Devoir Surveillé 5 - Avec morceau HEC
Séries, probabilités discrètes, applications linéaires en dimension finie

La calculatrice est interdite. Durée de l'épreuve : 4h.

1. Lire le sujet en entier et en détail avant de commencer.
2. Une importance particulière sera attribuée à la rigueur de la rédaction et des justifications, qui devront être concises mais précises.
3. Une prèsentation agréable sera appréciée : laissez une marge à gauche, la première page libre, et encadrez vos résultats. **L'absence de respect de cette consigne entraîne une minoration de la note.**

Exercice 1 (Questions de cours)

1. Rappeler le théorème du rang et sa preuve.
2. Convergence absolue des séries : résultat et preuve.
3. Rappeler les formules d'espérance et de variance pour les lois de Poissons et les lois géométriques, et les démontrer.

Exercice 2 (EML ECS)
Soit n un entier naturel supérieur ou égal à 2. On note $E = \mathbb{R}_n[x]$ et $\mathscr{B} = (1, x, \ldots, x^n)$ la base canonique de E. On note, pour tout polynôme P de E :

$$\varphi(P) = \frac{1}{n}x(1-x)P' + xP. \qquad (P' \text{ étant le polynôme dérivé de } P)$$

1. a) Montrer que φ est une application linéaire.
   b) Calculer $\varphi(x^n)$.
   c) Montrer que φ est un endomorphisme de E.
2. Déterminer la matrice A de φ dans la base $\mathscr{B}$. Préciser le rang de cette matrice.
3. a) L'endomorphisme φ est-il injectif ? Justifier votre réponse.
   b) Soit P un polynôme non nul de Ker (φ).
   Montrer que P admet 1 comme racine puis montrer que l'ordre de multiplicité de 1 est n.
   c) En déduire une base de Ker (φ).
4. On pose, pour tout k de $[\![0, n]\!]$: $P_k = x^k(1-x)^{n-k}$.
   a) Pour tout k de $[\![0, n]\!]$, calculer $\varphi(P_k)$.
   b) Montrer que la famille $(P_0, P_1, \ldots, P_n)$ est une base de E.
   c) Donner la matrice B de φ dans cette base.

1. a) Soient $P, Q \in E$ et $\lambda, \mu \in \mathbb{R}$. On a :

$$\begin{aligned}\varphi(\lambda P + \mu Q) &= \frac{1}{n}x(1-x)(\lambda P + \mu Q)' + x(\lambda P + \mu Q)\\ &= \frac{1}{n}x(1-x)(\lambda P' + \mu Q') + x(\lambda P + \mu Q)\\ &= \lambda\left(\frac{1}{n}x(1-x)P' + xP\right) + \mu\left(\frac{1}{n}x(1-x)Q' + xQ\right)\\ &= \lambda\varphi(P) + \mu\varphi(Q)\end{aligned}$$

Donc φ est une application linéaire.

b) On a $(x^n)' = nx^{n-1}$, donc :

$$\begin{aligned}\varphi(x^n) &= \frac{1}{n}x(1-x)\cdot nx^{n-1} + x\cdot x^n \\ &= x^n(1-x) + x^{n+1} \\ &= x^n - x^{n+1} + x^{n+1} \\ &= x^n\end{aligned}$$

c) Soit $P \in E$ de degré $d \leq n$. Alors P' est de degré $d-1 \leq n-1$ (ou nul si $d=0$).

Le terme $\frac{1}{n}x(1-x)P'$ est de degré au plus $(d-1)+2 = d+1$.

Le terme xP est de degré $d+1$.

Calculons le coefficient dominant de $\varphi(P)$. Si $P = a_d x^d + \dots$ avec $a_d \neq 0$:

- Le coefficient de x^{d+1} dans xP est a_d.
- Le coefficient de x^{d+1} dans $\frac{1}{n}x(1-x)P'$ provient de $-\frac{1}{n}x^2P'$, soit $-\frac{1}{n}\cdot d\cdot a_d$.

Donc le coefficient de x^{d+1} dans $\varphi(P)$ est $a_d\left(1-\frac{d}{n}\right)$.

Si $d < n$, ce coefficient est non nul, donc $\deg(\varphi(P)) = d+1 \leq n$.

Si $d = n$, on a vu que $\varphi(x^n) = x^n$, et plus généralement le coefficient de x^{n+1} s'annule.

Donc pour tout $P \in E$, on a $\varphi(P) \in E$: φ est un endomorphisme de E.

2. Calculons $\varphi(x^k)$ pour $k \in [\![0, n]\!]$:

$$\begin{aligned}\varphi(x^k) &= \frac{1}{n}x(1-x)\cdot kx^{k-1} + x^{k+1} \\ &= \frac{k}{n}x^k(1-x) + x^{k+1} \\ &= \frac{k}{n}x^k - \frac{k}{n}x^{k+1} + x^{k+1} \\ &= \frac{k}{n}x^k + \left(1-\frac{k}{n}\right)x^{k+1}\end{aligned}$$

Pour $k=0$: $\varphi(1) = x$.

La matrice A de φ dans la base $\mathscr{B}$ est :

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & \frac{1}{n} & 0 & \cdots & 0 & 0 \\ 0 & 1-\frac{1}{n} & \frac{2}{n} & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1-\frac{n-2}{n} & \frac{n-1}{n} & 0 \\ 0 & 0 & \cdots & 0 & 1-\frac{n-1}{n} & 1 \end{pmatrix}$$

La matrice A est triangulaire inférieure. Ses coefficients diagonaux sont $0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1$.

Par un pivot de Gauss (On décale toutes les lignes vers le haut et on mets la ligne de 0 en bas), on retrouve une matrice échelonnée d'où $\boxed{\mathrm{rg}(A) = n.}$

3. a) L'endomorphisme φ n'est pas injectif car $\text{rg}(\varphi) = n < n+1 = \dim(E)$.

Autrement dit, Ker $(\varphi) \neq \{0\}$ puisque d'après le théorème du rang, $\dim(E) = \text{rg}(\varphi) + \dim(\text{Ker }(\varphi))$. On a même :

$$\dim(\text{Ker }(\varphi)) = \dim(E) - \text{rg}(\varphi) = (n+1) - n = 1$$

b) Soit $P \in \text{Ker }(\varphi)$ non nul. On a $\varphi(P) = 0$, soit :

$$\frac{1}{n}x(1-x)P' + xP = 0$$

En évaluant en $x = 1$: $1 \cdot P(1) = 0$, donc $P(1) = 0$. Ainsi 1 est racine de P.

Supposons que 1 est racine d'ordre $m \geq 1$ de P, c'est-à-dire $P = (x-1)^m Q$ avec $Q(1) \neq 0$.

Alors $P' = m(x-1)^{m-1}Q + (x-1)^m Q'$.

On a :

$$\begin{aligned}
0 &= \frac{1}{n}x(1-x)P' + xP \\
&= \frac{1}{n}x(1-x)\left[m(x-1)^{m-1}Q + (x-1)^m Q'\right] + x(x-1)^m Q \\
&= -\frac{1}{n}x(x-1)^m\left[mQ + (x-1)Q'\right] + x(x-1)^m Q \\
&= x(x-1)^m\left[-\frac{m}{n}Q - \frac{1}{n}(x-1)Q' + Q\right] \\
&= x(x-1)^m\left[\left(1-\frac{m}{n}\right)Q - \frac{1}{n}(x-1)Q'\right]
\end{aligned}$$

Comme c'est une égalité de polynômes, et que le premier facteur $x(x-1)^m$ n'est pas nul, on a :

$$\left(1-\frac{m}{n}\right)Q - \frac{1}{n}(x-1)Q' = 0$$

En évaluant en $x = 1$: $\left(1-\frac{m}{n}\right)Q(1) = 0$.

