaut

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 0,997.$$

4. La probabilité d'obtenir trois ou plus d'individus du type  $\gamma$  vaut

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) \\ &= 1 - P(X = 0) - P(X = 1) - P(X = 2) = 0,022. \end{aligned}$$

## Exercice 2.6

Un producteur de semences constitue, à partir d'un lot important de graines dont le pouvoir germinatif est de 95 (c'est-à-dire qu'en moyenne 5 graines sur 100 ne germent pas), des paquets de 20 graines. On choisit au hasard un paquet. Calculer la probabilité qu'au moins 18 graines germent.

## Exercice 2.6 — correction

Le nombre de graines qui germent est une variable aléatoire réelle  $X$  suivant une loi binomiale de paramètre  $(20, 0,95)$ .

La probabilité qu'au moins 18 graines germent vaut

$$\begin{aligned}P(X \geq 18) &= P(X = 18) + P(X = 19) + P(X = 20) \\&= C_{20}^{18} \times 0,95^{18} \times 0,05^2 + C_{20}^{19} \times 0,95^{19} \times 0,05^1 + C_{20}^{20} \times 0,95^{20} \times 0,05^0 \\&= 0,925.\end{aligned}$$

## Exercice 2.7

Le sang d'un être humain possède une caractéristique, appelée facteur Rhésus, qui peut prendre deux valeurs notées  $R^+$  et  $R^-$ . En France, pour chacun des deux sexes, la proportion des individus  $R^+$  est égale à 0,85 et celle des individus  $R^-$  est égale à 0,15.

1. La formation d'un couple ne dépend pas du facteur Rhésus des conjoint·e·s. Énumérer les appariements possibles et indiquer leurs probabilités.
2. Chez les seuls couples où l'homme est  $R^+$  et la femme  $R^-$ , il se produit dans 8 % des naissances des accidents qui nécessitent un traitement spécial du nouveau-né. Déterminer la proportion des naissances en France qui nécessitent ce traitement
3. Dans une maternité, il y a en moyenne vingt accouchements par semaine. Calculer la probabilité pour que, dans une semaine prise au hasard,
  - 3.1 aucun nouveau-né n'ait besoin du traitement,
  - 3.2 un seul nouveau-né ait besoin du traitement,
  - 3.3 il se présente (strictement) plus d'un cas à traiter pendant une semaine.

## Exercice 2.7 — correction (questions 1 & 2)

**Question 1.** Les appariements possibles et leurs probabilités associées sont données dans le tableau suivant.

	$R^+$	$R^-$
$R^+$	$0,85 \times 0,85 = 0,7225$	$0,15 \times 0,85 = 0,1275$
$R^-$	$0,85 \times 0,15 = 0,1275$	$0,15 \times 0,15 = 0,0225$

**Question 2.** La proportion des naissances qui nécessitent un traitement vaut

$$0,08 \times 0,85 \times 0,15 = 0,0102.$$

## Exercice 2.7 — correction (question 3)

Le nombre de nouveaux-nés ayant besoin d'un traitement est une variable aléatoire réelle suivant une loi binomiale de paramètre  $(20, 0,0102)$ .

1. La probabilité qu'aucun nouveau-né n'ait besoin du traitement vaut

$$P(X = 0) = 0,815.$$

2. La probabilité qu'un seul nouveau-né ait besoin du traitement vaut

$$P(X = 1) = 0,168.$$

3. La probabilité qu'il se présente (strictement) plus d'un cas à traiter pendant une semaine vaut

$$P(X > 1) = 1 - P(X = 0) - P(X = 1) = 0,017.$$

## Exercice 2.8

Le nombre de particules visibles dans les flacons d'une solution injectable suit une loi de Poisson. Sachant que, sur 10 000 flacons contrôlés, on a décelé 2 particules dans 109 flacons et 3 particules dans 6 flacons, estimer

1. le nombre total de particules décelées dans l'ensemble des 10 000 flacons,
2. le nombre de flacons sans particule.

