# ESSEC II 2013 voie E

L'objet du problème est l'étude de la concentration en un type de bactéries d?un bassin destiné à la baignade.

Une municipalité doit effectuer un prélèvement et l'analyser afin de décider d'autoriser ou non l?utilisation du bassin. Dans une première partie, on étudiera des propriétés reliant la loi binomiale et la loi de Poisson. Dans une deuxième partie, on regardera la modélisation de la concentration en bactéries du bassin. Enfin, la troisième partie étudiera le principe d?un test destiné à prendre une décision d'utilisation.

**Les trois parties peuvent être traitées indépendamment en admettant les résultats des parties précédentes.**

Toutes les variables aléatoires sont définies sur un même espace probabilisé $(\Omega; \mathcal{A}; P)$. Si elles existent, on note $E(T)$ et $V(T)$ l'espérance et la variance d'une variable aléatoire $T$.

### I. Lien entre loi binomiale et loi de Poisson.

1. Pour tout entier $n$ strictement positif, on se donne un réel $p_n$ strictement positif et $n$ variables aléatoires $(X_k)_{k\in[\![1;n]\!]}$ indépendantes et suivant une loi de Bernoulli de paramètre $p_n$. On suppose que $np_n$ a une limite finie strictement positive et on pose $\lim\limits_{n\to+\infty} np_n = \lambda$ .

   (a) Quelle est la loi de $S_n = \displaystyle\sum_{k=1}^{n} X_k$ ?

   (b) Soit $k$ un entier naturel.

   i. Donner l'expression de $P(S_n = k)$ pour $n$ supérieur ou égal à $k$.

   ii. Que peut-on dire de la limite de $p_n$ quand $n$ tend vers l'infini ?
   Étudier la limite de $(1 - p_n)^n$ quand $n$ tend vers l'infini.

   iii. Montrer que $\lim\limits_{n\to+\infty} P(S_n = k) = e^{\lambda} \dfrac{\lambda^k}{k!}$ .

   (c) On pose $N_n = \max(X_1; \ldots; X_n)$ .

   i. Quelles sont les valeurs que peut prendre la variable aléatoire $N_n$ ?

   ii. Calculer $\lim\limits_{n\to+\infty} P(N_n = 0)$.

   iii. En déduire la limite en loi de la suite $(N_n)_{n\in\mathbb{N}^*}$. (Autrement dit reconnaître la loi derrière la limite précédente).

2. Soit $\lambda$ un réel strictement positif et $n$ un entier strictement positif tels que $0 < \lambda \leqslant n$ . On considère $U$ une variable aléatoire de loi de Bernoulli de paramètre $\dfrac{\lambda}{n}$ . On a donc $P(U = 1) = \dfrac{\lambda}{n}$ , $P(U = 0) = 1 - \dfrac{\lambda}{n}$ et pour tout $i$ entier naturel supérieur ou égal à 2, $P(U = i) = 0$ .

   (a) Montrer que pour tout réel positif $u$, $1 - u \leqslant e^{-u}$ .

   (b) Montrer que $\displaystyle\sum_{k=0}^{+\infty} \left| P(U = k) - \left(\frac{\lambda}{n}\right)^k \frac{1}{k!} e^{-\frac{\lambda}{n}} \right| = \frac{2\lambda}{n} \left(1 - e^{-\frac{\lambda}{n}}\right)$ .

   (c) En déduire que $\displaystyle\sum_{k=0}^{+\infty} \left| P(U = k) - \left(\frac{\lambda}{n}\right)^k \frac{1}{k!} e^{-\frac{\lambda}{n}} \right| \leqslant 2 \left(\frac{\lambda}{n}\right)^2$ .

3. Soit $\lambda$ un réel strictement positif et $n$ un entier strictement positif tels que $0 < \lambda \leqslant n$. Soient $Z$, $U$ et $V$ trois variables aléatoires indépendantes à valeurs dans $\mathbb{N}$. On suppose que $U$ suit une loi de Bernoulli de paramètre $\dfrac{\lambda}{n}$ et que $V$ suit une loi de Poisson de paramètre $\dfrac{\lambda}{n}$ . On observe en particulier que pour tout entier $p$ strictement négatif $P(Z = p) = P(U = p) = P(V = p) = 0$ .