Comme $Q(1) \neq 0$, on a $1 - \frac{m}{n} = 0$, soit $m = n$.

Donc 1 est racine d'ordre exactement n de P.

(En fait ici, on aurait pu "tricher" en calculant $\varphi((x-1)^n)$. On trouve $\varphi((x-1)^n) = 0$, donc $(x-1)^n \in$ Ker (φ), et comme $\dim($*Ker* $(\varphi)) = 1$*, on en déduit que* Ker $(\varphi) = \text{Vect}((x-1)^n)$*, d'où le résultat.)*

c) On a $P \in \text{Ker }(\varphi)$ non nul $\Rightarrow P$ est divisible par $(x-1)^n$.

Comme $\deg(P) \leq n$, on a $P = \lambda(x-1)^n$ pour un certain $\lambda \in \mathbb{R}^*$.

Mais alors Ker $(\varphi) \subset \text{Vect}((1-x)^n)$ et par égalité de dimensions, on a Ker $(\varphi) = \text{Vect}((1-x)^n)$.

Donc Ker $(\varphi) = \text{Vect}((1-x)^n)$ et $\boxed{\{(1-x)^n\}}$ est une base de Ker (φ).

4. a) Pour $k \in [\![0, n]\!]$, on a $P_k = x^k(1-x)^{n-k}$.

Calculons P_k' pour $k \in [\![1, n]\!]$:

$$P_k' = kx^{k-1}(1-x)^{n-k} + x^k \cdot (-(n-k))(1-x)^{n-k-1} = kx^{k-1}(1-x)^{n-k} - (n-k)x^k(1-x)^{n-k-1}$$

Donc :

$$\begin{aligned}
\varphi(P_k) &= \frac{1}{n}x(1-x)P_k' + xP_k \\
&= \frac{1}{n}x(1-x)\left[kx^{k-1}(1-x)^{n-k} - (n-k)x^k(1-x)^{n-k-1}\right] + x^{k+1}(1-x)^{n-k} \\
&= \frac{k}{n}x^k(1-x)^{n-k+1} - \frac{n-k}{n}x^{k+1}(1-x)^{n-k} + x^{k+1}(1-x)^{n-k} \\
&= \frac{k}{n}x^k(1-x)^{n-k+1} + \frac{k}{n}x^{k+1}(1-x)^{n-k} \\
&= \frac{k}{n}x^k(1-x)^{n-k}\left[(1-x)+x\right] \\
&= \frac{k}{n}x^k(1-x)^{n-k} \\
&= \frac{k}{n}P_k
\end{aligned}$$

On a aussi $\varphi(P_0) = \varphi((1-x)^n) = 0$ d'après la question précédente, ce qui correspond à $k = 0$.

Donc $\boxed{\varphi(P_k) = \frac{k}{n}P_k}$ pour tout $k \in [\![0, n]\!]$.

b) La famille $(P_0, P_1, \ldots, P_n)$ contient $n+1$ vecteurs dans E de dimension $n+1$.

Il suffit de montrer qu'elle est libre.

Remarquons que pour tout $k \in [\![0, n]\!]$, on a $P_k = x^k(1-x)^{n-k} = x^k \cdot (1 + xQ_k(x))$ pour un polynôme Q_k.

Donc le terme de plus bas degré de P_k est x^k, c'est-à-dire que $P_k = x^k +$ (termes de degré $\geq k+1$).

Soit $(\alpha_0, \alpha_1, \ldots, \alpha_n) \in \mathbb{R}^{n+1}$ tel que $\sum_{k=0}^{n} \alpha_k P_k = 0$.

Montrons par récurrence que $\alpha_k = 0$ pour tout k.

Le terme constant de $\sum_{k=0}^{n} \alpha_k P_k$ ne provient que de P_0 (car P_k a un terme de degré $\geq k$ pour $k \geq 1$).

Or le terme constant de $P_0 = (1-x)^n$ est 1. Donc $\alpha_0 \cdot 1 = 0$, soit $\alpha_0 = 0$.

Supposons que $\alpha_0 = \alpha_1 = \cdots = \alpha_{j-1} = 0$ pour un certain $j \in [\![1, n]\!]$.

Le coefficient de x^j dans $\sum_{k=j}^{n} \alpha_k P_k$ ne provient que de P_j (car P_k a un terme de degré $\geq k > j$ pour $k > j$).

Or le coefficient de x^j dans $P_j = x^j(1-x)^{n-j}$ est 1. Donc $\alpha_j = 0$.

Par récurrence, $\alpha_k = 0$ pour tout $k \in [\![0, n]\!]$.

Donc $(P_0, P_1, \ldots, P_n)$ est libre. Comme elle contient $n+1$ vecteurs dans un espace de dimension $n+1$, c'est une base de E.

c) Puisque $\varphi(P_k) = \frac{k}{n}P_k$, chaque P_k est vecteur propre de φ pour la valeur propre $\frac{k}{n}$.

La matrice de φ dans la base $(P_0, P_1, \ldots, P_n)$ est diagonale :

$$\boxed{B = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{n} & 0 & \cdots & 0 \\ 0 & 0 & \frac{2}{n} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} = \operatorname{diag}\left(0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right)}$$

Problème - morceau de sujet HEC

Dans ce problème, toutes les variables aléatoires sont supposées définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. *Si* X *est une variable aléatoire, on note respectivement* $\mathrm{E}(X)$ *et* $\mathrm{Var}(X)$ *son espérance et sa variance, sous réserve d'existence.*

Le but de ce problème est de mettre en évidence quelques résultats asymptotiques liés au modèle du collectionneur de vignettes. Dans chaque paquet de céréales se trouve une vignette et il y a en tout des vignettes de n types différents, où n est un entier supérieur ou égal à 1 . Chacun des n types de vignettes se retrouve avec la même fréquence dans les paquets de céréales. Une collection est alors complète lorsqu'elle comporte n vignettes de types différents.

On modélise le nombre total de paquets de céréales qu'il est nécessaire d'acheter pour obtenir la collection complète de n vignettes de types différents par la variable aléatoire notée C_n.

On pose par convention $C_0 = 0$ et pour tout $i \in [\![1, n]\!]$, on note C_i le nombre d'achats de paquets de céréales nécessaires pour obtenir i vignettes de types différents.

De même, pour tout $i \in [\![1, n]\!]$, on pose $X_i = C_i - C_{i-1}$, qui représente le nombre d'achats supplémentaires de paquets de céréales qu'il est nécessaire d'effectuer pour obtenir une nouvelle vignette d'un type différent des $(i-1)$ vignettes de types différents déjà obtenues. Par convention, on pose $X_1 = C_1 = 1$.

On suppose que les variables aléatoires $X_1, \ldots, X_n$ sont mutuellement indépendantes. Enfin, on pose :

$$V_n = \frac{C_n}{n} - \ln(n).$$

Questions préliminaires

1. a) Montrer que :

$$C_n = X_1 + \cdots + X_n = \sum_{i=1}^{n} X_i.$$

b) Justifier que pour tout $i \in [\![1, n]\!]$, la variable aléatoire X_i suit la loi géométrique $\mathcal{G}\left(\frac{n-i+1}{n}\right)$.

