



PRÉPASIMPLE

prepasimple.fr

ESSEC

Mathématiques 2 — Option économique

2013

Mathématiques appliquées

Mathématiques appliquées — Prépa ECG

Annales de concours · Corrigé

I — Lien entre loi binomiale et loi de Poisson

1)(a) Les n variables sont indépendantes et ont la même loi de Bernoulli de paramètre p_n . Leur somme compte le nombre de succès parmi n épreuves indépendantes de même probabilité de succès. Ainsi S_n suit la loi binomiale $\mathcal{B}(n, p_n)$.

1)(b)(i) La formule de la loi binomiale donne, pour $0 \leq k \leq n$,

$$P(S_n = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k}.$$

Si un paramètre vaut 1, la loi est concentrée en n ; la formule s'interprète alors avec la convention usuelle $0^0 = 1$ dans cette écriture de la loi binomiale.

1)(b)(ii) Puisque $np_n \rightarrow \lambda$ et $n \rightarrow +\infty$, on a $p_n = (np_n)/n \rightarrow 0$. En particulier, pour n assez grand, $0 < p_n < 1$, et on peut prendre le logarithme.

La dérivabilité de $t \mapsto \ln(1 - t)$ en 0 donne $\ln(1 - t)/t \rightarrow -1$. Par conséquent,

$$n \ln(1 - p_n) = (np_n) \frac{\ln(1 - p_n)}{p_n} \rightarrow -\lambda.$$

Par continuité de l'exponentielle, $(1 - p_n)^n \rightarrow e^{-\lambda}$. On a appliqué la continuité après avoir établi une limite du logarithme, sans composer abusivement des équivalents.

1)(b)(iii) Fixons $k \in \mathbb{N}$. Pour $n \geq k$ assez grand,

$$P(S_n = k) = \frac{(np_n)^k}{k!} \left(\prod_{j=0}^{k-1} \left(1 - \frac{j}{n} \right) \right) (1 - p_n)^n (1 - p_n)^{-k}.$$

Le produit contient un nombre fixé de facteurs; chacun tend vers 1. Le premier facteur tend vers $\lambda^k/k!$, le troisième vers $e^{-\lambda}$ et le dernier vers 1. Pour $k = 0$, le produit vide et les puissances d'exposant nul valent 1. Donc, pour tout $k \in \mathbb{N}$,

$$P(S_n = k) \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}.$$

Ces probabilités sont celles d'une loi de Poisson de paramètre λ . La caractérisation de la convergence en loi des variables à valeurs entières donne aussi $S_n \xrightarrow{\mathcal{L}} \mathcal{P}(\lambda)$.

1)(c)(i) Chaque X_k vaut 0 ou 1, donc N_n prend ses valeurs dans $\{0, 1\}$. Plus

précisément, $N_n = 0$ si et seulement si tous les X_k valent 0, et $N_n = 1$ si au moins un X_k vaut 1. Si $p_n = 1$, seule la valeur 1 est effectivement prise ; pour n assez grand, $0 < p_n < 1$ et les deux probabilités sont strictement positives.

1)(c)(ii) Par indépendance,

$$P(N_n = 0) = P(X_1 = 0, \dots, X_n = 0) = \prod_{k=1}^n P(X_k = 0) = (1 - p_n)^n \rightarrow e^{-\lambda}.$$

1)(c)(iii) On a $P(N_n = 1) = 1 - P(N_n = 0) \rightarrow 1 - e^{-\lambda}$. Pour tout entier différent de 0 et 1, la probabilité est nulle. La caractérisation ponctuelle de la convergence en loi pour les variables à valeurs entières montre donc

$$N_n \xrightarrow{\mathcal{L}} \mathcal{B}(1 - e^{-\lambda}).$$

La limite est une variable de Bernoulli, et non une variable de Poisson : le maximum reste à valeurs dans $\{0, 1\}$.

2)(a) Méthode 1 — Étude d'une fonction

Posons $g(u) = e^{-u} - 1 + u$ pour $u \geq 0$. On a $g(0) = 0$ et $g'(u) = 1 - e^{-u} \geq 0$. La fonction g est donc croissante sur $[0, +\infty[$ et positive ou nulle. Ainsi $1 - u \leq e^{-u}$.

Méthode 2 — Tangente à une fonction convexe

La fonction $f(u) = e^{-u}$ est deux fois dérivable sur $\mathbb{R}$, avec $f''(u) = e^{-u} > 0$. Elle est convexe et son graphe est au-dessus de sa tangente en 0, d'équation $y = f(0) + f'(0)u = 1 - u$. Ainsi $e^{-u} \geq 1 - u$, même pour tout réel u .

