Contrôle de rattrapage
5 mars 2017

Durée de l'épreuve : **2 heures**.
Aucun document n'est autorisé. La calculatrice est autorisée.
Le barème est donné à titre indicatif et est susceptible d'être modifié.

Remarques :

— Tout calcul menant à une probabilité strictement supérieure à 1 ou strictement négative *sans aucun commentaire de votre part* donnera lieu à une pénalité d'un point.

— Pour la plupart des questions, si vous ne justifiez pas correctement une réponse (même juste), vous n'aurez aucun point.

— Une copie mal présentée ou dont les résultats sont trop souvent non justifiés ou très mal justifiés se verra pénalisée d'un maximum de 2 points.

— En revanche, il est possible dans certaines questions d'obtenir une partie des points en donnant un début de réponse ou de raisonnement.

— Ce sujet ne comporte aucun piège.

Exercice 1 Alice, Bob et Charlie, virologistes **10,5 points**

Le mystérieux virus Z, qui menace, à terme, de transformer des humains en zombies s'est répandu dans la population. Exactement 20% des personnes sont infectées. On ne peut savoir si un individu est infecté qu'après l'avoir examiné.

Alice, Bob et Charlie sont chargés de mener des examens sur des individus tirés au hasard dans la population, dans l'espoir de trouver un vaccin. Ces tirages sont uniformes parmi la population et sont indépendants.

1. Si l'on fait un seul tirage, et que l'on note X la variable aléatoire égale à 1 si l'individu est infecté par le virus Z et 0 sinon, quelle est la loi de X ? Son espérance ? Sa variance ? **/1**

   Solution : La variable aléatoire X vaut soit 0, soit 1, donc sa loi est la loi de Bernoulli. Le paramètre p de cette loi est la probabilité que X soit 1. Or, d'après l'énoncé, comme le choix de l'individu est uniforme,

$$p = \mathbb{P}(X = 1) = \frac{\text{Nombre d'individus infectés}}{\text{Nombre total d'individus}} = \text{Proportion d'individus infectés} = 20\% = 0,2.$$

2. Alice décide d'examiner des individus jusqu'à tomber sur une personne infectée. On note N le nombre (aléatoire) de personnes examinéees.

   a) Quelle est la loi de N ? **/1**

   Solution : On répète la même expérience de Bernoulli de paramètre $p = 0,2$, de façon indépendante et N est le temps du premier succès (« l'individu est infecté »), donc la loi de N est la loi géométrique de paramètre $0,2$.

   b) Quelle est son espérance ? **/0,5**

   Solution : Comme N suit la loi géométrique de paramètre $p = 0,2$, son espérance vaut

$$\mathbb{E}[N] = \frac{1}{p} = \frac{1}{0,2} = 5.$$

   c) Quelle est la probabilité que N soit égal à 6 ? **/1**

   Solution : Comme N suit la loi géométrique de paramètre $p = 0,2$, on a

$$\mathbb{P}(N = 6) = (1-p)^{6-1}\, p = 0,8^5 \times 0,2 \approx 0,066.$$

3. Soit n un entier supérieur ou égal à 1. Bob examine n individus. On note S le nombre de personnes infectées parmi ces n individus.

a) Quelle est la loi de S ? Son espérance ? Sa variance ? **/1,5**

Solution : La variable aléatoire S est le nombre de succès dans la répétition de n expériences de Bernoulli indépendantes de même paramètre $p = 0,2$, donc S suit la loi binomiale de paramètres n et $p = 0,2$.

b) Si $n = 6$, quelle est la probabilité que S soit égal à 2 ? **/0,5**

Solution : Comme $S \sim \mathcal{B}(6; 0,2)$, on a

$$\mathbb{P}(S = 2) = \binom{6}{2} p^2 (1-p)^{6-2} = \frac{6 \times 5}{2 \times 1} 0,2^2 \times 0,8^4 \approx 0,246.$$

c) Combien de personnes Bob doit-il examiner pour que la probabilité de tomber sur *au moins une* personne infectée soit supérieure ou égale à 0,95 ? **/1**