Partie I

Pour tout entier $n \geq 1$, on pose :

$$H_n = \sum_{k=1}^{n} \frac{1}{k} \quad , \quad S_n = \sum_{k=1}^{n} \frac{1}{k^2}.$$

Nous allons démontrer la convergence de la suite $(S_n)_{n\geq 1}$ *et déterminer une valeur approchée de sa limite* S.

2. a) Montrer que pour tout entier $k \geq 2$, on a l'encadrement :

$$\frac{1}{k} - \frac{1}{k+1} \leq \frac{1}{k^2} \leq \frac{1}{k-1} - \frac{1}{k}.$$

b) Montrer que pour tout entier $n \geq 2$, on a :

$$\frac{3}{2} - \frac{1}{n+1} \leq S_n \leq 2 - \frac{1}{n}.$$

c) Montrer que la suite $(S_n)_{n\geq 1}$ est convergente et donner un encadrement de sa limite S.

d) Montrer que pour tout entier $n \geq 1$, on a l'encadrement :

$$\frac{1}{n+1} \leq S - S_n \leq \frac{1}{n}.$$

e) En déduire un programme `Python` qui permet d'obtenir une valeur approchée de S à 10^{-7}-près.

3. Pour tout entier $n \geq 1$, on pose $u_n = H_n - \ln(n)$.

   a) Montrer que pour tout entier $n \geq 1$, on a l'encadrement :

   $$\frac{1}{n+1} \leq \ln\left(\frac{n+1}{n}\right) \leq \frac{1}{n}.$$

   b) Montrer que pour tout entier $n \geq 1$, on a : $0 \leq u_n \leq 1$.

   c) Montrer que la suite $(u_n)_{n\geq 1}$ converge vers une certaine limite $\gamma \in [0,1]$.

4. Justifier l'existence de l'espérance de C_n et montrer que $E(C_n) = nH_n$.

5. Justifier l'existence de la variance de C_n et exprimer $\mathrm{Var}(C_n)$ en fonction de n, S_n et H_n.

6. a) Montrer que pour tout réel $a > 0$, on a :

   $$\mathbb{P}(|C_n - nH_n| \geq an) \leq \frac{S}{a^2}.$$

   b) Montrer que pour tout réel $c > 1$, on a :

   $$\mathbb{P}\left(\left|\frac{C_n}{n} - \ln(n)\right| \geq c\right) \leq \frac{S}{(c-1)^2}.$$

   c) Pour cette question, on suppose que $n = 10^6$.

   On donne l'approximation $\ln(n) \approx 13.816$.

   Montrer que :

   $$\mathbb{P}\left(\frac{C_n}{n} \in [7.81, 19.82]\right) \geq 0.92.$$

Partie III - la partie II n'est pas faisable en ECG1

Dans cette partie, on suppose que $n \geq 2$. On suppose de plus que les vignettes sont numérotées de 1 à n. Pour tout $i \in [\![1,n]\!]$ et tout entier $m \geq 1$, on note $A_{i,m}$ la variable aléatoire donnant le nombre de vignettes numérotées i obtenues dans les m premiers paquets de céréales achetés.

7. Justifier que pour tout $i \in [\![1,n]\!]$, la variable aléatoire $A_{i,m}$ suit la loi binomiale $\mathcal{B}\left(m, \frac{1}{n}\right)$.

8. a) Montrer que si $(E_i)_{i\geq 1}$ est une famille d'événements, alors pour tout entier $r \geq 1$:

   $$\mathbb{P}\left(\bigcup_{i=1}^{r} E_i\right) \leq \sum_{i=1}^{r} \mathbb{P}(E_i).$$

   b) Montrer que :

   $$\mathbb{P}(C_n > m) \leq n\left(1 - \frac{1}{n}\right)^m \leq n\mathrm{e}^{-\frac{m}{n}}.$$

   c) **Il y a ici une erreur de conception de sujet. Admettez cette question.** Soit c un réel strictement positif. Montrer qu'on a la majoration :

   $$\mathbb{P}(C_n > cn\ln(n)) \leq n^{1-c}.$$

d) Montrer que pour tout réel $x > -\ln(n)$, on a :

$$\mathbb{P}(V_n > x) \leq \mathrm{e}^{-x}.$$

Dans la suite de cette partie, on introduit un modèle légèrement différent : le nombre N de paquets achetés est décrit par une variable aléatoire N de loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda > 0$. On cherche à calculer la probabilité de compléter, à partir des N vignettes obtenues, la collection de vignettes. On suppose toujours que les vignettes sont numérotées de 1 à n. On note $\widetilde{A}_i$ la variable aléatoire donnant le nombre de vignettes numérotées i obtenues dans les N paquets de céréales achetés.

9. a) Soit $p \in \mathbb{N}^*$. Justifier que pour tout $i \in [\![1, n]\!]$, la loi conditionnelle de $\widetilde{A}_i$ sachant l'événement $[N = p]$ est la loi binomiale $\mathcal{B}\left(p, \frac{1}{n}\right)$ et en déduire que la variable aléatoire $\tilde{A}_i$ suit une loi de Poisson dont on déterminera le paramètre.

b) Montrer que pour tout n-uplet $(k_1, \ldots, k_n)$ avec $k_i \in \mathbb{N}$ et $k_1 + \cdots + k_n = p$, on a:

$$\mathbb{P}_{[N=p]}\left(\left[\tilde{A}_1 = k_1\right] \cap \cdots \cap \left[\tilde{A}_n = k_n\right]\right) = \frac{p!}{k_1! \cdots k_n!} \times \frac{1}{n^p}.$$

c) Montrer que pour tout n-uplet $(k_1, \ldots, k_n)$ avec $k_i \in \mathbb{N}$, on a :

$$\mathbb{P}\left(\left[\tilde{A}_1 = k_1\right] \cap \cdots \cap \left[\tilde{A}_n = k_n\right]\right) = \prod_{i=1}^{n} \frac{\mathrm{e}^{-\frac{\lambda}{n}} \left(\frac{\lambda}{n}\right)^{k_i}}{k_i!}.$$

d) Montrer que les variables aléatoires $\tilde{A}_1, \ldots, \tilde{A}_n$ sont mutuellement indépendantes. Commenter en comparant avec le résultat de la question 15.(c).

10. Soit D_n l'événement « À l'issue des N achats de paquets de céréales, la collection de vignettes est complète ».

Montrer que :

$$\mathbb{P}(D_n) = \left(1 - \mathrm{e}^{-\frac{\lambda}{n}}\right)^n.$$

On admet le résultat suivant : pour tout $p \in \mathbb{N}^*$, on a $\mathbb{P}_{[N=p]}(D_n) = \mathbb{P}(C_n \leq p)$.

11. a) Montrer que :

$$\mathbb{P}(D_n) = \sum_{p=0}^{\infty} \mathbb{P}(C_n \leq p)\, \mathbb{P}(N = p).$$

b) On suppose maintenant que $\lambda > 1$.

Soit $a \in]\sqrt{\lambda}, \lambda[$. On pose $k_1 = \lfloor \lambda - a \rfloor$ et $k_2 = \lfloor \lambda + a \rfloor + 1$.

