MATHÉMATIQUES II - ESSEC E 2013

Proposition de corrigé par David Meneu

Lycée Champollion - Grenoble, pour Major-Prépa

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 pouvaient être traitées indépendamment en admettant les résultats des parties précédentes.

Toutes les variables aléatoires sont définies sur le 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 la loi de Poisson

1) Pour tout entier n strictement positif, on se donne un réel p_n strictement positif et n variables aléatoires $(X_k)_{1\leqslant k\leqslant n}$ 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) La variable aléatoire $S_n = \sum\limits_{k=1}^{n} X_k$ a pour valeur le nombre de variables aléatoires X_k qui prennent la valeur 1, c'est-à-dire que S_n compte le nombre de succès obtenus en n épreuves indépendantes de même probabilité de succès p_n : le cours dit que S_n **suit donc la loi binomiale de paramètres** (n, p_n).

b) Soit k un entier naturel.

i) $S_n(\Omega) = [\![0, n]\!]$ et si $0 \leqslant k \leqslant n$, $P(S_n = k) = \binom{n}{k}(p_n)^k(1-p_n)^{n-k}$.

ii) On sait que $\lim\limits_{n\to+\infty} np_n = \lambda > 0$, donc $\lim\limits_{n\to+\infty} \frac{np_n}{\lambda} = 1$, c'est-à-dire que : $p_n \underset{n\to+\infty}{\sim} \frac{\lambda}{n}$.

Comme deux suites équivalentes ont la même limite, on a : $\boxed{\lim\limits_{n\to+\infty} p_n = \lim\limits_{n\to+\infty} \frac{\lambda}{n} = 0}$.

La limite de $(1-p_n)^n$ quand n tend vers l'infini, s'étudie classiquement en revenant à la forme exponentielle de l'expression :

$(1-p_n)^n = e^{n\ln(1-p_n)}$, où, puisque $\lim\limits_{n\to+\infty} p_n = 0$: $\ln(1-p_n) \underset{n\to+\infty}{\sim} -p_n \Longrightarrow n\ln(1-p_n) \underset{n\to+\infty}{\sim} -np_n$, et :

$\lim\limits_{n\to+\infty} n\ln(1-p_n) = \lim\limits_{n\to+\infty} -np_n = -\lambda$, d'où : $\boxed{\lim\limits_{n\to+\infty} (1-p_n)^n = \lim\limits_{n\to+\infty} e^{n\ln(1-p_n)} = e^{-\lambda}}$ par continuité de l'exponentielle sur $\mathbb{R}$.

iii) Le calcul initié en *ii)* correspond au résultat de cours sur l'approximation d'une loi binomiale par une loi de Poisson !

Pour tout entier k fixé et tout entier $n \geqslant k$, on a :

$$P(S_n = k) = \binom{n}{k}(p_n)^k.(1-p_n)^{n-k} = \frac{n!}{k!(n-k)!}.(p_n)^k(1-p_n)^{n-k}$$
$$= \frac{1}{k!}.\left(\frac{1}{1-p_n}\right)^k.n.(n-1)...(n-k+1).(p_n)^k.(1-p_n)^n$$

où :

$\star$ $\lim\limits_{n\to+\infty} \left(\frac{1}{1-p_n}\right)^k = 1^k = 1$ vu que $\lim\limits_{n\to+\infty} p_n = 0$,

⋆ $\lim_{n\to+\infty}(1-p_n)^n = e^{-\lambda}$ comme on l'a vu en *ii*),

⋆ $n.(n-1)...(n-k+1) \underset{n\to+\infty}{\sim} n^k$ puisque ce produit de k facteurs a pour coefficient dominant 1, d'où : $n.(n-1)...(-k+1).(p_n)^k \underset{n\to+\infty}{\sim} n^k.(p_n)^k = (np_n)^k$, et :
$\lim_{n\to+\infty} n.(n-1)...(n-k+1).(p_n)^k = \lim_{n\to+\infty}(np_n)^k = \lambda^k$ (k est fixé).

Par produit de limites, on obtient donc bien : $\boxed{\forall k\in\mathbb{N},\ \lim_{n\to+\infty} P(S_n=k) = e^{-\lambda}.\frac{\lambda^k}{k!}}$.

Rappelons que ce résultat exprime la *convergence en loi* de la suite $(S_n)_{n\in\mathbb{N}}$, vers la loi de Poisson $\mathcal{P}(\lambda)$.

c) On pose $N_n = \max(X_1,\cdots,X_n)$.

i) Pour tout entier $i\in[\![1,n]\!]$, $X_i(\Omega)=\{0;1\}$ (variables de Bernoulli), donc les seules valeurs possibles de N_n sont 0 (si et seulement si toutes les variables X_i prend la valeur 0), et 1 (dès que l'une des n variables $X_1,\cdots,X_n$ prennent la valeur 1) : $\boxed{N_n(\Omega)=\{0,1\}}$.

ii) Comme on le voit en déterminant l'univers-image de N_n, l'événement $[N_n=0]$ s'écrit :

$$[N_n=0]=[X_1=0]\cap[X_2=0]\cap\cdots\cap[X_n=0].$$

L'indépendance des n variables $(X_k)_{1\leqslant k\leqslant n}$ permet d'écrire :
$P(N_n=0)=P(X_1=0).P(X_2=0)\cdots P(X_n=0)=(1-p_n)^n$ et ainsi, selon les calculs déjà faits :

$$\boxed{\lim_{n\to+\infty} P(S_n=0)=e^{-\lambda}.}$$

iii) La suite $(N_n)_{n\in\mathbb{N}^*}$ est constituée de variables de Bernoulli, toutes d'univers-image $\{0;1\}$, où :
$\lim_{n\to+\infty} P(S_n=0)=e^{-\lambda}$ et par conséquent :
$\lim_{n\to+\infty} P(S_n=1)=\lim_{n\to+\infty} 1-P(S_n=0)=1-e^{-\lambda}$.
Par conséquent, **la suite $(N_n)_{n\in\mathbb{N}^*}$ converge en loi vers une variable aléatoire qui suit la loi de Bernoulli de paramètre** $\boxed{1-e^{-\lambda}.}$

2) Soit λ 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 $\frac{\lambda}{n}$. On a donc $P(U=1)=\frac{\lambda}{n}$, $P(U=0)=1-\frac{\lambda}{n}$ et pour tout i entier naturel supérieur ou égal à 2, $P(U=i)=0$.