## Exercice 2.8 — correction

On note  $X$  le nombre de particules visibles dans la solution injectable. On sait que cette variable aléatoire  $X$  suit une loi de Poisson dont on note  $\lambda > 0$  le paramètre. Sa loi est donc donnée par l'expression

$$e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} \delta_k.$$

Avec les hypothèses, on doit trouver le réel  $\lambda$  vérifiant

$$P(X = 2) = \frac{109}{10\,000} \quad \text{et} \quad P(X = 3) = e^{-\lambda} \frac{\lambda^3}{6},$$

c'est-à-dire

$$e^{-\lambda} \frac{\lambda^2}{2} = \frac{109}{10\,000} \quad \text{et} \quad e^{-\lambda} \frac{\lambda^3}{6} = \frac{6}{10\,000}.$$

## Exercice 2.8 — correction

On se retrouve avec le système d'équations

$$\begin{cases} e^{-\lambda} \frac{\lambda^2}{2} = \frac{109}{10\,000}, \\ e^{-\lambda} \frac{\lambda^3}{6} = \frac{6}{10\,000}. \end{cases}$$

En divisant la seconde ligne par la première, on obtient

$$\frac{e^{-\lambda} \frac{\lambda^3}{6}}{e^{-\lambda} \frac{\lambda^2}{2}} = \frac{\frac{6}{10\,000}}{\frac{109}{10\,000}},$$

c'est-à-dire

$$\frac{\lambda}{3} = \frac{6}{109}$$

On trouve alors  $\lambda = 3 \times 6/109 = 0,1651$ .

## Exercice 2.8 — correction

Le nombre total de particules décelées dans l'ensemble des 10 000 flacons vaut

$$10\,000 \times E[X] = 10\,000 \times \lambda = 1651.$$

Le nombre de flacons sans particules vaut

$$10\,000 \times P(X = 0) = 10\,000 \times e^{-\lambda} = 8478.$$

## Exercice 2.9

Pour estimer la densité de bactéries présentes dans un liquide, on en prélève des échantillons d'un volume constant et on les dépose chacun dans un tube à essai. Un réactif coloré indique si des bactéries se sont développées ou si le milieu est resté stérile. On admet que les bactéries sont réparties au hasard et que leur nombre suit une loi de Poisson. On réalise 100ensemencements, dont 25 restent stériles. Estimer le nombre de bactéries par échantillon de la taille de l'inoculât.

## Exercice 2.9 — correction

On note  $X$  le nombre de bactéries par échantillon de la taille de l'inoculat. Cette variable aléatoire  $X$  suit une loi de Poisson de paramètre  $\lambda > 0$ .

On sait que

$$P(X = 0) = 25/100 = 0,25,$$

c'est-à-dire  $e^{-\lambda} = 0,25$ . On en déduit que  $\lambda = -\ln 0,25$ .

Le nombre moyen de bactéries par échantillon vaut donc

$$E[X] = \lambda = -\ln 0,25 = 1,39.$$

## Exercice 2.14

Dans une population humaine, la glycémie peut être considéré comme une variable gaussienne de moyenne 1,05 g/l et de variance 0,0144 g<sup>2</sup>/l<sup>2</sup>. Calculer la proportion d'individus dont le taux de glycémie est compris entre 0,95 g/l et 1,30 g/l.