La première méthode ne demande qu'une étude de signe ; la seconde permet de reconnaître une inégalité de tangente réutilisable pour d'autres fonctions convexes.

2)(b) Posons $a = \lambda/n$, de sorte que $0 < a \leq 1$, et $q_k = e^{-a} a^k / k!$. La série de terme général q_k converge et sa somme est 1, par la série exponentielle.

La question 2)(a) donne $e^{-a} \geq 1 - a$, donc $|P(U = 0) - q_0| = e^{-a} - 1 + a$. Pour $k = 1$, on a $|P(U = 1) - q_1| = a(1 - e^{-a})$. Pour $k \geq 2$, la probabilité $P(U = k)$ est nulle et la valeur absolue vaut q_k . Par conséquent, la série demandée converge et

$$\begin{aligned} \sum_{k=0}^{+\infty} |P(U = k) - q_k| &= (e^{-a} - 1 + a) + a(1 - e^{-a}) + \sum_{k=2}^{+\infty} q_k \\ &= (e^{-a} - 1 + a) + a(1 - e^{-a}) + 1 - e^{-a} - ae^{-a} \\ &= 2a(1 - e^{-a}). \end{aligned}$$

En remplaçant a par λ/n , on obtient exactement l'égalité annoncée.

2)(c) L'inégalité de 2)(a) équivaut à $1 - e^{-a} \leq a$. Le membre $1 - e^{-a}$ est positif et $a > 0$, donc

$$2a(1 - e^{-a}) \leq 2a^2.$$

La somme de 2)(b) est ainsi majorée par $2(\lambda/n)^2$.

3)(a) Fixons $i \in \mathbb{N}$. Pour tout $k \in \mathbb{N}$,

$$0 \leq |P(U = k - i) - P(V = k - i)| \leq P(U = k - i) + P(V = k - i).$$

Les termes sont nuls pour $k < i$. En posant $j = k - i$ pour $k \geq i$, chacune des deux séries du membre de droite a pour somme 1. La comparaison prouve donc la convergence de la série définissant A_i .

Le même changement d'indice donne

$$A_i = \sum_{j=0}^{+\infty} |P(U = j) - P(V = j)| = 2 \frac{\lambda}{n} (1 - e^{-\lambda/n}) \leq 2(\lambda/n)^2,$$

d'après 2)(b) et 2)(c). En particulier, A_i ne dépend pas de i .

3)(b) Pour chaque i , on a

$$0 \leq A_i P(Z = i) \leq 2(\lambda/n)^2 P(Z = i).$$

Puisque Z est à valeurs dans $\mathbb{N}$,

$$\sum_{i=0}^{+\infty} P(Z = i) = 1.$$

La comparaison des séries positives prouve la convergence et donne

$$\sum_{i=0}^{+\infty} A_i P(Z = i) \leq 2(\lambda/n)^2.$$

3)(c) Pour $i > k$, on a $k - i < 0$: les probabilités $P(U = k - i)$ et $P(V = k - i)$ sont nulles. La série considérée n'a donc qu'un nombre fini de termes éventuellement non nuls, ceux d'indices 0 à k . Elle est convergente.

3)(d)(i) Fixons $k \geq 2$. Dans la somme définissant B_k , le terme $i = k$ correspond à $k - i = 0$, celui d'indice $i = k - 1$ correspond à $k - i = 1$, et ceux d'indices $0 \leq i \leq k - 2$ correspondent à $k - i \geq 2$. Pour ces derniers, $P(U = k - i) = 0$. Les termes d'indice $i > k$ sont nuls. Il vient donc

$$B_k = |1 - a - e^{-a}|P(Z = k) + |a - ae^{-a}|P(Z = k - 1) + \sum_{i=0}^{k-2} P(V = k - i)P(Z = i),$$

avec $a = \lambda/n$. C'est la décomposition demandée.

3)(d)(ii) Les événements $[Z = i, V = k - i]$, pour $0 \leq i \leq k - 2$, sont deux à deux incompatibles et leur réunion est incluse dans $[Z + V = k]$. Comme Z et V sont indépendantes,

$$\sum_{i=0}^{k-2} P(V = k - i)P(Z = i) = \sum_{i=0}^{k-2} P(Z = i, V = k - i) \leq P(Z + V = k).$$

Cette preuve utilise une réunion finie d'événements et l'indépendance, sans supposer une formule générale sur les densités de sommes.