(a) Une inégalité très classique dans cette question : montrons que pour tout réel u positif, $1-u\leqslant e^{-u}$.
On peut le faire en étudiant les variations, puis le signe de la fonction différence $u\mapsto e^{-u}-1+u$, où plus rapidement par un argument de convexité :
$g:\ u\mapsto e^{-u}$ est de classe $\mathcal{C}^2$ sur $\mathbb{R}^+$, avec : $\forall u\in\mathbb{R}^+,\ g'(u)=-e^{-u}$ et $g''(u)=-(-e^{-u})=e^{-u}>0$.
g est donc convexe sur $\mathbb{R}^+$, et sa courbe se trouve au-dessus de toutes ses tangentes, en particulier celle au point d'abscisse $u=0$, qui a pour équation : $y=g'(0).(u-0)+g(0) \iff y=-u+1$.
Ainsi : $\forall u\in\mathbb{R}^+,\ e^{-u}\geqslant 1-u$.

(b) Vérifions que la série $\sum_{k\geqslant 0}\left|P(U=k)-\left(\frac{\lambda}{n}\right)^k\frac{1}{k!}e^{-\frac{\lambda}{n}}\right|$ converge et calculons sa somme :

Pour tout entier $n\geqslant 2$, on a :

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

puisque $P(U=k)=0$ si $k\geqslant 2$, ce qu'on réécrit :

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

On reconnaît une série exponentielle, toujours convergente, donc la série proposée converge bien et a pour somme :

$$\sum_{k=0}^{+\infty}\left|P(U=k)-\left(\frac{\lambda}{n}\right)^k\frac{1}{k!}e^{-\frac{\lambda}{n}}\right|=\left|1-\frac{\lambda}{n}-e^{-\frac{\lambda}{n}}\right|+\left|\frac{\lambda}{n}-\frac{\lambda}{n}.e^{-\frac{\lambda}{n}}\right|+e^{-\frac{\lambda}{n}}.\left(e^{\frac{\lambda}{n}}-1-\frac{\lambda}{n}\right),\text{ où :}$$

$\left|1-\dfrac{\lambda}{n}-e^{-\frac{\lambda}{n}}\right| = e^{-\frac{\lambda}{n}}-1+\dfrac{\lambda}{n}$, puisque d'après (a) : $1-\dfrac{\lambda}{n} \leqslant e^{-\frac{\lambda}{n}}$ avec $u=\dfrac{\lambda}{n}>0$,
$\left|\dfrac{\lambda}{n}-\dfrac{\lambda}{n}.e^{-\frac{\lambda}{n}}\right| = \dfrac{\lambda}{n}.\left|1-e^{-\frac{\lambda}{n}}\right| = \dfrac{\lambda}{n}.\left(1-e^{-\frac{\lambda}{n}}\right)$ puisque $e^{-\frac{\lambda}{n}} \leqslant 1$ $(-\dfrac{\lambda}{n}<0)$. Finalement :

$$\sum_{k=0}^{+\infty}\left|P(U=k)-\left(\frac{\lambda}{n}\right)^k\frac{1}{k!}e^{-\frac{\lambda}{n}}\right| = e^{-\frac{\lambda}{n}}-1+\frac{\lambda}{n}+\frac{\lambda}{n}-\frac{\lambda}{n}.e^{-\frac{\lambda}{n}}+1-e^{-\frac{\lambda}{n}}-\frac{\lambda}{n}.e^{-\frac{\lambda}{n}}\boxed{=\frac{2\lambda}{n}.[1-e^{-\frac{\lambda}{n}}]}$$

(c) L'inégalité : $1-\dfrac{\lambda}{n} \leqslant e^{-\frac{\lambda}{n}}$ utilisée en (b) et obtenue à partir de (a) se réécrit aussi : $1-e^{-\frac{\lambda}{n}} \leqslant \dfrac{\lambda}{n}$.
On obtient donc, en effet : $\displaystyle\sum_{k=0}^{+\infty}\left|P(U=k)-\left(\frac{\lambda}{n}\right)^k\frac{1}{k!}e^{-\frac{\lambda}{n}}\right| = \frac{2\lambda}{n}.[1-e^{-\frac{\lambda}{n}}] \leqslant \frac{2\lambda}{n}.\frac{\lambda}{n} = 2\left(\frac{\lambda}{n}\right)^2$.

3) Soit λ 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 $\frac{\lambda}{n}$ et que V suit une loi de Poisson de paramètre $\frac{\lambda}{n}$. On observe en particulier que pour tout entier p strictement négatif :

$$P(Z=p)=P(U=p)=P(V=p)=0$$

(a) Soit i un entier naturel fixé. Pour tous réels a et b, l'*inégalité triangulaire* donne en effet :

$$|a-b|=|a+(-b)|\leqslant |a|+|-b|=|a|+|b|.$$

Pour tout entier $k\in\mathbb{N}$, on a alors :
$|P(U=k-i)-P(V=k-i)|=0$ si $k<i$ car alors $k-i<0$, et sinon :
$\forall k\geqslant i,\ |P(U=k-i)-P(V=k-i)|\leqslant |P(U=k-i)|+|P(V=k-i)|=P(U=k-i)+P(V=k-i)$
d'après l'inégalité triangulaire, les probabilités étant toujours positives.
Or :
$\displaystyle\sum_{k\geqslant i}P(U=k-i)=\sum_{j\geqslant 0}P(U=j)$ est une série convergente : seuls ses deux premiers termes sont non-nuls ! La série $\displaystyle\sum_{k\geqslant i}P(V=k-i)=\sum_{j\geqslant 0}P(V=j)$ est aussi convergente, de somme 1 :
la série $\displaystyle\sum_{k\geqslant 0}P(U=k-i)+P(V=k-i)$ est donc convergente (somme de deux séries convergentes).
L'inégalité précédente et le *théorème de comparaison des séries à termes positifs* assure alors que la série $\displaystyle\sum_{k\geqslant 0}|P(U=k-i)-P(V=k-i)|$ est convergente.