## Exercice 2.14 — correction

On note  $X$  la glycémie. La variable aléatoire  $X$  suit une loi normale de moyenne 1,05 et de variance 0,0144, c'est-à-dire

$$X \sim \mathcal{N}(1,05, 0,0144).$$

On obtient donc

$$Y := \frac{X - 1,05}{\sqrt{0,0144}} = \frac{X - 1,05}{0,12} \sim \mathcal{N}(0, 1).$$

On veut calculer la probabilité  $P(0,95 < X < 1,30)$  qui vaut

$$\begin{aligned} P(0,95 < X < 1,30) &= P\left(\frac{0,95 - 1,05}{0,12} < Y < \frac{1,30 - 1,05}{0,12}\right) \\ &= P(-0,83 < Y < 2,08) \\ &= \Phi(2,08) - \Phi(-0,83) \\ &= 0,779. \end{aligned}$$

## Exercice 2.15

Le quotient intellectuel, noté QI, est une variable de loi normale de moyenne 100 et d'écart-type 16.

1. Calculer le nombre  $a$  tel que

$$P(|QI - 100| \leq a) = 0,5.$$

2. Calculer le nombre  $a$  tel que

$$P(|QI - 100| \leq a) = 0,75.$$

## Exercice 2.15 — correction

On a  $QI \sim \mathcal{N}(100, 16^2)$ , donc

$$X := \frac{QI - 100}{16} \sim \mathcal{N}(0, 1).$$

Pour un nombre  $a$  quelconque, on calcule

$$\begin{aligned} \mathbf{P}(|QI - 100| \leq a) &= \mathbf{P}(|X| \leq a/16) \\ &= \Phi(a/16) - \Phi(-a/16) \\ &= \Phi(a/16) - (1 - \Phi(a/16)) \\ &= 2\Phi(a/16) - 1. \end{aligned}$$

## Exercice 2.15 — correction

On sait que

$$P(|QI - 100| \leq a) = 2\Phi(a/16) - 1.$$

1. Alors

$$P(|QI - 100|) = 0,5 \iff \Phi(a/16) = \frac{1 + 0,5}{2} = 0,75.$$

D'après la table, on a

$$\Phi(0,67) = 0,7486 \quad \text{et} \quad \Phi(0,68) = 0,7517.$$

On fait une interpolation linéaire pour trouver le nombre  $z := a/16$  tel que  $\Phi(z) = 0,75$ .

On veut trouver un réel  $\alpha \in [0; 1]$  tel que

$$(1 - \alpha) \times 0,7486 + \alpha \times 0,7517 = 0,75.$$

On obtient

$$\alpha = \frac{0,7500 - 0,7486}{0,7517 - 0,7486} = 0,4516.$$

## Exercice 2.15 — correction

1. On obtient

$$\alpha = \frac{0,7500 - 0,7486}{0,7517 - 0,7486} = 0,4516.$$

On en déduit que

$$z = (1 - \alpha) \times 0,67 + \alpha \times 0,68 = 0,6745$$

si bien que

$$a = 16z = 10,79.$$

2. D'après la table, on a

$$\Phi(0,67) = 0,7486 \quad \text{et} \quad \Phi(0,68) = 0,7517.$$

Avec les mêmes types de calculs, on trouve

$$a = 18,41.$$

## Exercice 2.16

Le pH de l'urine d'un·e adulte sain·e est une variable aléatoire normale de moyenne 6,25 et d'écart-type 0,36. Déterminer un intervalle centré sur la moyenne où se trouvera la mesure du pH urinaire pour 75 % des adultes sain·e·s.

## Exercice 2.16 — correction

Le pH de l'urine est une variable aléatoire  $X$  suivant la loi normale  $\mathcal{N}(6,25; 0,36^2)$ . On veut trouver un nombre  $a$  tel que

$$P(|X - 6,25| < a) = 0,75.$$

Comme précédemment, on doit résoudre l'équation

$$2\Phi(a/0,36) - 1 = 0,75,$$

c'est-à-dire  $\Phi(a/16) = 0,875$ .