3)(d)(iii) Posons $C = |1 - a - e^{-a}|$ et $D = |a - ae^{-a}|$. Les deux questions précédentes donnent, pour $k \geq 2$,

$$0 \leq B_k \leq CP(Z = k) + DP(Z = k - 1) + P(Z + V = k).$$

Les trois séries du membre de droite sont convergentes : leurs sommes sont majorées respectivement par C , D et 1. Les termes B_0 et B_1 sont finis par 3)(c). Par comparaison,

$$\sum_{k=0}^{+\infty} B_k$$

converge.

L'égalité admise dans l'énoncé est

$$\sum_{k=0}^{+\infty} B_k = \sum_{i=0}^{+\infty} A_i P(Z = i).$$

Elle permet, avec 3)(b), de majorer la première somme par $2a^2$. Les manipulations de sommes doubles discrètes sont admises dans le programme ; ici l'égalité est, en outre, explicitement fournie par le sujet.

4)(a) Pour tout k ,

$$0 \leq |P(Z + U = k) - P(Z + V = k)| \leq P(Z + U = k) + P(Z + V = k).$$

Les variables $Z + U$ et $Z + V$ sont à valeurs dans $\mathbb{N}$. Chacune des deux séries de probabilités a donc pour somme 1. La comparaison donne la convergence de la série demandée, avec une somme majorée par 2.

4)(b) Pour chaque k , la décomposition selon les valeurs de Z et l'indépendance donnent les sommes finies

$$P(Z + U = k) = \sum_{i=0}^k P(Z = i)P(U = k - i),$$

$$P(Z + V = k) = \sum_{i=0}^k P(Z = i)P(V = k - i).$$

Il suffit de considérer les valeurs i dont la probabilité est non nulle ; les autres contributions valent directement zéro. L'inégalité triangulaire donne

$$|P(Z + U = k) - P(Z + V = k)| \leq \sum_{i=0}^k |P(U = k - i) - P(V = k - i)|P(Z = i) = B_k.$$

En sommant les inégalités entre termes positifs, puis en utilisant l'égalité admise après 3)(d)(iii), on obtient

$$\begin{aligned} \sum_{k=0}^{+\infty} |P(Z + U = k) - P(Z + V = k)| &\leq \sum_{k=0}^{+\infty} B_k \\ &= \sum_{i=0}^{+\infty} \sum_{k=0}^{+\infty} |P(U = k - i) - P(V = k - i)|P(Z = i) \\ &= \sum_{i=0}^{+\infty} A_i P(Z = i) \leq 2(\lambda/n)^2. \end{aligned}$$

Ainsi, ajouter une même variable indépendante aux deux lois ne dégrade pas la majoration obtenue en question 2.

5)(a) Pour $j \in \{0, \dots, n\}$, définissons

$$T_j = U_1 + \dots + U_{n-j} + V_{n-j+1} + \dots + V_n,$$

où une somme vide vaut 0. Alors $T_0 = U_1 + \dots + U_n$ et $T_n = V_1 + \dots + V_n$. Pour tout entier naturel k ,

$$P(T_0 = k) - P(T_n = k) = \sum_{j=1}^n (P(T_{j-1} = k) - P(T_j = k)).$$

En prenant la valeur absolue et en appliquant l'inégalité triangulaire à cette somme finie, on obtient

$$|P(T_0 = k) - P(T_n = k)| \leq \sum_{j=1}^n |P(T_{j-1} = k) - P(T_j = k)|,$$

ce qui est exactement l'inégalité développée dans l'énoncé : on remplace successivement $U_n, U_{n-1}, \dots, U_1$ par $V_n, V_{n-1}, \dots, V_1$.

5)(b) Pour $1 \leq j \leq n$, posons

$$Z_j = U_1 + \dots + U_{n-j} + V_{n-j+2} + \dots + V_n.$$

On a $T_{j-1} = Z_j + U_{n-j+1}$ et $T_j = Z_j + V_{n-j+1}$. Par indépendance de l'ensemble des $2n$ variables et par le lemme des coalitions, les trois variables $Z_j, U_{n-j+1}, V_{n-j+1}$ sont indépendantes. Elles sont à valeurs dans $\mathbb{N}$, et les deux dernières ont les lois requises dans la question 4. Si $n = 1$, $Z_1 = 0$ est une variable constante, ce qui satisfait aussi ces conditions.