Solution : La probabilité que Bob n'examine aucune personne infectée est

$$\mathbb{P}(S = 0) = (1-p)^n = 0,8^n.$$

Donc la probabilité qu'il examine au moins une personne infectée est

$$\mathbb{P}(S \geq 1) = 1 - \mathbb{P}(S = 0) = 1 - 0,8^n.$$

On veut que $1 - 0,8^n \geq 0,95$, ce qui revient à $0,8^n \leq 0,05$. En passant au logarithme, c'est équivalent à

$$n \ln(0,8) \leq \ln(0,05).$$

En divisant par $\ln(0,8)$ *qui est négatif*, c'est équivalent à,

$$n \geq \frac{\ln(0,05)}{\ln(0,8)} \approx 13,4.$$

Comme n est entier, on a donc

$$\mathbb{P}(S \geq 1) \geq 0,95 \iff n \geq 14.$$

4. Charlie travaille beaucoup. Il examine 400 personnes. Pour i entre 1 et 400, soit X_i la variable aléatoire égale à 1 si la personne numéro i est infectée et 0 sinon. On note

$$\overline{X} = \frac{1}{400} \sum_{i=1}^{400} X_i$$

la proportion de personnes infectées parmi cet échantillon de 400 personnes.

a) On s'attend à ce que la variable aléatoire $\overline{X}$ soit « proche » d'un certain nombre (non aléatoire). Quel est ce nombre ? Et pourquoi ? (Donner le nom du théorème utilisé et justifier brièvement son utilisation). **/1**

Solution : La variable aléatoire $\overline{X}$ est la moyenne empirique d'un « grand » nombre de variables aléatoires i.i.d. intégrables. D'après la loi des grands nombres, on a, presque sûrement,

$$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow[n \to \infty]{} \mathbb{E}[X_1] = p = 0,2.$$

En supposant que 400 est suffisamment grand, on s'attend donc à ce que $\overline{X}$ soit donc « proche » de l'espérance de X_1 qui est, d'après la question 1, égale à $0,2$.

b) D'après un autre théorème (dont on donnera le nom), on peut approcher la loi de $\overline{X}$ par une loi à densité. Laquelle ? Quels sont ses paramètres ? **/1,5**

Solution : D'après le théorème central limite lorsque n est grand et que les variables aléatoires X_1, X_2, ..., X_n sont i.i.d. et ont une variance finie, la moyenne empirique $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ a une loi « proche » de la loi normale d' espérance X_1 et de variance $\mathbb{V}\mathrm{ar}(X_1)/n$.

D'après la question 1, $\mathbb{V}\mathrm{ar}(X_1) = p(1-p) = 0,2 \times 0,8 = 0,16$, donc la loi approchée de $\overline{X}$ est $\mathcal{N}(0,2; 0,16/400) = \mathcal{N}\left(0,2; 0,02^2\right)$.

c) En utilisant cette approximation, donner une valeur approchée de $\mathbb{P}\left(\overline{X} \leq 0,2392\right)$ et une valeur approchée de $\mathbb{P}\left(0,14848 \leq \overline{X} \leq 0,25152\right)$. **/1, 5**

Solution : On fait donc l'approximation $\overline{X} \sim \mathcal{N}\left(0,2; 0,02^2\right)$. On a d'abord,

$$\mathbb{P}\left(\overline{X} \leq 0,2392\right) = \mathbb{P}\left(\frac{\overline{X} - 0,2}{0,02} \leq \frac{0,2392 - 0,2}{0,02}\right) = \mathbb{P}\left(N \leq 1,96\right),$$

où $N = \frac{\overline{X}-0,2}{0,02}$ suit, par définition, la loi normale centrée réduite. D'après la table 1 de l'annexe, on a