Montrer que :

$$\sum_{p=k_1}^{k_2} \mathbb{P}(N = p) \geq 1 - \frac{\lambda}{a^2}.$$

c) Montrer l'encadrement :

$$\mathbb{P}(C_n \leq k_1)\left(1 - \frac{\lambda}{a^2}\right) \leq \mathbb{P}(D_n) \leq \mathbb{P}(C_n \leq k_2) + \frac{\lambda}{a^2}.$$

12. Soit $(c_n)_{n \geq 1}$ une suite convergente de nombres réels.

Montrer que pour tout entier n suffisamment grand, on a l'encadrement :

$$\mathbb{P}\left(V_n \leq c_n - \frac{1}{n^{1/3}} - \frac{1}{n}\right)\left(1 - \frac{\ln(n) + c_n}{n^{1/3}}\right) \leq \left(1 - \frac{e^{-c_n}}{n}\right)^n \leq \mathbb{P}\left(V_n \leq c_n + \frac{1}{n^{1/3}} + \frac{1}{n}\right) + \frac{\ln(n) + c_n}{n^{1/3}}.$$

On pourra appliquer la question précédente avec $\lambda = n\ln(n) + c_n n$ et $a = n^{2/3}$.

Questions préliminaires

1. a) Pour tout $i \in [\![1, n]\!]$, $X_i = C_i - C_{i-1}$. En sommant pour $i \in [\![1, n]\!]$, on obtient une somme télescopique :

$$\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} C_i - C_{i-1} = C_n - C_0 = \boxed{C_n}.$$

b) Pour $i \in [\![1, n]\!]$, la variable aléatoire X_i correspond au nombre d'achats à effectuer, lorsque l'on possède déjà $i-1$ vignettes différentes, pour obtenir l'une des $n-i+1$ vignettes que l'on ne possède pas encore.
À chaque achat d'un paquet de céréales, la probabilité d'obtenir l'une de ces nouvelles vignettes est $\dfrac{n-i+1}{n}$. Par ailleurs, chaque achat d'un paquet de céréales est indépendant des achats précédents et des vignettes déjà obtenues. On reconnaît donc la loi géométrique de paramètre $\dfrac{n-i+1}{n}$:

$$\boxed{\forall i \in [\![1, n]\!], \quad X_i \hookrightarrow \mathcal{G}\left(\frac{n-i+1}{n}\right).}$$

Partie I

2. a) Pour $k \geq 2$, on remarque que :

$$(0 <) \quad k(k-1) \leq k^2 \leq k(k+1)$$

et la fonction inverse étant décroissante sur $]0, +\infty[$, on obtient :

$$\frac{1}{k} - \frac{1}{k+1} = \frac{1}{k(k+1)} \leq \frac{1}{k} \leq \frac{1}{k(k-1)} = \frac{1}{k-1} - \frac{1}{k}.$$

b) Soit $n \geq 2$. On somme les inégalités ci-dessus pour $k \in [\![2, n]\!]$:

$$\sum_{k=2}^{n} \left(\frac{1}{k} - \frac{1}{k+1}\right) \leq \sum_{k=2}^{n} \frac{1}{k} \leq \sum_{k=2}^{n} \left(\frac{1}{k-1} - \frac{1}{k}\right).$$

Dans les membres de gauche et de droite, on reconnaît des sommes télescopiques. Dans le membre du milieu, on reconnaît la somme définissant S_n privé de son premier terme :

$$\frac{1}{2} - \frac{1}{n+1} \leq S_n - 1 \leq 1 - \frac{1}{n}$$

et on en déduit :

$$\boxed{\frac{3}{2} - \frac{1}{n+1} \leq S_n \leq 2 - \frac{1}{n}.}$$

c) *La convergence de la suite $(S_n)_{n\in\mathbb{N}}$ est un résultat de cours puisqu'il s'agit de la somme partielle d'une série de Riemann convergente. Mais l'énoncé nous demande de prouver ce résultat ici.*

La suite $(S_n)_{n\in\mathbb{N}}$ est croissante car :

$$\forall n \geq 1, \quad S_{n+1} - S_n = \frac{1}{(n+1)^2} \geq 0$$

et elle est majorée (par 2) d'après la question précédente. Ainsi, $\boxed{\text{la suite } (S_n)_{n\in\mathbb{N}} \text{ converge.}}$
L'encadrement déterminé pour S_n en question précédente donne: $\boxed{\frac{3}{2} \leq S \leq 2}$.

La valeur exacte de S, qu'il n'était bien sûr pas question de calculer ici, est $\dfrac{\pi^2}{6}$.

d) Soit $n \geq 1$, et prenons $N \geq n+1$. Sommons les inégalités obtenues à la question 2.(a) pour $k = n+1, \ldots, N$:

$$\sum_{k=n+1}^{N} \left(\frac{1}{k} - \frac{1}{k+1}\right) \leq \sum_{k=n+1}^{N} \frac{1}{k^2} \leq \sum_{k=n+1}^{N} \left(\frac{1}{k-1} - \frac{1}{k}\right).$$

D'où :

$$\frac{1}{n+1} - \frac{1}{N+1} \leq S_N - S_n \leq \frac{1}{n} - \frac{1}{N}.$$

Et en faisant tendre N vers $+\infty$ dans ces inégalités (en notant bien que tout converge) :

$$\boxed{\frac{1}{n+1} \leq S - S_n \leq \frac{1}{n}.}$$

e) Pour $n = 10^7$, on a $\dfrac{1}{n} = 10^{-7}$ et $\dfrac{1}{n+1} < 10^{-7}$, donc le nombre S_{10^7} est, d'après la question précédente, une valeur approchée de S à 10^{-7} près. Ce nombre peut être obtenu avec le code `Python` suivant :

```
S = 0
for i in range(1,10**7+1)
    S = S+1/(i**2)
print(S)
```

3. a) *Les questions 3.(a) et 3.(b) s'enchaînent de la même manière que les questions 2.(a) et 2.(b).*
Il est plus lisible ici d'utiliser la lettre k au lieu de n, comme l'énoncé l'avait d'ailleurs bien fait en question 2.(a).

Soit $k \geq 1$. En remarquant que :

$$\forall t \in \left[\frac{1}{k+1}, \frac{1}{k}\right], \quad \frac{1}{k+1} \leq \frac{1}{t} \leq \frac{1}{k}$$

la fonction $t \mapsto \dfrac{1}{t}$ étant décroissante sur $\left[\dfrac{1}{k+1}, \dfrac{1}{k}\right]$, la croissance de l'intégrale donne :

$$\int_k^{k+1} \frac{1}{k+1}\,\mathrm{d}t \leq \int_k^{k+1} \frac{1}{t}\,\mathrm{d}t \leq \int_k^{k+1} \frac{1}{k}\,\mathrm{d}t$$

soit le résultat demandé :

$$\boxed{\frac{1}{k+1} \leq \ln\left(\frac{k+1}{k}\right) \leq \frac{1}{k}.}$$

b) Soit $n \geq 1$. On somme les inégalités précédentes membre à membre pour $k \in [\![1, n]\!]$:

$$\sum_{k=1}^{n} \frac{1}{k+1} \leq \sum_{k=1}^{n} \underbrace{\ln\left(\frac{k+1}{k}\right)}_{=\ln(k+1)-\ln(k)} \leq \sum_{k=1}^{n} \frac{1}{k}.$$

On obtient une somme télescopique dans le membre du milieu :

$$H_{n+1} - 1 \leq \ln(n+1) \leq H_n$$

ce qui donne :

- avec l'inégalité de gauche :

$$\forall n \geq 2, \quad u_n = H_n - \ln(n) \leq 1$$

ce qui reste vrai avec $n = 1$;

- avec l'inégalité de droite :

$$\ln(n) \leq \ln(n+1) \leq H_n$$

donc

$$\boxed{\forall n \geq 1, \quad 0 \leq u_n \leq 1.}$$

c) Pour tout $n \geq 1$:

$$\begin{aligned} u_{n+1} - u_n &= H_{n+1} - H_n - \ln(n+1) + \ln(n) \\ &= \frac{1}{n+1} - \ln\left(\frac{n+1}{n}\right) \\ &\geq 0 \qquad \text{(d'après 3.(a)).} \end{aligned}$$

La suite u est donc croissante. Elle est de plus majorée (par 1) d'après 3.(b). Ainsi, $\boxed{\text{la suite } u \text{ converge.}}$
Sa limite vérifie le même encadrement large : $\boxed{\gamma \in [0,1]}$.