La question 4)(b) donne donc, pour chaque j ,

$$\sum_{k=0}^{+\infty} |P(T_{j-1} = k) - P(T_j = k)| \leq 2(\lambda/n)^2.$$

On somme l'inégalité de 5)(a) sur k . La somme sur j étant finie et chacune des séries positives étant convergente, on peut écrire

$$\sum_{k=0}^{+\infty} |P(T_0 = k) - P(T_n = k)| \leq \sum_{j=1}^n 2(\lambda/n)^2 = \frac{2\lambda^2}{n}.$$

5)(c) La somme T_0 des n variables de Bernoulli indépendantes de même paramètre λ/n suit la loi $\mathcal{B}(n, \lambda/n)$. Elle a donc la même loi que X . Par stabilité de la loi de Poisson pour des sommes de variables indépendantes, T_n suit la loi $\mathcal{P}(n\lambda/n) = \mathcal{P}(\lambda)$. En remplaçant

les probabilités dans 5)(b), on obtient

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

La majoration mesure ici l'écart total des probabilités ponctuelles ; elle ne suppose pas que X et une variable de Poisson soient définies par la même expérience.

II — Modélisation de la concentration en bactéries

6)(a) D'après le modèle, X_i vaut 1 avec la probabilité $\Delta V/V$ et vaut 0 avec la probabilité $1 - \Delta V/V$. Comme $0 < \Delta V \leq V$, ces paramètres définissent bien une loi de Bernoulli. Ainsi $X_i \sim \mathcal{B}(\Delta V/V)$.

6)(b) Chaque bactérie présente dans le prélèvement apporte exactement une unité au comptage. On a donc

$$N = \sum_{i=1}^n X_i.$$

Les X_i sont indépendantes par hypothèse, avec le même paramètre $\Delta V/V$. Par conséquent,

$$N \sim \mathcal{B}(n, \Delta V/V).$$

En particulier, $E(N) = n\Delta V/V = c\Delta V$, ce qui indique le paramètre naturel de l'approximation de Poisson.

7) Question adaptée au programme actuel. Code adapté en Python.

Pour $x < 0$, la fonction de répartition vaut 0, puisque $U \geq 0$. Pour $x \geq 0$, posons $m = \lfloor x \rfloor$. Alors

$$F(x) = P(U \leq m) = \sum_{k=0}^m e^{-\lambda} \frac{\lambda^k}{k!}.$$

Méthode 1 — Récurrence sur les probabilités ponctuelles

En notant $p_k = e^{-\lambda} \lambda^k / k!$, on a $p_0 = e^{-\lambda}$ et, pour $k \geq 1$, $p_k = p_{k-1} \lambda / k$. La récurrence évite de recalculer une puissance et une factorielle pour chaque terme.

Python

```
import numpy as np

def poisson(x, lam):
    if x < 0:
        return 0.0
    m = int(np.floor(x))
    p = np.exp(-lam)
    s = p
    for k in range(1, m + 1):
        p = p * lam / k
        s = s + p
    return s
```

À l'entrée de l'itération d'indice k , la variable p représente p_{k-1} ; après sa mise à jour, elle représente p_k . La variable s contient alors la somme des probabilités de 0 à k . La borne supérieure exclusive de range est $m + 1$, de sorte que p_m est bien ajouté. Si $m = 0$, la boucle est vide et le résultat est p_0 .

Le nombre d'itérations est m , avec une multiplication, une division et une addition par terme. Ce programme est la traduction directe de la méthode demandée. En arithmétique flottante, l'initialisation peut toutefois devenir nulle lorsque λ est grand : pour $\lambda = 2000$, le nombre e^{-2000} est trop petit pour être représenté par un flottant usuel. Une somme proche de 1/2 ne doit alors pas être calculée à partir de cette initialisation.

Méthode 2 — Récurrences à partir d'un terme central

Pour éviter cette difficulté dans les paramètres du test, on peut choisir $r = \min(m, \lfloor \lambda \rfloor)$, calculer le terme p_r par son logarithme, puis utiliser les relations

$$p_{k-1} = p_k \frac{k}{\lambda} \quad (k \geq 1), \quad p_{k+1} = p_k \frac{\lambda}{k+1}.$$

La valeur r est un indice où la probabilité est maximale parmi les indices de 0 à m : en effet $p_k/p_{k-1} = \lambda/k$. On ne part donc plus d'une probabilité artificiellement minuscule à l'intérieur d'une somme de taille appréciable. La commande `np.log(t)` renvoie $\ln(t)$ pour $t > 0$.