$$\mathbb{P}\left(N \leq 1,96\right) \approx 0,975,$$

donc $\mathbb{P}\left(\overline{X} \leq 0,2392\right) \approx 0,975$.
Ensuite, on a, avec le même raisonnement,

$$\begin{aligned}\mathbb{P}\left(0,14848 \leq \overline{X} \leq 0,25152\right) &= \mathbb{P}\left(\frac{0,14848 - 0,2}{0,02} \leq \frac{\overline{X} - 0,2}{0,02} \leq \frac{0,25152 - 0,2}{0,02}\right) \\ &= \mathbb{P}\left(-2,576 \leq N \leq 2,576\right).\end{aligned}$$

Toujours d'après la table 1, on a $\mathbb{P}\left(N \leq 2,576\right) \approx 0,995$. On a donc $\mathbb{P}\left(N > 2,576\right) \approx 0,005$ et par symétrie $\mathbb{P}\left(N < -2,576\right) \approx 0,005$. Finalement,

$$\mathbb{P}\left(-2,576 \leq N \leq 2,576\right) = \mathbb{P}\left(N \leq 2,576\right) - \mathbb{P}\left(N < -2,576\right) \approx 0,995 - 0,005 = 0,990,$$

donc $\mathbb{P}\left(0,14848 \leq \overline{X} \leq 0,25152\right) \approx 0,990$.

Exercice 2 Daphnée, inventrice d'histoire drôles **6 points**

Le métier de Daphnée est d'inventer des histoires drôles. Malheureusement, l'inspiration ne se contrôle pas et pour trouver une histoire drôle, Daphnée doit réfléchir un temps (en heures) aléatoire S dont une densité est la fonction f donnée par

$$f(t) = \begin{cases} 0 & \text{si } t < 0\,; \\ \frac{1}{3}e^{-\frac{1}{3}t} & \text{sinon.} \end{cases}$$

1. La loi de S est-elle une loi de référence ? Si oui, laquelle, avec quel(s) paramètre(s) ? **/0, 5**
Solution : D'après le cours, une densité de la loi exponentielle de paramètre $\lambda > 0$ est la fonction f donnée par
$$f(t) = \begin{cases} 0 & \text{si } t < 0\,; \\ \lambda e^{-\lambda t} & \text{sinon.} \end{cases}$$
On reconnaît donc ici la loi exponentielle de paramètre $\lambda = 1/3$.

2. Quelle est l'espérance de S ? **/0, 5**
Solution : Comme S suit la loi exponentielle de paramètre $\lambda = 1/3$, son espérance, d'après le cours, vaut
$$\mathbb{E}\left[S\right] = \frac{1}{\lambda} = \frac{1}{\frac{1}{3}} = 3.$$

3. Calculer $\mathbb{P}\left(S > 2\right)$. **/1**
Solution : On calcule
$$\mathbb{P}\left(S > 2\right) = \int_2^{\infty} f(t)\,\mathrm{d}t = \int_2 \frac{1}{3}e^{-\frac{1}{3}t}\,\mathrm{d}t = \left[-e^{-\frac{1}{3}t}\right]_2^{\infty} = 0 - (-e^{-\frac{2}{3}}) = e^{-\frac{2}{3}} \approx 0,513.$$

Le contrat de travail de Daphnée indique que chaque jour, elle doit trouver 2 histoires drôles. On note T la variable aléatoire égale au temps (en heures) qu'il faut à Daphnée pour trouver ces deux histoires drôles. On admet que T est une variable aléatoire de densité g donnée par

$$g(t) = \begin{cases} 0 & \text{si } t < 0\,; \\ \frac{1}{9}te^{-\frac{1}{3}t} & \text{sinon.} \end{cases}$$

4. Vérifier que g est bien une densité de probabilité. **/1**
Solution : La fonction g est clairement positive. On doit vérifier que son intégrale vaut 1. On fait une intégration par partie en dérivant la fonction $t \mapsto t$ et en « primitivant » la fonction $t \mapsto \frac{1}{9}e^{-\frac{1}{3}t}$.