On note A_i sa somme : $\displaystyle A_i=\sum_{k=0}^{+\infty}|P(U=k-i)-P(V=k-i)|=\sum_{k=i}^{+\infty}|P(U=k-i)-P(V=k-i)|$
$\displaystyle\overset{[j=k-i]}{=}\sum_{j=0}^{+\infty}|P(U=j)-P(V=j)|=\sum_{j=0}^{+\infty}\left|P(U=j)-\frac{1}{k!}\left(\frac{\lambda}{n}\right)^k.e^{-\frac{\lambda}{n}}\right|\leqslant 2\left(\frac{\lambda}{n}\right)^2$ en effet, lorsqu'on prend en compte la nullité des termes pour $k<i$, et qu'on constate que la somme correspond au résultat obtenu en 2)(c) vu la loi de V.

(b) Pour tout entier i, on a : $0\leqslant A_i\times P(Z=i)\leqslant 2.\left(\dfrac{\lambda}{n}\right)^2.P(Z=i)$, où : $2.\left(\dfrac{\lambda}{n}\right)^2$ est indépendant de l'indice i (considéré donc comme une constante), et $P(Z=i)$ le terme général d'une série convergente, de somme 1 puisque $Z(\Omega)\subset\mathbb{N}$.
On utilise donc encore une fois, le théorème de comparaison des séries à termes positifs, pour conclure que la série $\displaystyle\sum_{i\geqslant 0}A_i\times P(Z=i)$ est convergente.

(c) Soit k un entier naturel fixé. Lorsqu'on prend un seul terme $|P(U=k-i)-P(V=k-i)|$ de la série à termes positifs étudiée en (a), ce terme unique est bien sûr inférieur à la somme de tous :
$\forall i\in\mathbb{N},\ |P(U=k-i)-P(V=k-i)|\leqslant A_i \Longrightarrow |P(U=k-i)-P(V=k-i)|\times P(Z=i)\leqslant A_i\times P(Z=i)$
(on multiplie par une probabilité, toujours positive).
La série $\displaystyle\sum_{i\geqslant 0}A_i\times P(Z=i)$ étant convergente d'après (b), le théorème de comparaison des séries à termes positifs, permet une fois de plus de conclure que la série $|P(U=k-i)-P(V=k-i)|\times P(Z=i)$, est

convergente.

Remarque : en fait, on peut remarquer bien plus simplement que puisque $P(U = k-i) = P(V = k-i) = 0$ dès que $k - i < 0 \iff i > k$, les termes de la série sont nuls à partir du rang k, et la série converge forcément puisque sa somme ne comporte qu'un nombre fini (pour $0 \leqslant i \leqslant k$) de termes non-nuls !

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

i) Soit $k \geqslant 2$ un entier fixé : puisque U est une variable de Bernoulli, $P(U = k - i)$ est non-nul pour deux valeurs de i seulement : lorsque $k - i = 0 \iff i = k$, et lorsque $k - i = 1 \iff i = k - 1$. De même, et comme on l'a remarqué à la question précédente, $P(V = k - i)$ est nul dès que $k - i < 0 \iff i > k$.

La somme B_k ne comprend donc que les termes d'indices $0 \leqslant i \leqslant k$, et on réécrit :

$$B_k = \sum_{i=0}^{k} |P(U = k - i) - P(V = k - i)| \times P(Z = i)$$
$$= \sum_{i=0}^{k-2} |0 - P(V = k-i)| \times P(Z = i) + |P(U = 1) - P(V = 1)| \times P(Z = k-1) + |P(U = 0) - P(V = 0)| \times P(Z = k)$$
$$= \sum_{i=0}^{k-2} P(V = k - i) \times P(Z = i) + |\frac{\lambda}{n} - e^{-\frac{\lambda}{n}}.\frac{\lambda}{n}| + |1 - \frac{\lambda}{n} - e^{-\frac{\lambda}{n}}|,$$ ce qui est bien la forme demandée par l'énoncé.

ii) Pour tout entier $k \geqslant 2$, vu que V et Z sont à valeurs dans $\mathbb{N}$ l'événement $[V + Z = k]$ se décompose sous la forme :
$$[V + Z = k] = \bigcup_{i=0}^{k} [Z = i] \cap [V = k - i]$$
réunion d'événements incompatibles deux à deux, qui donne, en utilisant également l'indépendance des variables aléatoires V et Z :

$P(V + Z = k) = \displaystyle\sum_{i=0}^{k} P(Z = i) \times P(V = k - i) \geqslant \sum_{i=0}^{k-2} P(Z = i) \times P(V = k - i)$ puisque dans la somme de droite, on a écrit deux termes (positifs) de moins !

iii) On écrit donc finalement, d'après les questions précédentes :

$\forall k \geqslant 2,\ 0 \leqslant B_k \leqslant \left|1 - \dfrac{\lambda}{n} - e^{-\frac{\lambda}{n}}\right| .P(Z = k) + \left|\dfrac{\lambda}{n} - \dfrac{\lambda}{n}.e^{-\frac{\lambda}{n}}\right| .P(Z = k - 1) + P(V + Z = k)$, où :

La série dont le terme général est le membre de droite est convergente, comme somme de trois séries convergentes, puisqu'à des facteurs indépendants de k près, ce sont les lois de trois variables aléatoires discrètes à valeurs dans $\mathbb{N}$!