(a) Soit un entier naturel fixé. En remarquant que pour tout réels $a$ et $b$, on a: $|a-b| \leqslant |a|+|b|$ , montrer que la série de terme général $|P(U=k-i)-P(V=k-i)|$ , où $k$ décrit $\mathbb{N}$ , est convergente. On note $A_i$ sa somme, $A_i = \sum_{k=0}^{+\infty} |P(U=k-i)-P(V=k-i)|$ .

Montrer que pour tout $i$ on a : $A_i \leqslant 2\left(\frac{\lambda}{n}\right)^2$.

(b) Montrer que la série de terme général $A_i P(Z=i)$ est convergente.

(c) Soit $k$ un entier naturel fixé.
Montrer que la série de terme général $|P(U=k-i)-P(V=k-i)| \times P(Z=i)$ , où $i$ décrit $\mathbb{N}$ , est convergente.

(d) On pose $B_k = \sum_{k=0}^{+\infty} |P(U=k-i)-P(V=k-i)| \times P(Z=i)$ .

i. Montrer que , pour $k \geqslant 2$ , on a
$$B_k = \left|1-\frac{\lambda}{n}-e^{-\frac{\lambda}{n}}\right| P(Z=k) + \left|\frac{\lambda}{n}-\frac{\lambda}{n}e^{-\frac{\lambda}{n}}\right| P(Z=k-1) + \sum_{i=0}^{k-2} P(V=k-i)\, P(Z=i) .$$

ii. Justifier que pour $k \geqslant 2$, on a $\sum_{i=0}^{k-2} P(V=k-i)\, P(Z=i) \leqslant P(V+Z=k)$ .

iii. Montrer finalement que la série de terme général $B_k$ , où $k$ décrit $\mathbb{N}$ , est convergente.

On admettra alors qu'on a : $\sum_{k=0}^{+\infty} B_k = \sum_{i=0}^{+\infty} A_i P(Z=i)$, c'est à dire :

$$\sum_{k=0}^{+\infty}\left(\sum_{i=0}^{+\infty} |P(U=k-i)-P(V=k-i)| \times P(Z=i)\right) = \sum_{i=0}^{+\infty}\left(\sum_{k=0}^{+\infty} |P(U=k-i)-P(V=k-i)|\right) \times P(Z=i)$$

4. On conserve dans cette question les notations et les hypothèses de la question 3 concernant les variables aléatoires $Z$ , $U$ et $V$.

(a) Montrer que la série de terme général: $|P(Z+U=k)-P(Z+V=k)|$ est convergente.

(b) Déduire de la question 3 que
$$\sum_{k=0}^{+\infty} |P(Z+U=k)-P(Z+V=k)| \leqslant \sum_{i=0}^{+\infty}\left(\sum_{k=0}^{+\infty} |P(U=k-i)-P(V=k-i)|\right) P(Z=i) ,$$
puis que
$$\sum_{k=0}^{+\infty} |P(Z+U=k)-P(Z+V=k)| \leqslant 2\left(\frac{\lambda}{n}\right)^2 .$$

5. Soient $U_1;\ldots;U_n;V_1;\ldots V_n$ , $2n$ variables aléatoires indépendantes, telles que pour tout entier $i$ tel que $1 \leqslant i \leqslant n$, $U_i$ suit une loi de Bernoulli de paramètre $\frac{\lambda}{n}$ et $V_i$ suit une loi de Poisson de paramètre $\frac{\lambda}{n}$ .