$$\int_{\mathbb{R}} g(t)\,\mathrm{d}t = \int_0^\infty \frac{1}{9}te^{-\frac{1}{3}t}\,\mathrm{d}t = \left[-\frac{1}{3}te^{-\frac{1}{3}t}\right]_0^\infty - \int_0^\infty \left(-\frac{1}{3}e^{-\frac{1}{3}t}\right)\mathrm{d}t = 0 + \int_0^\infty \frac{1}{3}e^{-\frac{1}{3}t}\,\mathrm{d}t = 1,$$

où, pour la dernière égalité, on a utilisé le fait que l'intégrale de la fonction f de la première partie de l'exercice est 1 (car c'est une densité de probabilité).

5. Calculer l'espérance de T. **/1**
Solution : D'après le cours, l'espérance de la variable T de densité g vaut

$$\mathbb{E}[T] = \int_{\mathbb{R}} tg(t)\,\mathrm{d}t = \int_0^\infty \frac{1}{9}t^2e^{-\frac{1}{3}t}\,\mathrm{d}t.$$

On fait encore une intégration par partie, en dérivant la fonction $t \mapsto t^2$ et en « primitivant » la fonction $t \mapsto \frac{1}{9}e^{-\frac{1}{3}t}$:

$$\begin{aligned}\int_0^\infty \frac{1}{9}t^2e^{-\frac{1}{3}t}\,\mathrm{d}t &= \left[2t \times \left(-\frac{1}{3}e^{-\frac{1}{3}t}\right)\right]_0^\infty - \int_0^\infty 2t\left(-\frac{1}{3}e^{-\frac{1}{3}t}\right)\mathrm{d}t \\ &= 0 + 2\int_0^\infty t\frac{1}{3}e^{-\frac{1}{3}t}\,\mathrm{d}t = 2 \times 3 = 6,\end{aligned}$$

où, pour voir que $\int_0^\infty t\frac{1}{3}e^{-\frac{1}{3}t}\,\mathrm{d}t = 3$, on s'est rappelé que la loi exponentielle de paramètre $1/3$ a pour espérance 3.

6. Vérifier que la fonction G définie par

$$G(t) = \begin{cases} 0 & \text{si } t < 0\,; \\ 1 - \left(\frac{1}{3}t + 1\right)e^{-\frac{1}{3}t} & \text{si } t \geq 0. \end{cases}$$

est bien la fonction de répartition de la variable aléatoire T. **/1**
Solution : Comme

$$\lim_{t\to\infty}\left(\frac{1}{3}t + 1\right)e^{-\frac{1}{3}t} = 0,$$

on a bien que $\lim_{t\to\infty} G(t) = 1$.
De plus, on a, pour tout $t > 0$,

$$G'(t) = 0 - \left[\frac{1}{3}e^{-\frac{1}{3}t} + \left(\frac{1}{3}t + 1\right)\left(-\frac{1}{3}\right)e^{-\frac{1}{3}t}\right] = -\frac{1}{3}e^{-\frac{1}{3}t} + \frac{1}{9}te^{-\frac{1}{3}t} + \frac{1}{3}e^{-\frac{1}{3}t} = \frac{1}{9}te^{-\frac{1}{3}t} = g(t).$$

Donc sur l'intervalle $]0\,;\infty[$, la fonction G est bien la primitive de g qui vaut 1 en $+\infty$. Sur l'intervalle $]-\infty\,;0]$, la fonction G est nulle donc est bien la primitive nulle en $-\infty$ de g. Finalement, G est bien la fonction de répartition de T.