4. Chaque variable aléatoire X_i ($i \in [\![1,n]\!]$) suit la loi géométrique de paramètre $\dfrac{n-i+1}{n}$, donc admet une espérance égale à $\dfrac{n}{n-i+1}$.
Comme somme de variables aléatoires admettant chacune une espérance, $C_n = X_1 + \cdots + X_n$ admet une espérance. Par linéarité de E :

$$\mathbb{E}(C_n) = \sum_{i=1}^{n} \mathbb{E}(X_i) = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \sum_{i=1}^{n} \frac{1}{n-i+1}.$$

Avec le changement d'indice $j = n - i + 1$:

$$\boxed{\mathbb{E}(C_n) = n \sum_{j=1}^{n} \frac{1}{j} = nH_n.}$$

L'énoncé aurait pu nous demander ici de vérifier que :

- $\mathbb{E}(C_n) \underset{n\to+\infty}{\sim} n\ln(n)$ *(ce qui justifie de s'intéresser à* $V_n = \dfrac{C_n}{n} - \ln(n)$*),*
- $\mathbb{E}(V_n) = u_n$*,*
- *et donc* $\mathbb{E}(V_n) \underset{n\to+\infty}{\longrightarrow} \gamma$*.*

où γ *est la constante d'Euler-Mascheroni, limite de la suite* $(u_n)_{n\in\mathbb{N}}$*.*

5. Chaque variable aléatoire X_i ($i \in [\![1,n]\!]$) admet aussi une variance égale à :

$$\frac{1 - \dfrac{n-i+1}{n}}{\left(\dfrac{n-i+1}{n}\right)^2} = \frac{i-1}{n} \times \frac{n^2}{(n-i+1)^2} = n \times \frac{i-1}{(n-i+1)^2}.$$

Comme somme de variables aléatoires admettant chacune une variance, $C_n = X_1 + \cdots + X_n$ admet une variance. Par indépendance mutuelle de $X_1, \ldots, X_n$:

$$\mathbb{V}(C_n) = \sum_{i=1}^{n} \mathbb{V}(X_i) = n \sum_{i=1}^{n} \frac{i-1}{(n-i+1)^2}$$

Avec le changement d'indice $j = n - i + 1$:

$$\boxed{\mathbb{V}(C_n) = n\sum_{i=1}^{n}\frac{n-j}{j^2} = n^2\sum_{i=1}^{n}\frac{1}{j^2} - n\sum_{i=1}^{n}\frac{1}{j} = n^2 S_n - nH_n.}$$

L'énoncé aurait pu nous demander ici de vérifier que:

$$\mathbb{V}(V_n) = S_n - \frac{H_n}{n} \underset{n\to+\infty}{\longrightarrow} S$$

6. a) **Cette question était difficile pour vous, car elle demande de se souvenir de Bienaymé-Tchébychev que vous avez vu en terminale.** *Nous reverrons cette inégalité et l'inégalité de Markov dans un prochain chapitre.* La variable aléatoire C_n admettant une variance, l'inégalité de Bienaymé-Tchébychev s'écrit :

$$\forall b > 0, \quad \mathbb{P}(|C_n - \mathbb{E}(C_n)| \geq b) \leq \frac{\mathbb{V}(C_n)}{b^2}.$$

Avec l'espérance et la variance de C_n trouvées en questions précédentes, et en choisissant $b = an$:

$$\forall a > 0, \quad \mathbb{P}(|C_n - nH_n| \geq an) \leq \overbrace{\frac{n^2 S_n - nH_n}{a^2 n^2}}^{H_n \geq 0} \leq \frac{S_n}{a^2}.$$

Enfin, la suite $(S_n)_{n\in\mathbb{N}}$ étant croissante, elle est majorée par sa limite. Donc :

$$\boxed{\forall a > 0, \quad \mathbb{P}(|C_n - nH_n| \geq an) \leq \frac{S}{a^2}.}$$

b) **Encore une fois, question difficile. C'est souvent le cas en fin de partie. Mon conseil : en devoir, si vous bloquez sur cette question, vous passez à la suite. Par contre je recommande fortement de la reprendre à la maison, c'est très typique des raisonnements en probabilités.**
Soit $c > 1$. On pose $a = c - 1$ (> 0). Avec la question précédente :

$$\mathbb{P}\Big(|C_n - nH_n| \geq (c-1)n\Big) \leq \frac{S}{(c-1)^2}.$$

Travaillons sur l'événement intervenant ici :

$$\begin{aligned}\Big[|C_n - nH_n| \geq (c-1)n\Big] &= \Big[\Big|\frac{C_n}{n} - H_n\Big| \geq (c-1)\Big] \\ &= \Big[\frac{C_n}{n} \notin \,]H_n - c + 1, H_n + c - 1[\Big]\end{aligned}$$

D'après la question 3. :

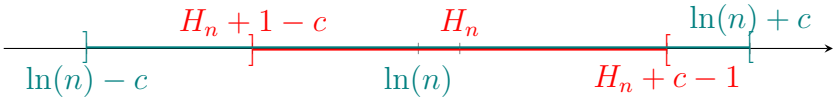
$$0 \leq u_n = H_n - \ln(n) \leq 1$$

donc notamment :

$$\ln(n) \leq H_n \leq H_n + 1 \quad \text{et} \quad H_n - 1 \leq \ln(n)$$

d'où :

$$\ln(n) - c \leq H_n + 1 - c \quad \text{et} \quad H_n + c - 1 \leq \ln(n) + c.$$



On a donc l'inclusion entre les intervalles :

$$\left]H_n - c + 1, H_n + c - 1\right[\subset \left]\ln(n) - c, \ln(n) + c\right[$$

On en déduit l'inclusion d'événements :

$$\left[\frac{C_n}{n} \notin \left]\ln(n) - c, \ln(n) + c\right[\right] \subset \left[\frac{C_n}{n} \notin \left]H_n - c + 1, H_n + c - 1\right[\right]$$

et finalement, par croissance de la probabilité :

$$\begin{aligned}\mathbb{P}\left(\left|\frac{C_n}{n} - \ln(n)\right| \geq c\right) &= \mathbb{P}\left(\frac{C_n}{n} \notin \left]\ln(n) - c, \ln(n) + c\right[\right) \\ &\leq \mathbb{P}\Big(|C_n - nH_n| \geq (c-1)n\Big) \\ &\leq \boxed{\frac{S}{(c-1)^2}.}\end{aligned}$$

c) On a montré en 2.(c) que $S \leq 2$. La question précédente donne :

$$\mathbb{P}\left(\frac{C_n}{n} \in [\ln(n) - c, \ln(n) + c]\right) \geq 1 - \frac{2}{(c-1)^2}$$

donc, avec $c = 6$ et sachant que $13.81 \leq \ln(n) \leq 13.82$:

$$\begin{aligned}\mathbb{P}\left(\frac{C_n}{n} \in [7.81, 19.82]\right) &= \mathbb{P}\left(\frac{C_n}{n} \in [13.81 - 5, 13.82 + 5]\right) \\ &\geq \mathbb{P}\left(\frac{C_n}{n} \in [\ln(n) - 5, \ln(n) + 5]\right) \\ &\geq 1 - \frac{2}{(6-1)^2} \;=\; 1 - 0.08 \;=\; \boxed{0.92.}\end{aligned}$$

Partie III

7. À chaque achat d'un paquet de céréales, la vignette numéro i est obtenue (succés) avec la probabilité $\frac{1}{n}$. Les achats étant indépendants les uns des autres, on reconnaît le modéle de la loi binomiale (on compte le nombre de succés pour m achats) :

$$\boxed{\forall i \in [\![1, n]\!], \quad A_{i,m} \hookrightarrow \mathcal{B}\left(m, \frac{1}{n}\right).}$$

8. a) On procéde par récurrence sur $r \geq 1$.