Python

```
import numpy as np

def poisson_stable(x, lam):
    if x < 0:
        return 0.0
    m = int(np.floor(x))
    r = min(m, int(np.floor(lam)))
    log_p = -lam
    for k in range(1, r + 1):
        log_p = log_p + np.log(lam / k)
    p_r = np.exp(log_p)
    s = p_r
    p = p_r
    for k in range(r, 0, -1):
        p = p * k / lam
        s = s + p
    p = p_r
    for k in range(r + 1, m + 1):
        p = p * lam / k
        s = s + p
    return min(1.0, s)
```

La première boucle calcule

$$-\lambda + \sum_{k=1}^r \ln(\lambda/k) = \ln(p_r),$$

puis la deuxième somme les termes d'indices $r - 1$ à 0 et la troisième ceux d'indices $r + 1$ à m . Chaque terme est ajouté une seule fois. Les divisions par λ sont légitimes car $\lambda > 0$. La dernière instruction limite à 1 un éventuel dépassement dû aux arrondis. Les probabilités beaucoup plus petites que les possibilités de représentation des flottants peuvent encore être arrondies à zéro ; aucun des deux algorithmes n'offre une arithmétique exacte sur les réels.

La première méthode minimise les opérations dans le calcul élémentaire attendu. La seconde ajoute le calcul initial d'un logarithme de probabilité pour permettre l'évaluation numérique autour de paramètres tels que $\lambda = 2000$. Les deux codes ont été exécutés sur des cas représentatifs ; la version stable a aussi été confrontée à une fonction de référence pour des paramètres élevés. Ces essais vérifient le calcul numérique, sans remplacer la démonstration de la formule de F .

8)(a) Prenons dans 5)(c) le paramètre $\lambda = c\Delta V$. Puisque $n = cV$, on a $\lambda/n = \Delta V/V \leq 1$. Les hypothèses $0 < \lambda \leq n$ sont donc satisfaites et N a bien la loi requise.

Les variables N et U étant à valeurs dans $\mathbb{N}$, on a

$$\begin{aligned} |P(N \leq K) - P(U \leq K)| &= \left| \sum_{k=0}^K (P(N = k) - P(U = k)) \right| \\ &\leq \sum_{k=0}^K |P(N = k) - P(U = k)| \\ &\leq \sum_{k=0}^{+\infty} |P(N = k) - P(U = k)| \\ &\leq \frac{2(c\Delta V)^2}{n} = \frac{2c(\Delta V)^2}{V}. \end{aligned}$$

On a utilisé uniquement une somme finie pour passer aux fonctions de répartition ;

aucune indépendance entre N et U n'est nécessaire.

8)(b) Avec les volumes exprimés en mètres cubes, $V = 1000$ et $\Delta V = 10^{-3}$. La majoration devient

$$|P(N \leq K) - P(U \leq K)| \leq \frac{2c \times 10^{-6}}{1000} = 2c \times 10^{-9}.$$

La différence de deux probabilités étant aussi majorée par 1, on peut retenir $\min(1, 2c \times 10^{-9})$. La concentration c n'est pas fixée dans cette sous-question ; aucun nombre indépendant de c plus précis ne peut être déduit de la seule majoration 8)(a).

8)(c) Puisque $c < 10^6$ et $\Delta V = 10^{-3}$, 8)(a) donne

$$|P(N \leq K) - P(U \leq K)| \leq \frac{2c \times 10^{-6}}{V} < \frac{2}{V}.$$

Pour obtenir une garantie uniforme à partir de cette majoration pour toutes les concentrations $c < 10^6$, on impose $2/V \leq 10^{-6}$, soit

$$V \geq 2 \times 10^6 \text{ m}^3.$$

Le seuil de volume fourni par la majoration est donc $2 \times 10^6 \text{ m}^3$. Au seuil lui-même, l'erreur est strictement inférieure à 10^{-6} grâce à l'hypothèse stricte $c < 10^6$. Pour $V < 2 \times 10^6$, des valeurs de c arbitrairement proches de 10^6 rendent la borne $2c \times 10^{-6}/V$ supérieure à 10^{-6} : cette majoration seule ne donne plus la garantie uniforme demandée. Cela ne prouve pas que l'erreur réelle dépasse ce niveau.

8)(d) L'amélioration admise dans l'énoncé affirme, pour $X \sim \mathcal{B}(n, \lambda/n)$ et $0 < \lambda \leq n$,

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

Avec $\lambda = c\Delta V$ et $n = cV$, le raisonnement sur les fonctions de répartition utilisé en 8)(a) donne

$$|P(N \leq K) - P(U \leq K)| \leq \frac{2c\Delta V \min(c\Delta V, 2)}{cV} = \frac{2\Delta V \min(c\Delta V, 2)}{V} \leq \frac{4\Delta V}{V}.$$

Le facteur c se simplifie : la dernière borne est indépendante de la concentration. L'amélioration initiale étant donnée comme admise dans le sujet source, sa démonstration n'est pas demandée ; c'est son application au modèle qui vient d'être établie.