Une dernière fois, le théorème de comparaison des séries à termes positifs, permet de conclure que la série $\displaystyle\sum_{k\geqslant 0} B_k$ est convergente.

L'énoncé demandait alors d'admettre qu'on a : $\displaystyle\sum_{k=0}^{+\infty} B_k = \sum_{i=0}^{+\infty} A_i \times P(Z = i)$, c'est-à-dire :
$$\sum_{k=0}^{+\infty} \left(\sum_{i=0}^{+\infty} |P(U = k - i) - P(V = k - i)| \times P(Z = i) \right) = \sum_{i=0}^{+\infty} \left(\sum_{k=0}^{+\infty} |P(U = k - i) - P(V = k - i)| \right) \times P(Z = i)$$
c'est-à-dire qu'on peut intervertir les deux symboles sommes, dans la somme double écrite ci-dessus.

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

(a) La série $\displaystyle\sum_{k\geqslant 0} |P(Z + U = k) - P(Z + V = k)|$ est à termes positifs, et l'inégalité triangulaire donne :
$$\forall k \in \mathbb{N},\ |P(Z + U = k) - P(Z + V = k)| \leqslant P(Z + U = k) + P(Z + V = k),$$
où $\displaystyle\sum_{k\geqslant 0} P(Z + U = k) + P(Z + V = k)$ est une série convergente comme somme de deux séries convergentes : $Z + U$ et $Z + V$ sont en effet des v.a.r. discrètes à valeurs dans $\mathbb{N}$.

Le théorème de comparaison des séries à termes positifs permet une fois de plus de conclure que la série $\sum_{k\geqslant 0}|P(Z+U=k)-P(Z+V=k)|$,est convergente.

(b) Pour tout entier $k\in\mathbb{N}$, la formule des probabilités totales avec le système complet d'événements associé à la v.a.r. Z donne :

$P(Z+U=k)=\sum_{i=0}^{+\infty}P(Z+U=k\cap Z=i)=\sum_{i=0}^{+\infty}P(U=k-i\cap Z=i)=\sum_{i=0}^{+\infty}P(U=k-i)\times P(Z=i)$ puisque Z et U sont indépendantes.

On obtient de même : $P(Z+V=k)=\sum_{i=0}^{+\infty}P(V=k-i)\times P(Z=i)$, d'où :

$\forall k\in\mathbb{N},\ |P(Z+U=k)-P(Z+V=k)|=\left|\sum_{i=0}^{+\infty}[P(U=k-i)-P(V=k-i)]\times P(Z=i)\right|$

$=\left|\sum_{i=0}^{+\infty}[P(U=k-i)-P(V=k-i)]\right|\times P(Z=i)$; l'inégalité triangulaire généralisée (les séries concernées convergent) donne alors :

$\left|\sum_{i=0}^{+\infty}P(U=k-i)-P(V=k-i)\right|\leqslant\sum_{i=0}^{+\infty}|P(U=k-i)-P(V=k-i)|$, d'où (puisque $P(Z=i)\geqslant 0$) :

$\forall k\in\mathbb{N},\ |P(Z+U=k)-P(Z+V=k)|\leqslant\sum_{i=0}^{+\infty}|P(U=k-i)-P(V=k-i)|\times P(Z=i)$.

Dans cette inégalité, le membre de gauche est le terme général d'une série convergente d'après 4)(a), le membre de droite aussi : on a reconnu l'expression de B_k de la question 3). L'inégalité est donc conservée par passage à la somme quand k décrit $\mathbb{N}$, soit :

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

ce qui donne bien le résultat attendu après interversion des symboles sommes dans le membre de droite, selon le résultat admis en fin de question 3).

Enfi, en reprenant les notations de la question 3) :

$\forall i\in\mathbb{N},\ \sum_{k=0}^{+\infty}|P(U=k-i)-P(V=k-i)|\times P(Z=i)=A_i\times P(Z=i)\leqslant 2\left(\frac{\lambda}{n}\right)^2$ d'après 3)(a).

Par passage à la somme dans cette inégalité, les séries correspondantes étant convergentes :

$$\sum_{i=0}^{+\infty}\sum_{k=0}^{+\infty}|P(U=k-i)-P(V=k-i)|\times P(Z=i)=\sum_{i=0}^{+\infty}A_i\times P(Z=i)\leqslant 2\left(\frac{\lambda}{n}\right)^2\times\sum_{i=0}^{+\infty}P(Z=i)=2\left(\frac{\lambda}{n}\right)^2$$

On a bien obtenu au final :

$$\sum_{k=0}^{+\infty}|P(Z+U=k)-P(Z+V=k)|\leqslant 2\left(\frac{\lambda}{n}\right)^2$$

5) Soient $U_1,U_2,\cdots,U_n,\ V_1,V_2,\cdots,V_n$, $2n$ variables aléatoires indépendantes telles que pour tout entier $i\in[\![1,n]\!]$, 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) Introduisons ici les notations qui permettront de présenter simplement le raisonnement : pour tout entier $i\in[\![1;n]\!]$, on pose : $Z_i=\sum_{k<i}U_k+\sum_{k>i}V_k$, et $\alpha_{i,k}=P(Z_i+U_i=k)-P(Z_i+V_i=k)$.