(a) Montrer que pour tout entier naturel $k$, on a
$|P(U_1+\cdots+U_n=k)-P(V_1+\cdots+V_n=k)| \leqslant$
$|P(U_1+\cdots+U_n=k)-P(U_1+\cdots+U_{n-1}+V_n=k)| +$
$|P(U_1+\cdots+U_{n-1}+V_n=k)-P(U_1+\cdots+U_{n-2}+V_{n-1}+V_n=k)| +$
$\cdots + |P(U_1+V_2+\cdots+V_n=k)-P(V_1+\cdots+V_n=k)|$.

(b) En déduire que
$$\sum_{k=0}^{-\infty} |P(U_1+\cdots+U_n=k)-P(V_1+\cdots+V_n=k)| \leqslant 2\frac{\lambda^2}{n} .$$

(c) Soit $X$ une variable aléatoire de loi binomiale de paramètres $n$ et $\frac{\lambda}{n}$ . Conclure de ce qui précède que
$$\sum_{k=0}^{+\infty} \left|P(X=k)-\frac{\lambda^k}{k!}e^{-\lambda}\right| \leqslant 2\frac{\lambda^2}{n} .$$

## II. Modélisation de la concentration en bactéries.

Le bassin qu'on étudie est supposé de volume $V$ , en $\mathrm{m}^3$ , et on effectue un prélèvement de volume $\Delta V$ , en $\mathrm{m}^3$. Dans cette situation, la probabilité pour qu'une bactérie spécifique du bassin se trouve dans le prélèvement est égale à $\dfrac{\Delta V}{V}$.

Supposons que 1e bassin contienne $n$ bactéries numérotées de 1 à $n$ , $n \in \mathbb{N}^*$. On considère alors $n$ variables aléatoires $X_1; \ldots; X_n$ à valeurs 0 ou 1 telles que $X_i = 1$ si la bactérie $i$ se trouve dans le prélèvement et 0 sinon. Les variables en question sont supposées indépendantes.

On pose $c = \dfrac{n}{V}$ qui représente 1a concentration en bactéries du bassin par $\mathrm{m}^3$.

1. (a) Quelle est la loi de $X_i$ pour $i \in [\![1; n]\!]$ ?

   (b) Soit $N$ le nombre de bactéries présentes dans le prélèvement.
   Montrer que $N$ suit une loi binomiale de paramètres $n$ et $\dfrac{\Delta V}{V}$

2. Soit $U$ une variable aléatoire qui suit la loi de Poisson de paramètre $\lambda$. On appelle $F$ sa fonction de répartition.
   Écrire en Python une fonction d'en-tête `def Poisson(x, lambda):`
   qui a pour résultat $F(x)$. On s'attachera à minimiser le nombre d'opérations.

3. Soit $U$ une variable aléatoire de loi de Poisson de paramètre $c \triangle V$. Soit $K \in \mathbb{N}^*$ fixé.

   (a) Montrer que $|P(N \leqslant K) - P(U \leqslant K)| \leqslant \dfrac{2c(\triangle V)^2}{V}$ .

   (b) On suppose que $V = 1000$ , $V$ volume en $m^3$, et que le prélèvement est de volume est égal à 1 litre soit $\Delta V = 10^{-3}$.
   Trouver un majorant de l'erreur commise en approximant $P(N \leqslant K)$ par $P(U \leqslant K)$ .

   (c) On suppose que la concentration dans le bassin reste inférieure à $10^6$ bactéries par mètre cube, et que le prélèvement réalisé est encore de 1 litre.
   Quelle est la valeur minimale du volume $V$ garantissant que l'erreur commise dans l'approximation précédente soit inférieure à $10^{-6}$ .
   On admettra que le résultat de 1a question 5. c) peut être amélioré de la façon suivante :
   $\displaystyle\sum_{k=0}^{+\infty} \left| P(X = k) - \frac{\lambda^k}{k!} e^{-\lambda} \right| \leqslant 2\lambda \frac{\min(2; \lambda)}{n}$ .

   (d) Montrer que si $U$ suit une loi de Poisson de paramètre $c\Delta V$, $|P(N \leqslant K) - P(U \leqslant K)| \leqslant 4\dfrac{\triangle V}{V}$ .