8)(e) Pour un prélèvement de un litre, la borne de 8)(d) vaut $4 \times 10^{-3}/V$. Elle est inférieure ou égale à 10^{-6} dès que

$$V \geq \frac{4 \times 10^{-3}}{10^{-6}} = 4000 \text{ m}^3.$$

Le seuil fourni par cette majoration est donc 4000 m^3 , au lieu de $2 \times 10^6 \text{ m}^3$ en 8)(c), et il ne nécessite plus de borne sur c .

Si le mot « inférieure » est interprété au sens strict et que l'on utilise seulement l'inégalité 8)(d), il faut prendre $V > 4000 \text{ m}^3$: cette condition a pour borne inférieure 4000, sans minimum dans les volumes réels. Au seuil $V = 4000$, l'inégalité écrite permet d'affirmer une erreur inférieure ou égale à 10^{-6} . Cette distinction concerne la garantie fournie par la borne, et non une valeur exacte de l'erreur.

III — Construction d'une procédure de test

9) Sur $] -1, +\infty[$, la fonction h est dérivable et

$$h'(x) = \ln(1+x) + 1 - 1 = \ln(1+x).$$

Pour $-1 < x < 0$, on a $0 < 1+x < 1$, donc $h'(x) < 0$. On a $h'(0) = 0$, puis $h'(x) > 0$ pour $x > 0$. La fonction est ainsi strictement décroissante sur $] -1, 0]$ et strictement croissante sur $[0, +\infty[$. Sa valeur minimale sur cet intervalle est $h(0) = 0$. Donc $h(x) > 0$ pour $x > -1$ et $x \neq 0$.

Pour préciser les bornes de variation, posons $t = 1+x > 0$. Lorsque $x \rightarrow -1^+$, on a $t \ln t \rightarrow 0$ par croissances comparées, d'où $h(x) = t \ln t - t + 1 \rightarrow 1$. Lorsque $x \rightarrow +\infty$, cette même expression tend vers $+\infty$.

Sur $] -\infty, -1]$, l'énoncé définit $h(x) = 0$. Au total, h est positive ou nulle sur $\mathbb{R}$ et ses zéros sont les réels $x \leq -1$ ainsi que $x = 0$. La définition donnée n'est pas le prolongement continu à gauche de la formule logarithmique : $h(-1) = 0$, alors que

$\lim_{x \rightarrow -1^+} h(x) = 1$. Cette particularité doit être respectée en 11)(d) et 12).

10) Le support de Y est un ensemble discret de réels positifs ou nuls. Comme Y admet une espérance et est positive, sa série de transfert a des termes positifs et une somme finie. Pour $\alpha > 0$,

$$E(Y) = \sum_y yP(Y = y) \geq \sum_{y \geq \alpha} yP(Y = y) \geq \alpha \sum_{y \geq \alpha} P(Y = y) = \alpha P(Y \geq \alpha).$$

La première inégalité supprime des termes positifs ; la seconde utilise $y \geq \alpha$ sur l'ensemble considéré. En divisant par le réel strictement positif α , on obtient l'inégalité demandée :

$$P(Y \geq \alpha) \leq E(Y)/\alpha.$$

Le résultat de Markov n'est donc pas supposé connu : il est établi ici dans le cas discret nécessaire à la suite.

11)(a) Pour $u \in \mathbb{R}$, les termes de la série de transfert sont positifs et

$$\sum_{k=0}^{+\infty} e^{uk} P(X = k) = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(\lambda e^u)^k}{k!} = e^{-\lambda} e^{\lambda e^u} < +\infty.$$

La convergence de la série exponentielle justifie à la fois l'existence de l'espérance et son calcul. Ainsi

$$E(e^{uX}) = e^{\lambda(e^u - 1)} \quad \text{pour tout } u \in \mathbb{R}.$$

11)(b) Par 11)(a), pour tout $u \in \mathbb{R}$,

$$E(e^{u(X-(1+\varepsilon)\lambda)}) = \exp(\lambda(e^u - 1 - u(1 + \varepsilon))).$$

Prenons $u_0 = \ln(1 + \varepsilon)$. Comme $\varepsilon > 0$, $u_0 > 0$. On a $e^{u_0} = 1 + \varepsilon$, donc

$$e^{u_0} - 1 - u_0(1 + \varepsilon) = \varepsilon - (1 + \varepsilon) \ln(1 + \varepsilon) = -h(\varepsilon).$$

Cela donne exactement $E(e^{u_0(X-(1+\varepsilon)\lambda)}) = e^{-\lambda h(\varepsilon)}$.