7. En déduire la valeur de $\mathbb{P}(T \leq 3)$ et de $\mathbb{P}(T > 9)$. **/1**
Solution : Par définition de la fonction de répartition,

$$\mathbb{P}(T \leq 3) = G(3) = 1 - (1+1)\,e^{-1} \approx 0,264.$$

Puis, en passant par l'événement contraire,

$$\mathbb{P}(T > 9) = 1 - \mathbb{P}(T \leq 9) = 1 - G(9) = 1 - \left(1 - (3+1)e^{-3}\right) = 4e^{-3} \approx 0,199.$$

Exercice 3 Eugénie, myrmécologue **5 points**

Eugénie, scientifique spécialiste des fourmis, a découvert une nouvelle espèce de fourmis. Elle souhaite en déterminer les principales caractéristiques et commence par étudier leurs tailles. Elle prélève un échantillon de $n = 25$ fourmis et les mesure. Pour i entre 1 et n, la taille (en mm) de la fourmi numéro i est notée X_i. On admet que les variables aléatoires X_1, X_2, ..., X_n sont indépendantes et ont la même loi normale $\mathcal{N}(m, \sigma^2)$.

1. Quelle est la loi (nom et paramètre(s) en fonction de m, σ et n) de la variable aléatoire $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$? **/1,5**

Solution : Comme $\overline{X}$ est une combinaison linéaire de variables aléatoires gaussiennes indépendantes, elle est encore gaussienne. Il ne reste plus qu'à chercher son espérance et sa variance pour connaître ses paramètres.
Par linéarité de l'espérance, on a

$$\mathbb{E}\left[\overline{X}\right] = \frac{1}{n}\left(\mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n]\right) = \frac{1}{n} \times n\mathbb{E}[X_1] = \mathbb{E}[X_1],$$

car, les variables aléatoires X_1, X_2, ..., X_n ayant la même loi, elles ont aussi la même espérance. Comme cette loi commune est la loi normale $\mathcal{N}(m; \sigma^2)$, on a $\mathbb{E}[X_1] = m$, donc également $\mathbb{E}[\overline{X}] = m$.
Pour la variance, comme les variables aléatoires X_1, ..., X_n sont indépendantes, on a

$$\mathbb{V}\mathrm{ar}(\overline{X}) = \frac{1}{n^2}\left(\mathbb{V}\mathrm{ar}(X_1) + \mathbb{V}\mathrm{ar}(X_2) + \cdots + \mathbb{V}\mathrm{ar}(X_n)\right).$$

Comme ces variables ont la même loi, elles ont en particulier la même variance, égale à σ^2. Finalement,

$$\mathbb{V}\mathrm{ar}(\overline{X}) = \frac{1}{n^2} \times n\sigma^2 = \frac{\sigma^2}{n}.$$

Finalement, la loi de $\overline{X}$ est $\mathcal{N}(m, \frac{\sigma^2}{n})$.

2. On suppose dans cette question que le paramètre m est inconnu mais que le paramètre σ est connu et vaut 2.

   a) Les données mesurées par Eugénie sont telles que

$$\sum_{i=1}^{25} X_i = 292,8.$$

   Donner une estimation du paramètre m. **/0,5**

   Solution : Comme m est l'espérance de la loi $\mathcal{N}(m, \sigma^2)$, un estimateur du paramètre m est la moyenne empirique $\overline{X}$. On trouve donc ici comme estimation de m,

$$\overline{X} = \frac{1}{25}\sum_{i=1}^{25} X_i = \frac{292,8}{25} = 11,712.$$

   b) Déterminer un intervalle de confiance au niveau de confiance 95% pour la taille moyenne des fourmis de cette nouvelle espèce. **/1**

   Solution : Un intervalle de confiance de niveau 95% pour m est donné par

$$[\overline{X} - z_{0,95}\frac{\sigma}{\sqrt{n}} ; \overline{X} + z_{0,95}\frac{\sigma}{\sqrt{n}}],$$

   où $z_{0,95}$ est le réel tel que

$$\mathbb{P}\left(-z_{0,95} \leq N \leq z_{0,95}\right) = 0,95.$$

   Par symétrie, $z_{0,95}$ est aussi tel que

$$\mathbb{P}\left(N \leq z_{0,95}\right) = 0,975.$$

   En regardant le tableau de la fonction de répartition, ou en le connaissant par cœur, on a $z_{0,95} \approx 1,96$. Finalement, on trouve l'intervalle $[10,928 ; 12,496]$.

