**HEC/ESCP 2, voie ECS - Épreuve de mathématiques du 9 mai 2022**

*Corrigé proposé par Romain Meurant pour l'APHEC. Toute suggestion permettant de l'améliorer est la bienvenue : romain.meurant@lycee-descartes.ma*

## 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 :

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

(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 $\dfrac{n-i+1}{n}$ :

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

## Partie I

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

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

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 :

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

(c) *La convergence de la suite $(S_n)$ 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 $(S_n)$ est croissante car :

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

et elle est majorée (par 2) d'après la question précédente. Ainsi, la suite $(S_n)$ converge.
L'encadrement déterminé pour $S_n$ en question précédente donne : $\dfrac{3}{2} \leq S \leq 2$.

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

(d) Il suffit ici d'utiliser les deux encadrements (celui de $S_n$ et celui de $S$) obtenus en (c) et (d) :

$$\frac{3}{2} - \frac{1}{n+1} \leq S_n \leq 2 - \frac{1}{n} \qquad \text{donc} \qquad \frac{1}{n+1} - \frac{3}{2} \leq -S_n \leq \frac{1}{n} - 2$$

et

$$\frac{3}{2} \leq S \leq 2,$$

en sommant membre à membre les inégalités :

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

(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 `Scilab` suivant :

```
S = 0
for i=1:10^7
    S = S+1/i^2
end
disp(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 :

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

(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

$$\forall n \geq 1, \quad u_n \geq 0.$$

(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, la suite $u$ converge.
Sa limite vérifie le même encadrement large : $\gamma \in [0; 1]$.

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 $E$ :

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

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

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

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

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

*Ce troisième point est particulièrement intéressant puisque $\gamma$ est l'espérance d'une variable aléatoire suivant la loi de Gumbel, étudiée en partie II.*

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 $X_1, \ldots, X_n$ :

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

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

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

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

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

*ce qui est particulièrement intéressant puisque $S$ est la variance d'une variable aléatoire suivant la loi de Gumbel, étudiée en partie II.*

6. (a) La variable aléatoire $C_n$ admettant une variance, l'inégalité de Bienaymé-Tchébychev s'écrit :

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

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

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

Enfin, la suite $(S_n)$ étant croissante, elle est majorée par sa limite. Donc :

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

(b) Soit $c > 1$. On pose $a = c - 1$ $(> 0)$ ; avec la question précédente :

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

Travaillons sur l'événement intervenant ici :

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

D'après la question 3. :

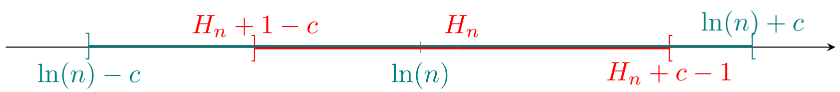
$$0 \leq u_n = H_n - \ln(n) \leq 1$$

donc notamment :

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