- *Initialisation.* On a égalité pour $r = 1$.
- *Hérédité.* Soit $r \geq 1$. Supposons que $\mathbb{P}\left(\bigcup_{i=1}^{r} E_i\right) \leq \sum_{i=1}^{r} \mathbb{P}(E_i)$.

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{r+1} E_i\right) &= \mathbb{P}\left(\left(\bigcup_{i=1}^{r} E_i\right) \bigcup E_{r+1}\right) \\ &= \mathbb{P}\left(\bigcup_{i=1}^{r} E_i\right) + \mathbb{P}\left(E_{r+1}\right) - \underbrace{\mathbb{P}\left(\left(\bigcup_{i=1}^{r} E_i\right) \bigcap E_{r+1}\right)}_{\geq 0} \\ &\leq \mathbb{P}\left(\bigcup_{i=1}^{r} E_i\right) + \mathbb{P}\left(E_{r+1}\right) \\ &\underset{\text{H.R.}}{\leq} \sum_{i=1}^{r} \mathbb{P}(E_i) + \mathbb{P}\left(E_{r+1}\right) = \sum_{i=1}^{r+1} \mathbb{P}(E_i).\end{aligned}$$

b) L'événement $[C_n > m]$ signifie que l'album n'est pas encore complet après m achats de paquets de céréales, ce qui veut dire qu'au moins l'une des vignettes numéro i ($i \in [\![1, n]\!]$) n'a encore jamais été obtenue :

$$(C_n > m) = \bigcup_{i=1}^{n} (A_{i,m} = 0).$$

Avec la question précédente :

$$\mathbb{P}(C_n > m) = \mathbb{P}\left(\bigcup_{i=1}^{n}(A_{i,m} = 0)\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_{i,m} = 0) = \sum_{i=1}^{n} \overbrace{\left(1 - \frac{1}{n}\right)^m}^{\text{indépendant de } i} = n\left(1 - \frac{1}{n}\right)^m.$$

L'inégalité classique (que l'on peut obtenir avec une étude de fonction)

$$\forall y \in \mathbb{R}, \quad 1 + y \leq \mathrm{e}^y$$

appliquée à $y = -\dfrac{1}{n}$ donne :

$$\boxed{\mathbb{P}(C_n > m) \leq n\left(1 - \frac{1}{n}\right)^m \leq n\mathrm{e}^{-\frac{m}{n}}.}$$

c) On pourrait être tenté de choisir $m = cn\ln(n)$ dans la question précédente, de sorte que :

$$\mathbb{P}\Big(C_n > cn\ln(n)\Big) \leq n\mathrm{e}^{-\frac{cn\ln(n)}{n}} \leq n\mathrm{e}^{-\frac{cn\ln(n)}{n}} = nn^{-c} = n^{1-c}.$$

Mais dans les questions précédentes, m est un entier naturel non nul.

Alors on essaie de s'en sortir, en vain, avec une partie entière...

- Pour $m = \lfloor cn\ln(n)\rfloor$, c'est un entier naturel qui peut être nul et surtout :

$$\mathbb{P}\Big(C_n > cn\ln(n)\Big) \leq \mathbb{P}\Big(C_n > \lfloor cn\ln(n)\rfloor\Big) \leq \underbrace{\mathrm{e}^{-\frac{\lfloor cn\ln(n)\rfloor}{n}} \not\leq \mathrm{e}^{-\frac{cn\ln(n)}{n}}}_{\text{(l'inégalité dans l'autre sens est vraie)}}$$

- Pour $m = \lfloor cn\ln(n)\rfloor + 1$, cette fois $m \in \mathbb{N}^*$, mais on est bloqué ici :

$$\mathbb{P}\Big(C_n > cn\ln(n)\Big) = \underbrace{\mathbb{P}(C_n \geq m) \not\leq \mathbb{P}(C_n > m)}_{\text{(l'inégalité dans l'autre sens est vraie)}}.$$

Dommage, car ça aurait bien fonctionné ensuite :

$$\mathbb{P}(C_n > m) \leq n\mathrm{e}^{-\frac{\lfloor cn\ln(n)\rfloor + 1}{n}} \leq n\mathrm{e}^{-\frac{cn\ln(n)}{n}} = \dots.$$

Conclusion: je ne sais pas comment répondre à cette question ! (sauf si on suppose que $cn\ln(n) \in \mathbb{N}^*$, mais on sera alors bloqué dans la question suivante). C'est une erreur de l'énoncé!

d) Pour tout $x \in \mathbb{R}$:

$$\begin{aligned}
\mathbb{P}(V_n > x) &= \mathbb{P}\left(\frac{C_n}{n} - \ln(n) > x\right) \\
&= \mathbb{P}\left(\frac{C_n}{n} > x + \ln(n)\right) \\
&= \mathbb{P}\left(C_n > \left(\frac{x}{\ln(n)} + 1\right) n\ln(n)\right).
\end{aligned}$$

On pose $c = \dfrac{x}{\ln(n)} + 1$.
Lorsque $x > -\ln(n)$, on a $c > 0$ et la question précédente s'applique :

$$\boxed{\mathbb{P}(V_n > x) \leq n^{1-\left(\frac{x}{\ln(n)}+1\right)} = n^{-\frac{x}{\ln(n)}} = \mathrm{e}^{-\frac{x}{\ln(n)}\ln(n)} = \mathrm{e}^{-x}.}$$

9. a) L'explication est la même qu'en 15.(a) : la loi de $\tilde{A}_i$ sachant l'événement $[N = p]$ est une loi $B\left(p, \dfrac{1}{n}\right)$.