11)(c)(i) Puisque $\lambda > 0$, l'événement $[\lambda^{-1}(X - \lambda) \geq \varepsilon]$ est $[X - \lambda \geq \lambda\varepsilon]$. Pour $u > 0$, la multiplication par u puis l'application de l'exponentielle préservent l'ordre. Ces opérations sont réversibles. Ainsi les deux événements

$$[\lambda^{-1}(X - \lambda) \geq \varepsilon] \quad \text{et} \quad [e^{u(X-\lambda)} \geq e^{\lambda\varepsilon u}]$$

sont égaux, d'où l'égalité de leurs probabilités.

11)(c)(ii) La variable $Y = e^{u(X-\lambda)}$ est discrète, positive et admet une espérance finie par 11)(a). L'inégalité de Markov, démontrée en 10), s'applique avec le seuil strictement positif $e^{\lambda\varepsilon u}$. Elle donne

$$\begin{aligned} P(\lambda^{-1}(X - \lambda) \geq \varepsilon) &\leq e^{-\lambda\varepsilon u} E(e^{u(X-\lambda)}) \\ &= E(e^{u(X-(1+\varepsilon)\lambda)}). \end{aligned}$$

En choisissant $u = u_0 = \ln(1 + \varepsilon)$, la question 11)(b) permet de conclure

$$P(\lambda^{-1}(X - \lambda) \geq \varepsilon) \leq e^{-\lambda h(\varepsilon)}.$$

11)(d) Si $\varepsilon > 1$, l'événement considéré équivaut à $[X \leq \lambda(1 - \varepsilon)]$ et son seuil est strictement négatif. Puisque $X \geq 0$, cet événement est impossible, donc sa probabilité vaut 0. La définition donne $h(-\varepsilon) = 0$ et le membre de droite vaut 1 ; l'inégalité est vérifiée.

Si $\varepsilon = 1$, l'événement est $[X = 0]$, de probabilité $e^{-\lambda}$. La définition originale donne encore $h(-1) = 0$, donc $e^{-\lambda} \leq 1 = e^{-\lambda h(-1)}$. La borne demandée est vraie, bien qu'elle soit ici peu précise.

Supposons maintenant $0 < \varepsilon < 1$ et posons $v = \ln(1 - \varepsilon) < 0$. Comme la multiplication par v inverse l'ordre, on a l'égalité d'événements

$$[X - \lambda \leq -\lambda\varepsilon] = [e^{v(X-\lambda)} \geq e^{-v\lambda\varepsilon}].$$

La variable exponentielle est positive et admet une espérance d'après 11)(a). Markov donne alors

$$\begin{aligned}
 P(X - \lambda \leq -\lambda\varepsilon) &\leq e^{v\lambda\varepsilon} E(e^{v(X-\lambda)}) \\
 &= \exp(\lambda(e^v - 1 - v(1 - \varepsilon))) \\
 &= \exp(\lambda(-\varepsilon - (1 - \varepsilon) \ln(1 - \varepsilon))) \\
 &= e^{-\lambda h(-\varepsilon)}.
 \end{aligned}$$

Les trois cas couvrent tous les réels $\varepsilon > 0$ et établissent l'inégalité demandée.

12)(a) Pour $\lambda > 2000$ et $0 \leq \alpha \leq 2000$, posons $\varepsilon = 1 - \alpha/\lambda$. On a $0 < \varepsilon \leq 1$. De plus,

$$[X \leq \alpha] = [\lambda^{-1}(X - \lambda) \leq \lambda^{-1}(\alpha - \lambda)] = [\lambda^{-1}(X - \lambda) \leq -\varepsilon].$$

La question 11)(d) s'applique et donne

$$P(X \leq \alpha) \leq e^{-\lambda h(-\varepsilon)} = e^{-\lambda h(\alpha/\lambda - 1)}.$$

Le cas $\alpha = 0$ correspond à $\varepsilon = 1$: la borne obtenue est 1 avec la définition de h , ce qui reste valide.

12)(b) Si $\alpha = 0$, les deux expressions exponentielles apparaissant dans 12)(a) et 12)(b) valent 1 puisque $h(-1) = 0$. La conclusion est donc immédiate à partir de 12)(a).