d'où :

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




On a donc l'inclusion entre les intervalles :

$$\big]H_n - c + 1; H_n + c - 1\big[ \subset \big]\ln(n) - c; \ln(n) + c\big[$$

On en déduit l'inclusion d'événements :

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

et finalement, par croissance de la probabilité :

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

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

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

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

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

*Il semblerait qu'il y ait une coquille dans l'énoncé (lire « 19.82 » au lieu de « 19.92 »). Néanmoins la réponse de l'énoncé est vraie puisque :*

$$\mathrm{P}\left(\frac{C_n}{n} \in [7.81; 19.92]\right) \geq \mathrm{P}\left(\frac{C_n}{n} \in [7.81; 19.82]\right).$$

## Partie II

7. Commençons par déterminer la fonction de répartition $F_{M_n}$ de $M_n$.
Pour tout $x \in \mathbb{R}$ :

$$(M_n \leq x) = (T_1 \leq x) \cap \cdots \cap (T_n \leq x).$$

Les variables aléatoires $T_1, \ldots, T_n$ étant mutuellement indépendantes :

$$F_{M_n}(x) = \mathrm{P}(M_n \leq x) = \mathrm{P}(T_1 \leq x) \times \cdots \times \mathrm{P}(T_n \leq x)$$

et comme ces variables suivent toutes la même loi $\mathcal{E}(1)$ :

$$F_{M_n}(x) = \big(F_{T_1}(x)\big) = \begin{cases} \left(1-\mathrm{e}^{-x}\right)^n & \text{si } x \geq 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

La fonction de répartition de $M_n$ est clairement de classe $\mathcal{C}^1$ sur $\mathbb{R}^*$ et continue également en 0 (car $\left(1-\mathrm{e}^{-0}\right)^n = 0$), donc continue sur $\mathbb{R}$, ce qui prouve que $M_n$ est une variable aléatoire à densité.
Une densité $f_n$ de $M_n$ est obtenue en dérivant $F_{M_n}$ sur $\mathbb{R}^*$ et en choisissant une valeur arbitraire en 0 (choisissons 0) :

$$\forall x \in \mathbb{R}, \quad f_n(x) = \begin{cases} n\mathrm{e}^{-x}\left(1-\mathrm{e}^{-x}\right)^{n-1} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

8. (a) La fonction $g$ est définie par :

$$g(x) = \begin{cases} (n+1)\mathrm{e}^{-(n+1)x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

Pour $x > 0$ et $t \in ]0; x[$, comme $t > 0$ et $x - t > 0$ :

$$\begin{aligned} f_n(t)g(x-t) &= n\mathrm{e}^{-t}\left(1-\mathrm{e}^{-t}\right)^{n-1} \times (n+1)\mathrm{e}^{-(n+1)(x-t)} \\ &= n\mathrm{e}^{-t}\left(1-\mathrm{e}^{-t}\right)^{n-1}(n+1)\mathrm{e}^{-(n+1)x}\mathrm{e}^{(n+1)t} \\ &= (n+1)\mathrm{e}^{-(n+1)x}n\mathrm{e}^{t}\left(\mathrm{e}^{t}-1\right)^{n-1}. \end{aligned}$$

(b) Par ailleurs, pour $x \leq 0$, ainsi que pour $x \geq 0$ et $t \notin ]0; x[$ :

$$f_n(t)g(x-t) = 0$$

(l'un des deux facteurs est nul).

Pour tout $n \in \mathbb{N}^*$, posons $Y_n = \sum_{i=1}^{n} Z_i$.
On va montrer par récurrence que, pour tout $n \in \mathbb{N}^*$, la variable aléatoire $Y_n$ admet $f_n$ pour densité.

— *Initilisation.* Pour $n = 1$, $Y_1 = Z_1$ suit la loi $\mathcal{E}(1)$ et $f_1$ en est bien une densité :

$$f_1(x) = \begin{cases} 1\mathrm{e}^{-x}\left(1-\mathrm{e}^{-x}\right)^0 = \mathrm{e}^{-x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}.$$

— *Hérédité.* On suppose pour un entier $n \in \mathbb{N}^*$ fixé, que $Y_n$ admet pour densité $f_n$. Déterminons une densité de $Y_{n+1}$.
Comme $Y_{n+1} = Y_n + Z_{n+1}$ et comme les variables aléatoires $Y_n = Z_1 + \cdots + Z_n$ et $Z_{n+1}$ sont indépendantes (d'après le lemme des coalitions, les variables aléatoires $Z_1, \ldots, Z_n, Z_{n+1}$ étant mutuellement indépendantes), on peut utiliser le produit de convolution.
La fonction $h$ définie ci-dessous, si elle est définie et continue sauf peut-être en un nombre fini de points, sera une densité de $Y_{n+1}$ :