On remarque ainsi que : $Z_n=U_1+\cdots+U_{n-1}$, que $Z_1=V_2+\cdots+V_n$, que $Z_i+V_i=Z_{i-1}+U_{i-1}$, et enfin que si par convention on pose : $Z_0=V_1+\cdots+V_n$ et $U_0=0$, on peut écrire :

$|P(U_1+\cdots+U_n=k)-P(V_1+\cdots+V_n=k)|=|P(Z_n+U_n=k)-P(Z_0+U_0=k)|$ qu'on peut voir comme le résultat du télescopage :

$|P(Z_n+U_n=k)-P(Z_0+U_0=k)|=\left|\sum_{k=1}^{n}P(Z_i+U_i=k)-P(Z_{i-1}+U_{i-1}=k)\right|=\left|\sum_{i=1}^{n}\alpha_{i,k}\right|\leqslant\sum_{i=1}^{n}|\alpha_{i,k}|$

selon l'inégalité triangulaire généralisée, ce qui est exactement la relation demandée par l'énoncé, au vu des notations choisies.

b) En reprenant les notations ci-dessus : Z_i est toujours une v.a.r. discrète à valeurs dans $\mathbb{N}$ (somme de v.a.r. ayant cette propriété), avec U_i et V_i ces trois v.a.r. vérifient donc les conditions et donc les propriétés établies en 3) et 4). Par conséquent et pour tout entier $i \in [\![1;n]\!]$, on peut écrire :

$$\sum_{k=0}^{+\infty} |\alpha_{i,k}| = \sum_{k=0}^{+\infty} |P(Z_i + U_i = k) - P(Z_i + V_i = k)| \leqslant 2\left(\frac{\lambda}{n}\right)^2.$$

Mais alors, d'après l'inégalité obtenue précédemment :

$$\sum_{k=0}^{+\infty} |P(U_1 + \cdots + U_n = k) - P(V_1 + \cdots + V_n = k)| \leqslant \sum_{k=0}^{+\infty}\sum_{i=1}^{n} |\alpha_{i,k}| = \sum_{i=1}^{n}\sum_{k=0}^{+\infty} |\alpha_{i,k}|$$
$$\leqslant \sum_{i=1}^{n} 2\left(\frac{\lambda}{n}\right)^2 = n \times 2\frac{\lambda^2}{n^2} = \frac{2\lambda^2}{n}$$

c) Soit maintenant X une variable aléatoire de loi binomiale de paramètres n et $\frac{\lambda}{n}$. Il existe donc n v.a.r. $U_1, \cdots, U_n$ mutuellement indépendantes et de même loi de Bernoulli $\mathcal{B}(\frac{\lambda}{n})$ telles que : $X = U_1 + \cdots + U_n$. Si $V_1, \cdots, V_n$ sont n v.a.r. indépendantes (entre elles, et des U_i) et de même loi de Poisson $\mathcal{P}(\frac{\lambda}{n})$, on sait qu'alors la v.a.r. $Y = V_1 + \cdots + V_n$ suit la loi de Poisson de paramètre $n \times \dfrac{\lambda}{n} = \lambda$.
Avec ces $2n$ v.a.r., les résultats précédents s'appliquent, notamment la toute dernière inégalité de 5)(b), qui se réécrit :

$$\sum_{k=0}^{+\infty} |P(X = k) - P(Y = k)| = \sum_{k=0}^{+\infty} \left|P(X = k) - e^{-\lambda}.\frac{\lambda^k}{k!}\right| \leqslant \frac{2\lambda^2}{n}.$$

II - Modélisation de la concentration en bactéries

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

Supposons que la bassin contienne n bactéries numérotées de 1 à n ($n \in \mathbb{N}^*$). On considère alors n variables aléatoires $X_1,\ X_2, \cdots, X_n$ à valeurs dans $\{0;1\}$ telles que : $X_i = 1$ si la bactérie i se trouve dans le prélèvement, et $X_i = 0$ sinon. Les variables en question sont supposées indépendantes.
On pose $c = \frac{n}{V}$ qui représente la concentration en bactéries du bassin par m^3.

6) (a) D'après les données de l'énoncé, il est clair que pour tout $i \in [\![1,n]\!]$, X_i suit la loi de Bernoulli de paramètre $\frac{\Delta V}{V}$.

(b) Soit N le nombre de bactéries présentes dans le prélèvement : on a donc $N = \displaystyle\sum_{i=1}^{n} X_i$, et N est la somme de n variables de Bernoulli, indépendantes et de même paramètre : on sait alors, d'après le cours, que N suit la loi binomiale de paramètres $(n, \frac{\Delta V}{V})$.

7) Soit U une variable aléatoire qui suit la loi de Poisson de paramètre λ. On appelle F sa fonction de répartition.
Rappelons donc, pour comprendre quel est le calcul demandé, que puisque U est une variable discrète, F est une fonction en escalier qui vérifie :
⋆ $\forall x < 0,\ F(x) = 0$ puisque $U(\Omega) = \mathbb{N}$, et
⋆ $\forall x \geqslant 0,\ F(x) = P(U \leqslant x) = P(U \leqslant k) = \displaystyle\sum_{i=0}^{k} \frac{\lambda^i}{i!}.e^{-\lambda}$ si $k = \text{Ent}(x)$ est la partie entière du réel x.

D'où la fonction en Turbo-Pascal :

```
Function Poisson (x, lambda : real) : real;
Var s,puiss : real;
   k,i,fact : longint;
Begin
```

```
if x < 0 then Poisson := 0 else
                    begin
                       k:=trunc(x); s:=1; fact := 1; puiss := 1;
                       for i:=1 to k do
                       begin
                       fact:=fact*i; puiss:=puiss*lambda;
                       s:= s+puiss/fact;
                       end;
Poisson := exp(-lambda)*s;
```

On minimise le nombre d'opérations en calculant de proche en proche les factorielles et les puissances (c'est-à-dire qu'on ne recommence pas le calcul du début lorsqu'on passe au terme suivant de la somme). La multiplication par $e^{-\lambda}$ intervient une seule fois à la fin du calcul, la variable `s` calcule de proche en proche, la somme $\sum_{i=0}^{k} \frac{\lambda^i}{i!}$ qui commence à $\frac{\lambda^0}{0!} = 1$.