Supposons $0 < \alpha \leq 2000$ et définissons, pour $t \geq 2000$,

$$g_\alpha(t) = t h(\alpha/t - 1) = \alpha \ln(\alpha/t) - \alpha + t.$$

L'égalité utilise $\alpha/t > 0$, de sorte que l'argument de h est strictement supérieur à -1 . La fonction est dérivable et

$$g'_\alpha(t) = 1 - \alpha/t \geq 0,$$

car $t \geq 2000 \geq \alpha$. Donc $g_\alpha(\lambda) \geq g_\alpha(2000)$ pour $\lambda > 2000$. La fonction $z \mapsto e^{-z}$ étant décroissante,

$$P(X \leq \alpha) \leq e^{-g_\alpha(\lambda)} \leq e^{-g_\alpha(2000)} = e^{-2000h(\alpha/2000 - 1)}.$$

Il fallait étudier le produit $t h(\alpha/t - 1)$; la seule monotonie de h ne suffirait pas à comparer les expressions complètes.

12)(c) Posons $q(x) = 2000h(x/2000 - 1)$. Pour $0 < x \leq 2000$, on peut écrire

$$q(x) = x \ln(x/2000) - x + 2000.$$

Cette fonction est continue sur $]0, 2000]$ et dérivable sur $]0, 2000[$, avec

$$q'(x) = \ln(x/2000) < 0.$$

Elle est donc strictement décroissante. Par croissances comparées, $x \ln(x/2000) \rightarrow 0$ lorsque $x \rightarrow 0^+$; ainsi $q(x) \rightarrow 2000$ lorsque $x \rightarrow 0^+$, tandis que $q(2000) = 0$. Comme $0 < \ln(100) < 2000$, le théorème des valeurs intermédiaires et la stricte décroissance donnent une unique solution $\alpha_0 \in]0, 2000[$ de $q(x) = \ln(100)$.

Il reste à examiner l'extrémité $x = 0$: la définition originale de h donne $q(0) = 2000h(-1) = 0$, donc $x = 0$ n'est pas solution. La solution est ainsi unique sur tout $[0, 2000]$. Il serait incorrect d'affirmer que q est continue ou décroissante sur cet intervalle fermé, puisque sa limite à droite en 0 diffère de sa valeur en 0.

12)(d) L'hypothèse admise donne $1865 < \alpha_0 < 2000$. Sur $]0, 2000[$, q est strictement décroissante, donc

$$q(1865) > q(\alpha_0) = \ln(100).$$

En appliquant 12)(b) avec $\alpha = 1865$, pour tout $\lambda > 2000$, on obtient

$$P(X \leq 1865) \leq e^{-q(1865)} < e^{-\ln(100)} = \frac{1}{100}.$$

Cela établit, en particulier, la majoration large demandée.

12)(e) Le comptage observé 1600 est inférieur au seuil 1865. La règle consistant à autoriser la baignade lorsque le nombre observé de bactéries est inférieur ou égal à 1865 conduit donc à autoriser le bassin.

Sous le modèle de Poisson retenu pour X , pour toute concentration dépassant la limite tolérée, le nombre moyen de bactéries dans un litre vérifie $\lambda > 2000$. La question 12)(d) garantit alors

$$P_\lambda(\text{autoriser le bassin}) = P_\lambda(X \leq 1865) \leq 1/100.$$

Ainsi, la procédure présente un risque d'autoriser à tort un bassin non conforme d'au plus 1 %, uniformément pour les paramètres $\lambda > 2000$. Ce risque concerne la répétition du prélèvement lorsque la concentration réelle est fixée et excessive. Il ne signifie pas que, conditionnellement à l'observation 1600, la probabilité que le bassin soit pollué est de 1 % : aucun modèle probabiliste sur la concentration inconnue n'est donné.

La garantie de cette troisième partie est établie dans le modèle de Poisson explicitement adopté. Si l'on revient au nombre N de la partie II, la probabilité d'autorisation erronée est majorée par $1/100$ augmenté d'une borne de l'erreur d'approximation issue de 8), pour le volume effectif du bassin. Les données de cette dernière question ne fixent pas ce volume ; on ne peut donc pas chiffrer davantage cette correction.

Relevé des adaptations

- 7) Question adaptée au programme actuel. Code adapté en Python : la demande de fonction dans l'ancien langage est remplacée par une fonction Python de même objectif, avec documentation des commandes utiles. Le corrigé fournit le calcul récursif élémentaire et une variante évitant l'initialisation numérique par une probabilité trop petite lorsque le paramètre est élevé.