Déterminons la loi de $\tilde{A}_i$. Notons d'abord que $\tilde{A}_i$ est à valeurs dans $\mathbb{N}$.
Soit $k \in \mathbb{N}$. On applique la formule des probabilités totales au systéme complet d'événements $\{(N = p)\}_{p\in\mathbb{N}}$:

$$\begin{aligned}
\mathbb{P}\left(\tilde{A}_i = k\right) &= \sum_{p=0}^{+\infty} \underbrace{\mathbb{P}_{(N=p)}\left(\tilde{A}_i = k\right)}_{=0 \text{ si } k>p} \times \mathbb{P}(N = p) \\
&= \sum_{p=k}^{+\infty} \binom{p}{k} \frac{1}{n^k} \left(1 - \frac{1}{n}\right)^{p-k} \mathrm{e}^{-\lambda} \frac{\lambda^p}{p!} \\
&= \sum_{p=k}^{+\infty} \frac{\cancel{p!}}{k!\ (p-k)!} \frac{1}{n^k} \left(1 - \frac{1}{n}\right)^{p-k} \mathrm{e}^{-\lambda} \frac{\lambda^p}{\cancel{p!}}.
\end{aligned}$$

On effectue le changement d'indice $q = p - k$:

$$\begin{aligned}
\mathbb{P}\left(\tilde{A}_i = k\right) &= \sum_{q=0}^{+\infty} \frac{1}{k!\ q!} \frac{1}{n^k} \left(1 - \frac{1}{n}\right)^{q} \mathrm{e}^{-\lambda} \lambda^{q+k} \\
&= \frac{\mathrm{e}^{-\lambda}\ \lambda^k}{k!\ n^k} \sum_{q=0}^{+\infty} \frac{\left(\lambda\left(1 - \frac{1}{n}\right)\right)^q}{q!}.
\end{aligned}$$

On reconnaît la somme d'une série exponentielle :

$$\mathbb{P}\left(\tilde{A}_i = k\right) = \frac{\cancel{\mathrm{e}^{-\lambda}}\ \lambda^k}{k!\ n^k} \times \exp\left(\lambda\left(\cancel{1} - \frac{1}{n}\right)\right) = \frac{\left(\frac{\lambda}{n}\right)^k\ \mathrm{e}^{-\frac{\lambda}{n}}}{k!}.$$

Par conséquent, $\boxed{\text{la variable aléatoire } \tilde{A}_i \text{ suit la loi de Poisson } \mathcal{P}\left(\dfrac{\lambda}{n}\right).}$

b) Sachant $[N = p]$ avec $p = k_1 + \cdots + k_n$, l'événement

$$\left[\left[\tilde{A}_1 = k_1\right] \cap \cdots \cap \left[\tilde{A}_n = k_n\right]\right]$$

correspond à :

- choisir k_1 vignettes numéro 1 parmi p,
- choisir k_2 vignettes numéro 2 parmi les $p - k_1$ restantes,
- etc.

ce qui donne :

$$\begin{aligned}
&\binom{p}{k_1} \times \binom{p-k_1}{k_2} \times \cdots \times \binom{p-k_1-\cdots-k_{n-1}}{k_n} \\
&= \frac{p!}{k_1!\cancel{(p-k_1)!}} \times \frac{\cancel{(p-k_1)!}}{k_2!\cancel{(p-k_1-k_2)!}} \times \cancel{\ddots} \times \cdots \times 1 \\
&= \frac{p!}{k_1!k_2!\ldots k_n!}
\end{aligned}$$

cas favorables, avec n^p cas possibles. D'où :

$$\boxed{\mathbb{P}_{(N=p)}\Big(\left[\tilde{A}_1 = k_1\right] \cap \cdots \cap \left[\tilde{A}_n = k_n\right]\Big) = \frac{p!}{k_1!k_2!\ldots k_n!}\frac{1}{n^p}.}$$

c) Soit $(k_1, \ldots, k_n) \in \mathbb{N}^n$. Notons $p = k_1 + \cdots + k_n$.
Le nombre total de vignettes obtenues correspond au nombre d'achats de paquets de céréales :

$$\left[\left[\tilde{A}_1 = k_1\right] \cap \cdots \cap \left[\tilde{A}_n = k_n\right]\right] \subset (N = p).$$

Donc :

$$\begin{aligned}
\mathbb{P}\Big(\left[\tilde{A}_1 = k_1\right] \cap \cdots \cap \left[\tilde{A}_n = k_n\right]\Big) &= \mathbb{P}\Big(\left[\tilde{A}_1 = k_1\right] \cap \cdots \cap \left[\tilde{A}_n = k_n\right] \cap [N = p]\Big) \\
&= \mathbb{P}_{(N=p)}\Big(\left[\tilde{A}_1 = k_1\right] \cap \cdots \cap \left[\tilde{A}_n = k_n\right]\Big) \times \mathbb{P}(N = p) \\
&= \frac{\cancel{p!}}{k_1!\ldots k_n!}\frac{1}{n^p} \times \frac{\lambda^p}{\cancel{p!}}\mathrm{e}^{-\lambda} \\
&= \frac{\mathrm{e}^{-\lambda}}{k_1!\ldots k_n!}\left(\frac{\lambda}{n}\right)^{k_1+\cdots+k_n} \\
&= \boxed{\prod_{i=1}^{n} \frac{\mathrm{e}^{-\frac{\lambda}{n}}\left(\frac{\lambda}{n}\right)^{k_i}}{k_i!}.}
\end{aligned}$$

d) Pour tout $i \in [\![1, n]\!]$, la variable aléatoire $\tilde{A}_i$ suit la loi de Poisson $P\left(\frac{\lambda}{n}\right)$:

$$\forall k_i \in \mathbb{N}, \quad \mathbb{P}\left(\tilde{A}_i = k_i\right) = \frac{\mathrm{e}^{-\frac{\lambda}{n}}\left(\frac{\lambda}{n}\right)^{k_i}}{k_i!}.$$

Avec la question précédente, on a donc :

$$\forall k_1, \ldots, k_n \in \mathbb{N}, \quad \mathbb{P}\Big(\left[\tilde{A}_1 = k_1\right] \cap \cdots \cap \left[\tilde{A}_n = k_n\right]\Big) = \mathbb{P}\left(\tilde{A}_1 = k_1\right) \times \cdots \times \mathbb{P}\left(\tilde{A}_n = k_n\right)$$

ce qui prouve $\boxed{\text{l'indépendance mutuelle des variables aléatoires } \tilde{A}_1, \ldots, \tilde{A}_n.}$

Remarque. Ainsi :

- *Si on fixe un nombre donné m d'achats de paquets de céréales*, alors les variables aléatoires $A_{1,m}, \ldots, A_{n,m}$, qui comptent le nombre de vignettes de chaque type obtenues, ne sont pas indépendantes, ce qui est normal puisque leur somme vaut m.
- Si en revanche on considère, avec les variables aléatoires $\tilde{A}_1, \ldots, \tilde{A}_n$, les mêmes comptabilisations, *non pas pour un nombre fixé d'achats mais pour un nombre d'achats aléatoire avec une loi bien choisie*, alors ces variables aléatoires sont indépendantes.

On est d'accord, c'est un peu compliqué à comprendre en lisant mais au final pas très surprenant.

10. Avoir une collection compléte signifie que chaque vignette numéro i ($i \in [\![1, n]\!]$) a été obtenue au moins une fois :

$$D_n = \bigcap_{i=1}^{n} \left[\tilde{A}_i \geq 1\right].$$

Avec l'indépendance mutuelle qui vient d'être prouvée :

$$\boxed{\mathbb{P}(D_n) = \prod_{i=1}^{n} \mathbb{P}\left(\tilde{A}_i \geq 1\right) = \prod_{i=1}^{n} \left(1 - \mathbb{P}\left(\tilde{A}_i = 0\right)\right) = \prod_{i=1}^{n} \left(1 - \mathrm{e}^{-\frac{\lambda}{n}}\right) = \left(1 - \mathrm{e}^{-\frac{\lambda}{n}}\right)^n.}$$

11. a) On applique la formule des probabilités totales au système complet d'événements $([N=p])_{p\in\mathbb{N}}$:

$$\boxed{\mathbb{P}(D_n)=\sum_{p=0}^{+\infty}\mathbb{P}_{[N=p]}(D_n)\mathbb{P}(N=p)=\sum_{p=0}^{+\infty}\mathbb{P}(C_n\leq p)\mathbb{P}(N=p).}$$

b) On a :

$$\sum_{p=k_1}^{k_2}\mathbb{P}(N=p)=\mathbb{P}\Big(\lfloor\lambda-a\rfloor\leq N\leq\lfloor\lambda+a\rfloor+1\Big).$$