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

(a) Comme on le fait classiquement pour des variables aléatoires discrètes à valeurs entières - c'est le cas pour N et U - , on peut écrire :

$$|P(N \leqslant K) - P(U \leqslant K)| = \left|\sum_{i=0}^{K} P(N=i) - \sum_{i=0}^{K} P(U=i)\right| = \left|\sum_{i=0}^{K} [P(N=i) - P(U=i)]\right| \leqslant \sum_{i=0}^{K} |P(N=i) - P(U=i)|$$

d'après l'inégalité triangulaire généralisée.

Comme le terme général $|P(N=i) - P(U=i)|$ est positif, et la série correspondante est convergente (selon un argument identique à I-3) (a)), on peut aussi écrire :

$\sum_{i=0}^{K} |P(N=i) - P(U=i)| \leqslant \sum_{i=0}^{+\infty} |P(N=i) - P(U=i)|$. Or :

N suit une loi binomiale de paramètres n et $\frac{\Delta V}{V}$, ce dernier pouvant aussi s'écrire, avec les notations de l'énoncé : $\frac{c\Delta V}{n}$, donc de la forme $\frac{\lambda}{n}$ avec $\lambda = c\Delta V$. Comme U suit justement la loi de Poisson de paramètre $\lambda = c\Delta V$, le résultat de I-5)(c) s'applique, et :

$\sum_{i=0}^{+\infty} |P(N=i) - P(U=i)| \leqslant \frac{2\lambda^2}{n} = \frac{2c^2(\Delta V)^2}{n} = \frac{2c(\Delta V)^2}{V}$ puisque $c = \frac{n}{V} \iff \frac{c}{n} = \frac{1}{V}$.

De toutes ces inégalités successives, on déduit en effet :

$$|P(N \leqslant K) - P(U \leqslant K)| \leqslant \frac{2c(\Delta V)^2}{V}$$

(b) On suppose que $V = 1000$ m^3 et que le prélèvement est de volume ΔV égal à 1 litre (soit 10^{-3} m^3). L'erreur commise en approximant $P(N \leqslant K)$ par $P(U \leqslant K)$ correspond à la différence absolue entre ces deux valeurs, soit $|P(N \leqslant K) - P(U \leqslant K)|$, qui est d'après la question précédente, majorée par : $\frac{2c(\Delta V)^2}{V} = \frac{2.(10^{-3})^2.c}{10^3} \boxed{= 2c.10^{-9}}$.

(c) On suppose que la concentration de bactéries dans le bassin reste inférieure à 10^6 bactéries par mètre cube, et que le prélèvement réalisé est encore de 1 litre. Avec ces données, le majorant de l'erreur d'approximation obtenue précédemment, est lui-même majoré par : $\frac{2.10^6.10^{-3}}{V} = \frac{2.10^3}{V}$.

L'erreur commise est inférieure à 10^{-6} dès que le majorant est lui-même inférieur à cette valeur, soit : $\frac{2.10^3}{V} \leqslant 10^{-6} \iff 2.10^3.10^6 \leqslant V \iff \boxed{V \geqslant 2.10^9}$.

Le bassin doit donc contenir au moins 2 millions de m^3 d'eau, ce qui est énorme !

L'énoncé <u>admet</u> dans la suite que la question 5)(c) peut être améliorée de la façon suivante :

$$\sum_{k=0}^{+\infty} \left|P(X=k) - e^{-\lambda}\frac{\lambda^k}{k!}\right| \leqslant \frac{2\lambda \min(\lambda, 2)}{n}$$

(d) Par le même raisonnement qu'à la question 8)(a), si U suit une loi de Poisson de paramètre $\lambda = c\Delta V$, le même résultat s'applique, mais avec le nouveau majorant :

$|P(N \leqslant K) - P(U \leqslant K)| \leqslant \dfrac{2\lambda.\min(\lambda, 2)}{n} = \dfrac{2c\Delta V.\min(c\Delta V, 2)}{n} = \dfrac{2\Delta V.\min(c\Delta V, 2)}{V} \leqslant \dfrac{4\Delta V}{V}$

puisque $\dfrac{c}{n} = \dfrac{1}{V}$, et $\min(c\Delta V, 2) \leqslant 2$.

(e) On suppose toujours que le prélèvement réalisé est de 1 litre. L'erreur commise dans l'approximation précédente de la loi binomiale de N, par la loi de Poisson de même espérance, est inférieure à 10^{-6} dès que le majorant connu de cette erreur est lui-même inférieur à 10^{-6}, c'est-à-dire :

$4\dfrac{\Delta V}{V} \leqslant 10^{-6} \iff 4.10^{-3} \leqslant V.10^{-6} \iff \boxed{V \geqslant 4.10^3}$.

Il suffit donc que le volume V dépasse 4000 m^3, ce qui permet de constater l'efficacité du nouveau majorant de l'erreur, par rapport au résultat obtenu en 8)(c) !

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 :

$$h(x) = \begin{cases} 0 & \text{si } x \leqslant -1 \\ (1+x)\ln(1+x) - x & \text{si } x > -1 \end{cases}$$

9) La fonction h est bien dérivable sur $]-1; +\infty[$ comme somme et produit de fonctions dérivables sur cet intervalle.

$\forall x > -1 :\ h'(x) = \ln(1+x) + (1+x).\dfrac{1}{1+x} - 1 = \ln(1+x)$.

$\ln(1+x) > 0 \iff 1+x > 1 \iff x > 0$, donc h est strictement décroissante sur $]-1; 0[$, et strictement croissante sur $]0; +\infty[$.

Par conséquent, la fonction h admet un minimum en $x = 0$ qui vaut : $h(0) = 1.\ln(1) - 0 = 0$.

On peut donc conclure que : $\forall x > -1,\ h(x) \geqslant h(0) = 0$.