3. On suppose maintenant que m et σ sont tous les deux inconnus.

   a) Les données mesurées par Eugénie sont telles que

$$\sum_{i=1}^{25} X_i^2 = 3545.$$

Donner une estimation de la variance σ^2 (deux possibilités, en choisir une), puis de l'écart-type σ. **/1**

Solution : Un estimateur de σ^2 est la variance empirique. *On rappelle qu'on peut choisir la version baisée ou la version débiaisée.* La version avec biais est

$$s_n^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \overline{X}^2.$$

Ici, on trouve

$$s_n^2 = 3545/25 - 11,712^2 \approx 4,629.$$

Si on veut la version débiaisée σ_n^2, on multiplie par $\frac{n}{n-1}$:

$$\sigma_n^2 = \frac{25}{24}s_n^2 \approx 4,822.$$

Pour l'estimation de l'écart-type, il suffit de prendre la racine carrée de l'estimation de la variance. En prenant, par exemple, la version sans biais, on trouve :

$$\hat{\sigma} \approx \sqrt{4,822} \approx 2,196.$$

b) Déterminer dans ce cadre (m et σ inconnus) un intervalle de confiance au niveau de confiance 95% du paramètre m, c'est-à-dire de la taille moyenne des fourmis de cette nouvelle espèce. **/1**

Solution : L'intervalle de confiance pour m a la même forme sauf qu'on remplace les quantiles de la loi normale par ceux de la loi de Student et qu'on remplace l'écart-type (maintenant inconnu) par l'estimateur $\hat{\sigma}$. On cherche la valeur $t_{0.95}$ qui est telle que

$$\mathbb{P}\left(-t_{0,95} \leq T \leq t_{0,95}\right) \approx 0,95,$$

où T est la loi de Student à $n-1 = 24$ degrés de liberté. Par symétrie, cela revient à trouver $t_{0.95}$ tel que

$$\mathbb{P}\left(T \leq t_{0.95}\right) \approx 0,975.$$

La table des quantiles de la loi de Student donne

$$t_{0,95} \approx 2,0639.$$

Finalement, l'intervalle de confiance cherché est

$$\begin{aligned}&[\overline{X} - t_{0,95}\frac{\hat{\sigma}}{\sqrt{n}}\,;\overline{X} + t_{0,95}\frac{\hat{\sigma}}{\sqrt{n}}]\\ &= [11,712 - 2,0639 \times \frac{2,196}{5}\,; 11,712 + 2,0639 \times \frac{2,196}{5}]\\ &= [10,806\,; 12,618].\end{aligned}$$

Annexes

x	$\mathbb{P}\left(\mathcal{N}(0;1) \leq x\right)$
0	$0,5$
1	$0,841$
$1,645$	$0,950$
$1,960$	$0,975$
$2,576$	$0,995$

TABLE 1 – Quelques valeurs de la fonction de répartition de la loi normale centrée réduite

Nombre n de degrés de libertés	t tel que $\mathbb{P}(\mathcal{T}(n) \leq t) = 0,95$	t' tel que $\mathbb{P}(\mathcal{T}(n) \leq t') = 0,975$
11	$1,7959$	$2,2010$
12	$1,7823$	$2,1788$
13	$1,7709$	$2,1604$
20	$1,7247$	$2,0860$
24	$1,7109$	$2,0639$
25	$1,7081$	$2,0595$
26	$1,7056$	$2,0555$

Table 2 – Quelques valeurs de la fonction quantile de la loi de Student $\mathcal{T}(n)$, pour différentes valeurs du nombre de degrés de liberté n