Tout réel x vérifie :

- $\lfloor x\rfloor\leq x$, donc :
$$\Big(\lfloor\lambda-a\rfloor\leq N\Big)\supset(\lambda-a\leq N),$$
- $x<\lfloor x\rfloor+1$, donc :
$$(N\leq\lambda+a)\subset\Big(N<\lfloor\lambda+a\rfloor+1\Big)\subset\Big(N\leq\lfloor\lambda+a\rfloor+1\Big).$$

On a donc l'inclusion d'événements :

$$(\lambda-a\leq N\leq\lambda+a)\subset\Big(\lfloor\lambda-a\rfloor\leq N\leq\lfloor\lambda+a\rfloor+1\Big).$$

On obtient :

$$\sum_{p=k_1}^{k_2}\mathbb{P}(N=p)\geq\mathbb{P}(\lambda-a\leq N\leq\lambda+a)=1-\mathbb{P}\Big(|N-\lambda|>a\Big).$$

Or N suit la loi de Poisson $\mathcal{P}(\lambda)$ donc admet une espérance et une variance, toutes deux égales à λ. Par l'inégalité de Bienaymé-Tchébychev :

$$\boxed{\sum_{p=k_1}^{k_2}\mathbb{P}(N=p)=1-\mathbb{P}\Big(|N-\mathbb{E}(N)|>a\Big)\geq1-\frac{\mathbb{V}(N)}{a^2}=1-\frac{\lambda}{a^2}.}$$

c) • *Première inégalité.*
On ne conserve que les termes pour $p\in[\![k_1,k_2]\!]$ dans la somme de la question (a) :

$$\mathbb{P}(D_n)=\sum_{p=0}^{+\infty}\mathbb{P}(C_n\leq p)\mathbb{P}(N=p)\geq\sum_{p=k_1}^{k_2}\mathbb{P}(C_n\leq p)\mathbb{P}(N=p)$$

Or :

$$p\geq k_1\implies[C_n\leq k_1]\subset[C_n\leq p]$$

et donc :

$$\mathbb{P}(C_n\leq p)\geq\mathbb{P}(C_n\leq k_1).$$

D'où :

$$\mathbb{P}(D_n)\geq\sum_{p=k_1}^{k_2}\overbrace{\mathbb{P}(C_n\leq k_1)}^{\text{indépendant de }p}\times\mathbb{P}(N=p)=\mathbb{P}(C_n\leq k_1)\sum_{p=k_1}^{k_2}\mathbb{P}(N=p)$$

et donc, avec la question (b):

$$\boxed{\mathbb{P}(D_n)\geq\mathbb{P}(C_n\leq k_1)\left(1-\frac{\lambda}{a^2}\right).}$$

- *Seconde inégalité.*
  On découpe la somme obtenue en question (a) :

$$\mathbb{P}(D_n) = \sum_{p=0}^{k_2} \mathbb{P}(C_n \leq p)\mathbb{P}(N=p) + \sum_{p=k_2+1}^{+\infty} \mathbb{P}(C_n \leq p)\mathbb{P}(N=p).$$

  * Pour le premier morceau, on raisonne comme précédemment en majorant $\mathbb{P}(C_n \leq p)$ par $\mathbb{P}(C_n \leq k_2)$:

$$\sum_{p=0}^{k_2} \mathbb{P}(C_n \leq p)\mathbb{P}(N=p) \leq \mathbb{P}(C_n \leq k_2) \underbrace{\sum_{p=0}^{k_2} \mathbb{P}(N=p)}_{\leq 1} \leq \mathbb{P}(C_n \leq k_2).$$

  * Pour le second morceau, on majore $\mathbb{P}(C_n \leq p)$ par 1 :

$$\sum_{p=k_2+1}^{+\infty} \mathbb{P}(C_n \leq p)\mathbb{P}(N=p) \leq \sum_{p=k_2+1}^{+\infty} \mathbb{P}(N=p) = 1 - \sum_{p=0}^{k_2} \mathbb{P}(N=p) \dots$$

  et on tronque cette dernière somme pour utiliser la question (b) :

$$\dots \leq 1 - \sum_{p=k_1}^{k_2} \mathbb{P}(N=p) \leq \frac{\lambda}{a^2}.$$

  On recolle les morceaux :

$$\boxed{\mathbb{P}(D_n) \leq \mathbb{P}(C_n \leq k_2) + \frac{\lambda}{a^2}.}$$

12. Laissons-nous guider par l'indication de l'énoncé en posant $\lambda = n\ln(n) + c_n n$ et $a = n^{2/3}$, et observons chacune des cinq expressions colorées que nous donne la question précédente :

$$\mathbb{P}(C_n \leq k_1)\left(1 - \frac{\lambda}{a^2}\right) \leq \mathbb{P}(D_n) \leq \mathbb{P}(C_n \leq k_2) + \frac{\lambda}{a^2}.$$

- On minore la partie entière :

$$k_1 = \lfloor \lambda - a \rfloor \geq \lambda - a - 1 = n\ln(n) + c_n n - n^{2/3} - 1$$

  et on obtient :

$$\begin{aligned}\mathbb{P}(C_n \leq k_1) &\geq \mathbb{P}\Big(C_n \leq n\ln(n) + c_n n - n^{2/3} - 1\Big) \\ &= \mathbb{P}\left(V_n \leq \frac{n\ln(n) + c_n n - n^{2/3} - 1}{n} - \ln(n)\right) \\ &= \mathbb{P}\left(V_n \leq c_n - \frac{1}{n^{1/3}} - \frac{1}{n}\right).\end{aligned}$$

- $1 - \dfrac{\lambda}{a^2} = 1 - \dfrac{n\ln(n) + c_n n}{\left(n^{2/3}\right)^2} = 1 - \dfrac{n\ln(n) + c_n n}{n^{4/3}} = 1 - \dfrac{\ln(n) + c_n}{n^{1/3}}$
- La question 18. donne :

$$\mathbb{P}(D_n) = \left(1 - \mathrm{e}^{-\dfrac{n\ln(n) + c_n n}{n}}\right)^n = \left(1 - \mathrm{e}^{-\ln(n) - c_n}\right)^n = \left(1 - \frac{\mathrm{e}^{-c_n}}{n}\right)^n.$$

- On procède comme en rouge, mais cette fois on *majore* la partie entière :

$$k_2 = \lfloor \lambda + a \rfloor + 1 \leq \lambda + a + 1 = n\ln(n) + c_n n + n^{2/3} + 1$$

et on obtient :

$$\begin{aligned}\mathbb{P}(C_n \leq k_2) &\leq \mathbb{P}\Big(C_n \leq n\ln(n) + c_n n + n^{2/3} + 1\Big) \\ &= \mathbb{P}\left(V_n \leq \frac{n\ln(n) + c_n n + n^{2/3} + 1}{n} - \ln(n)\right) \\ &= \mathbb{P}\left(V_n \leq c_n + \frac{1}{n^{1/3}} + \frac{1}{n}\right).\end{aligned}$$

- $\dfrac{\lambda}{a^2} = \dfrac{n\ln(n) + c_n n}{(n^{2/3})^2} = \dfrac{n\ln(n) + c_n n}{n^{4/3}} = \dfrac{\ln(n) + c_n}{n^{1/3}}$

On obtient bien la réponse demandée en recollant tous les morceaux.