10) Soit Y une variable aléatoire réelle à valeurs discrètes positives et admettant une espérance $E(Y)$, et soit α un réel strictement positif ; en notant $Y(\Omega) = \{y_k |\ k \in I\}$ où $I \subset \mathbb{N}$, on écrit :

$E(Y) = \displaystyle\sum_{k\in I} y_k.P(Y = y_k) = \sum_{k\in I | y_k \geqslant \alpha} y_k.P(Y = y_k) + \sum_{k\in I | y_k < \alpha} y_k.P(Y = y_k)$, où :

$\displaystyle\sum_{k\in I | y_k \geqslant \alpha} y_k.P(Y = y_k) \geqslant \alpha. \sum_{k\in I | y_k \geqslant \alpha} P(Y = y_k) = \alpha.P(Y \geqslant \alpha)$,

et $\displaystyle\sum_{k\in I | y_k < \alpha} y_k.P(Y = y_k) \geqslant 0$ car $y_k \geqslant 0$ et $P(Y = y_k) \geqslant 0$ (c'est une probabilité !).

De tout ceci, on déduit finalement : $E(Y) \geqslant \alpha.P(Y \geqslant \alpha) \iff \boxed{P(Y \geqslant \alpha) \leqslant \dfrac{1}{\alpha}.E(Y)}$ (car $\alpha > 0$).

11) Soit λ un réel strictement positif et soit X une variable aléatoire qui suit une loi de Poisson de paramètre λ. Soit ε un réel strictement positif.

(a) Pour tout réel u, d'après le théorème de transfert, et sous réserve de convergence absolue de la série :

$$\begin{aligned} E(e^{uX}) &= \sum_{k=0}^{+\infty} e^{uk}.\frac{\lambda^k}{k!}.e^{-\lambda} \\ &= e^{-\lambda}.\sum_{k=0}^{+\infty} \frac{(\lambda.e^u)^k}{k!} = e^{-\lambda}.e^{\lambda.e^u} = \boxed{e^{\lambda.(e^u - 1)}} \end{aligned}$$

La série (à termes positifs) est bien absolument convergente, puisqu'on a reconnu une série exponentielle.

(b) D'après le résultat précédent, et aussi la propriété de linéarité de l'espérance, pour tout réel u :
$E(e^{u(X-(1+\varepsilon)\lambda)}) = E(e^{uX-u.(1+\varepsilon)\lambda}) = e^{-u(1+\varepsilon)\lambda}.E(e^{uX}) = e^{-u(1+\varepsilon)\lambda}.e^{\lambda.(e^u-1)} = e^{-\lambda.(1+u(1+\varepsilon)-e^u)}$.
Il s'agit donc de trouver un réel u_0 tel que :
$1+u_0(1+\varepsilon)-e^{u_0} = h(\varepsilon) \iff 1+u_0(1+\varepsilon)-e^{u_0} = (1+\varepsilon)\ln(1+\varepsilon)-\varepsilon$.
Il est alors assez clair que $\boxed{u_0 = \ln(1+\varepsilon)}$ convient.

(c) i) Pour tout réel $u>0$, il y a égalité des événements :
$[\lambda^{-1}.(X-\lambda) \geqslant \varepsilon] = [X-\lambda \geqslant \lambda.\varepsilon] = [u(X-\lambda) \geqslant \lambda\varepsilon u] = [e^{u(X-\lambda)} \geqslant e^{\lambda\varepsilon u}]$, notamment parce que u, ε et λ sont strictement positifs, et par croissance de la fonction exponentielle sur $\mathbb{R}$.
Il y a donc en particulier, égalité des probabilités : $\boxed{P(\lambda^{-1}(X-\lambda) \geqslant \varepsilon) = P(e^{u(X-\lambda)} \geqslant e^{\lambda\varepsilon u})}$

ii) D'après le résultat précédent, et d'après l'inégalité de Markov redémontrée à la question 10), on peut écrire :
$P(\lambda^{-1}(X-\lambda) \geqslant \varepsilon) = P(e^{u(X-\lambda)} \geqslant e^{\lambda\varepsilon u}) \leqslant \dfrac{1}{e^{\lambda\varepsilon u}}E(e^{u(X-\lambda)}) = E(e^{-\lambda\varepsilon u}.e^{u(X-\lambda)}) = E(e^{u(X-(1+\varepsilon)\lambda)})$
d'après la linéarité de l'espérance.
Comme on pouvait écrire ce qui précède pour tout réel $u>0$, le cas particulier $u=u_0$ donne, d'après (b) :
$$P(\lambda^{-1}(X-\lambda) \geqslant \varepsilon) \leqslant E(e^{u_0(X-(1+\varepsilon)\lambda)}) = e^{-\lambda h(\varepsilon)}.$$

(d) Lorsque $\varepsilon \geqslant 1$: $-\varepsilon \leqslant -1$ et $h(-\varepsilon)=0$, donc $e^{-\lambda h(-\varepsilon)} = e^0 = 1$. Et comme bien sûr une probabilité est majorée par 1, on a bien $P(\lambda^{-1}(X-\lambda)) \leqslant e^{-\lambda h(-\varepsilon)}$.
Si $0<\varepsilon<1$: comme précédemment, pour tout réel $u>0$,
$$\begin{aligned}P(\lambda^{-1}(X-\lambda) \leqslant -\varepsilon) &= P(-\lambda^{-1}(X-\lambda) \geqslant \varepsilon) = P(-u(X-\lambda) \geqslant \lambda\varepsilon u)\\ &= P(e^{-u(X-\lambda)} \geqslant e^{\lambda\varepsilon u})\\ &\leqslant \frac{1}{e^{\lambda\varepsilon u}}.E(e^{-u(X-\lambda)})\\ &\leqslant e^{-\lambda\varepsilon u}.e^{\lambda u}.E(e^{-uX})\\ &\leqslant e^{\lambda u(1-\varepsilon)}.e^{\lambda.(e^{-u}-1)} = e^{-\lambda(1-u(1-\varepsilon)-e^{-u})}\end{aligned}$$
Avec $u=-\ln(1-\varepsilon)$ qui est bien strictement positif puisque $0<1-\varepsilon<1$, on obtient :
$P(\lambda^{-1}(X-\lambda) \leqslant -\varepsilon) \leqslant e^{-\lambda(1+\ln(1-\varepsilon)(1-\varepsilon)-(1-\varepsilon))} = e^{-\lambda((1-\varepsilon)\ln(1-\varepsilon)-(-\varepsilon)}$, soit :
$$P(\lambda^{-1}(X-\lambda) \leqslant -\varepsilon) \leqslant e^{-\lambda.h(-\varepsilon)}$$

12) 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 la question 11) et on suppose de nouveau que $\Delta V = 10^{-3}$ m^3. Fixons un réel α compris entre 0 et 2000.