Avec le tableau, on trouve  $a/0,36 = 1,15$ , c'est-à-dire  $a = 0,41$ . L'intervalle recherché vaut

$$[6,25 - a; 6,25 + a] = [5,84; 6,66].$$

## Exercice 2.17

On a constaté que la répartition du taux de cholestérol pour un grand nombre de personnes est la suivante.

taux	pourcentage
inférieur à 165 cg	58
compris entre 165 cg et 180 cg	38
supérieur à 180 cg	4

1. Sachant que la répartition suit une loi normale, calculer la valeur moyenne et l'écart-type.
2. On admet que les personnes dont le taux est supérieur à 183 cg doivent suivre un traitement. Calculer le nombre d'individus à soigner dans une population de 100 000 personnes.

## Exercice 2.17 — correction (question 1)

Le taux de cholestérol est une variable aléatoire  $X$  suivant une loi normale de moyenne  $\mu$  et d'écart-type  $\sigma$  inconnus. D'après les hypothèses, on dispose des trois égalités

$$P(X < 165) = 0,58,$$

$$P(165 < X < 180) = 0,38,$$

$$P(X > 180) = 0,04.$$

Avec la première et la dernière, on peut écrire

$$P\left(\frac{X - \mu}{\sigma} < \frac{165 - \mu}{\sigma}\right) = 0,58,$$

$$P\left(\frac{X - \mu}{\sigma} > \frac{180 - \mu}{\sigma}\right) = 0,04,$$

c'est-à-dire

$$\Phi\left(\frac{165 - \mu}{\sigma}\right) = 0,58,$$

$$1 - \Phi\left(\frac{180 - \mu}{\sigma}\right) = 0,04.$$

## Exercice 2.17 — correction (question 1)

On veut donc trouver deux nombres  $\mu$  et  $\sigma$  tel que

$$\Phi\left(\frac{165 - \mu}{\sigma}\right) = 0,58 \quad \text{et} \quad \Phi\left(\frac{180 - \mu}{\sigma}\right) = 0,96.$$

Avec le tableau de valeurs, on a

$$\begin{aligned} \Phi(0,20) &= 0,5793, & \Phi(0,21) &= 0,5832, \\ \Phi(1,75) &= 0,9599, & \Phi(1,76) &= 0,9608 \end{aligned}$$

et, par interpolation linéaire, on trouve

$$\begin{aligned} \frac{165 - \mu}{\sigma} &= 0,20 + 0,01 \times \frac{0,5800 - 0,5793}{0,5832 - 0,5793} = 0,202, \\ \frac{180 - \mu}{\sigma} &= 1,75 + 0,01 \times \frac{0,9600 - 0,9599}{0,9608 - 0,9599} = 1,751. \end{aligned}$$

## Exercice 2.17 — correction (question 1)

On a obtenu

$$\frac{165 - \mu}{\sigma} = 0,202 \quad \text{et} \quad \frac{180 - \mu}{\sigma} = 1,751.$$

Cela se réécrit sous la forme du système

$$\begin{cases} 0,202\sigma + \mu = 165, \\ 1,751\sigma + \mu = 180. \end{cases}$$

En soustrayant les deux lignes, on obtient

$$(1,751 - 0,202)\sigma = 180 - 165, \quad \text{c'est-à-dire} \quad \sigma = \frac{180 - 165}{1,751 - 0,202} = 9,68$$

puis

$$\mu = 165 - 0,202\sigma = 163.$$

## Exercice 2.17 — correction (question 2)

On veut calculer

$$100\,000 \times P(X > 163).$$

On trouve

$$\begin{aligned} P(X > 183) &= P\left(\frac{X - \mu}{\sigma} > \frac{183 - \mu}{\sigma}\right) \\ &= P\left(\frac{X - \mu}{\sigma} > 2,066\right) \\ &= 1 - \Phi(2,066) \\ &= \Phi(-2,066) \\ &= 0,0197 - (0,0197 - 0,0192) \times \frac{2,070 - 2,066}{2,070 - 2,060} \\ &= 0,0167. \end{aligned}$$

La quantité recherchée vaut donc

$$100\,000 \times P(X > 163) = 16\,700.$$