   (e) On suppose toujours que le prélèvement réalisé est de un litre. À l'aide de l?inégalité précédente, trouver la valeur minimale du volume $V$ garantissant que l'erreur commise dans l'approximation précédente soit inférieure a $10^{-6}$.

## III. Construction d?une procédure de test.

Un prélèvement est réalisé et mis en culture pendant 24 heures. Les bactéries se multiplient et on obtient ainsi une concentration plus importante et plus facile à estimer. On conserve les notations de la partie précédente et on posera $\lambda = c\Delta V$. En outre, on définit la fonction $h : \mathbb{R} \to \mathbb{R}$ , par $h(x) = \begin{cases} 0 & \text{si } x \leqslant -1 \\ (1+x)\ln(1+x) - x & \text{si } x > 1 \end{cases}$ .

1. Déterminer les variations de la fonction $h$sur l'intervalle$]1; +\infty[$ et préciser le signe de $h$.

2. Démontrer l'inégalité de Markov :
   *Si $Y$ est une variable aléatoire réelle à valeurs discrètes positives admettant une espérance $E(Y)$, pour tout réel $\alpha$ strictement positif, on a $P(Y \geqslant \alpha) \leqslant \dfrac{1}{\alpha} E(Y)$.*

3. Soit $\lambda$ un réel strictement positif et soit $X$ une variable aléatoire qui suit une loi de Poisson de paramètre$\lambda$. Soit $\varepsilon$ un réel strictement positif.

   (a) Pour tout réel $u$ , calculer $E\left(e^{uX}\right)$.

   (b) Trouver un réel strictement positif $u_0$ tel que $E\left(e^{u_0(X-(1+\varepsilon)\lambda)}\right) = e^{-\lambda h(\varepsilon)}$ .

   (c) i. Montrer que pour tout $u > 0$ , $P\left(\lambda^{-1}(X - \lambda) \geqslant \varepsilon\right) = P\left(e^{u(X-\lambda)} \geqslant e^{\lambda\varepsilon u}\right)$

   ii. En déduire $P\left(\lambda^{-1}(X - \lambda) \geqslant \varepsilon\right) \leqslant e^{-\lambda h(\varepsilon)}$ .

   (d) Montrer de même que $P\left(\lambda^{-1}(X - \lambda) \leqslant -\varepsilon\right) \leqslant e^{-\lambda h(-\varepsilon)}$ (on traitera successivement les cas a $\varepsilon \geqslant 1$ et $\varepsilon \in ]0; 1[$ .

4. Si la concentration en bactéries est trop élevée, on doit interdire le bassin. La limite maximale de tolérance est fixée à 2000 bactéries par litre. On garde les notations de 1a question III.3. et on suppose de nouveau que $\Delta V = 10^{-3}$ . Fixons un réel $\alpha$ compris entre 0 et 2000.

   (a) Montrer que pour tout réel $\lambda$ strictement supérieur à 2000,
   $P(X \leqslant \alpha) = P\left(\lambda^{-1}(X-\lambda) \leqslant \lambda^{-1}(\alpha-\lambda)\right) \leqslant e^{-\lambda h\left(\lambda^{-1}\alpha-1\right)}$

   (b) En déduire que si $\lambda$ est un réel strictement supérieur à 2000, $P(X \leqslant \alpha) \leqslant e^{-2000\, h\left(\frac{\alpha}{2000}-1\right)}$.

   (c) Montrer que l'équation d'inconnue $x$ , $2000\, h\left(\dfrac{x}{2000}-1\right) = \ln(100)$admet une unique solution $\alpha_0$comprise entre 0 et 2000.

   (d) On admettra que $1865 < \alpha_0$ . Déduire que si $\lambda > 2000$ , $P(X < 1865) \leqslant \dfrac{1}{100}$ .

   (e) Exemple d'application : on compte 1600 bactéries dans [e prélèvement de l litre. Que peut-on dire du risque que l'on prend en autorisant le bassin ?