(a) Pour tout réel $\lambda > 2000$: on a alors $\varepsilon = \lambda^{-1}(\alpha-\lambda) < 0$, donc d'après 11)(d) :
$P(\lambda^{-1}(X-\lambda) \leqslant \lambda^{-1}(\alpha-\lambda)) \leqslant e^{-\lambda.h(-\varepsilon)} = e^{-\lambda.h(-\lambda^{-1}(\alpha-\lambda))} = e^{\lambda.h(\lambda^{-1}\alpha-1)}$.
Et comme $\lambda>0$, il y a égalité des événements : $[X \leqslant \alpha] = [X-\lambda \leqslant \alpha-\lambda] = [\lambda^{-1}(X-\lambda) \leqslant \lambda^{-1}(\alpha-\lambda)]$, on a bien :
$$P(X \leqslant \alpha) = P(\lambda^{-1}(X-\lambda) \leqslant \lambda^{-1}(\alpha-\lambda)) \leqslant e^{-\lambda.h(\lambda^{-1}\alpha-1)}$$

(b) Toujours avec $\lambda > 2000$: $\lambda^{-1}(\alpha-\lambda) = \dfrac{\alpha}{\lambda}-1$, où :
$0<\alpha<2000<\lambda \Longrightarrow 0 < \dfrac{\alpha}{\lambda} < \dfrac{\alpha}{2000} < 1 \Longrightarrow -1 < \dfrac{\alpha}{\lambda}-1 < \dfrac{\alpha}{2000}-1 < 0$.
Or la fonction h est strictement décroissante et positive sur l'intervalle $]-1;0[$, donc :
$0 < h(\frac{\alpha}{\lambda}-1) > h(\frac{\alpha}{2000}-1)$, et comme $\lambda < 2000$, et comme $\lambda > 2000$, on obtient par produit d'inégalités entre réels tous positifs :
$\lambda.h(\frac{\alpha}{\lambda}-1) > 2000.h(\frac{\alpha}{2000}-1) \Longrightarrow -\lambda.h(\frac{\alpha}{\lambda}-1) < -2000.h(\frac{\alpha}{2000}-1) \Longrightarrow e^{-\lambda.h(\lambda^{-1}.\alpha-1)} < e^{-2000.h(\frac{\alpha}{2000}-1)}$
par croissance stricte de l'exponentielle sur $\mathbb{R}$.
Vu l'inégalité obtenue à la question précédente, on a bien : $\boxed{P(X \leqslant \alpha) \leqslant e^{-2000.h(\frac{\alpha}{2000}-1)}}$.

(c) La fonction affine : $x \mapsto \dfrac{x}{2000} - 1$ est strictement croissante (car $2000 > 0$) sur $]0; 2000]$, à valeurs dans $]-1; 0]$; la fonction h est strictement décroissante sur $]-1; 0]$, donc par composition : $x \mapsto h(\dfrac{x}{2000} - 1)$ est strictement décroissante sur $]0; 2000]$.
Comme $2000 > 0$, la fonction $g : \ x \mapsto 2000.h(\dfrac{x}{2000} - 1) - \ln(100)$ est encore strictement décroissante sur $]0; 2000]$, à valeurs dans :
$[2000.h(0) - \ln(100); 2000. \lim\limits_{x \to -1^+} h(x) - \ln(100)[= [-\ln(100); 2000 - \ln(100)[$, puisque :
$\lim\limits_{x \to -1^+} h(x) = \lim\limits_{x \to -1^+} (1+x)\ln(1+x) - x = \lim\limits_{x \to 0^+} x\ln(x) - x + 1 = 1$ par croissances comparées.
La fonction g est également continue sur $]0; 2000]$, comme composée de fonctions continues.
Puisque bien sûr : $2000 - \ln(100) > 0 > -\ln(100)$, le théorème de la bijection assure que l'équation : $2000.h(\dfrac{x}{2000} - 1) - \ln(100) = 0 \iff 2000.h(\dfrac{x}{2000} - 1) = \ln(100)$, admet une unique solution α_0 dans $]0; 2000]$.

(d) On admettra que $\alpha_0 > 1865$. Du résultat de la question (b) on tire, pour $\lambda > 2000$ et $\alpha = \alpha_0$:
$P(X < 1865) \leqslant P(X \leqslant \alpha_0) \leqslant e^{-2000h(\frac{\alpha_0}{2000} - 1)} = e^{-\ln(100)} = \dfrac{1}{100}$.

(e) Exemple d'application : on compte 1600 bactéries dans le prélèvement de 1 litre ; le résultat de la question précédente exprime que si le nombre moyen réel λ de bactéries par litre est supérieur à 2000 (donc, au-delà de la limite autorisée), la probabilité que le prélèvement présente une quantité de bactéries inférieure à 1865 (et donc, donne faussement l'idée que la concentration en bactéries est dans la norme autorisée), est inférieure à 1%. C'est le risque d'erreur pris lorsqu'on autorise le bassin avec, comme ici, 1600 bactéries dans le prélèvement d'un litre.

★★★ FIN DU SUJET ★★★