$$h(x) = \int_0^x f_n(t)g(x-t)\mathrm{d}t.$$

Cette fonction est nulle pour $x \leq 0$, tout comme $f_{n+1}(x)$.
Pour $x > 0$, la question précédente donne :

$$\begin{aligned} h(x) &= \int_0^x (n+1)\mathrm{e}^{-(n+1)x}n\mathrm{e}^{t}\left(\mathrm{e}^{t}-1\right)^{n-1}\mathrm{d}t \\ &= (n+1)\mathrm{e}^{-(n+1)x}\int_0^x n\mathrm{e}^{t}\left(\mathrm{e}^{t}-1\right)^{n-1}\mathrm{d}t \\ &= (n+1)\mathrm{e}^{-(n+1)x}\left[\left(\mathrm{e}^{t}-1\right)^n\right]_0^x \\ &= (n+1)\mathrm{e}^{-(n+1)x} \times \left(\mathrm{e}^{x}-1\right)^n \\ &= (n+1)\mathrm{e}^{-x} \times \left(\mathrm{e}^{-x}\right)^n\left(\mathrm{e}^{x}-1\right)^n \\ &= (n+1)\mathrm{e}^{-x}\left(1-\mathrm{e}^{-x}\right)^n \;=\; f_{n+1}(x). \end{aligned}$$

Par conséquent, $f_{n+1}$ est une densité de $Y_{n+1}$.

On a finalement prouvé par récurrence que, pour tout entier $n \in \mathbb{N}^*$, la variable aléatoire $Z_1 + \cdots + Z_n$ admet $f_n$ pour densité.

9. Par les règles sur le produit et la composée de fonctions continues, la fonction $f$ est continue sur $\mathbb{R}^*$.
Elle est de plus clairement positive.
Déterminons, si elle converge, son intégrale entre $-\infty$ et $+\infty$. Pour $A > 0$, on reconnaît sous l'intégrale $u'\mathrm{e}^{u}$ avec $u(x) = -\mathrm{e}^{-x}$ :

$$\int_{-A}^{A} f(x)\mathrm{d}x = \int_{-A}^{A} \mathrm{e}^{-x}\mathrm{e}^{-\mathrm{e}^{-x}}\mathrm{d}x = \left[\mathrm{e}^{-\mathrm{e}^{-x}}\right]_{-A}^{A} = \mathrm{e}^{-\mathrm{e}^{-A}} - \mathrm{e}^{-\mathrm{e}^{A}} \underset{A\to+\infty}{\longrightarrow} 1 - 0 = 1.$$

Ainsi, l'intégrale $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x$ converge et vaut 1.
En conclusion, $f$ est une densité de probabilité.

10. (a) Pour tout $x \in \mathbb{R}$ :

$$F_{W_n}(x) = \mathrm{P}(W_n \leq x) = \mathrm{P}\left(\sum_{i=1}^{n} Z_i - \ln(n) \leq x\right) = \mathrm{P}\left(\sum_{i=1}^{n} Z_i \leq x + \ln(n)\right)$$

D'après les questions 7. et 8.(b), la fonction de répartition de la variable aléatoire $\sum_{i=1}^{n} Z_i$ est la même que celle de $M_n$ (la fonction de répartition caractérise la loi) :

$$F_{W_n}(x) = F_{M_n}\big(x + \ln(n)\big)$$

$$= \begin{cases} \left(1-\mathrm{e}^{-x-\ln(n)}\right)^n & \text{si } x+\ln(n)>0 \\ 0 & \text{si } x+\ln(n)\leq 0 \end{cases}$$
$$= \begin{cases} \left(1-\dfrac{\mathrm{e}^{-x}}{n}\right)^n & \text{si } x>-\ln(n) \\ 0 & \text{si } x\leq -\ln(n) \end{cases}.$$

(b) Soit $x \in \mathbb{R}$.
Comme $-\ln(n) \underset{n\to+\infty}{\longrightarrow} -\infty$, on a :

$$\exists N \in \mathbb{N}^*, \quad \forall n \geq N, \quad x > -\ln(n).$$

Ainsi, pour $n \geq N$ :

$$F_{W_n}(x) = \left(1-\frac{\mathrm{e}^{-x}}{n}\right)^n = \exp\left(n\ln\left(1-\frac{\mathrm{e}^{-x}}{n}\right)\right)$$

Comme $-\dfrac{\mathrm{e}^{-x}}{n} \underset{n\to+\infty}{\longrightarrow} 0$, on a l'équivalent :

$$\ln\left(1-\frac{\mathrm{e}^{-x}}{n}\right) \underset{n\to+\infty}{\sim} -\frac{\mathrm{e}^{-x}}{n} \quad \text{donc} \quad n\ln\left(1-\frac{\mathrm{e}^{-x}}{n}\right) \underset{n\to+\infty}{\sim} -\mathrm{e}^{-x}$$

Ce dernier équivalent étant constant et non nul, il s'agit de la limite de l'expression :

$$n\ln\left(1-\frac{\mathrm{e}^{-x}}{n}\right) \underset{n\to+\infty}{\longrightarrow} -\mathrm{e}^{-x}.$$

Ainsi, par composition des limites :

$$\forall x \in \mathbb{R}, \quad F_{W_n}(x) \underset{n\to+\infty}{\longrightarrow} \mathrm{e}^{-\mathrm{e}^{-x}}$$

Or la fonction $G$ définie, pour $x \in \mathbb{R}$, par $G(x) = \mathrm{e}^{-\mathrm{e}^{-x}}$ est la fonction de répartition de la loi de Gumbel (elle est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ de dérivée $f$).

On en conclut que la suite de variables aléatoires $(W_n)$ converge en loi vers la loi de Gumbel.

11. (a)

```
function simulV(n)
    C = 0
    for i=1:n
        C = C + grand(1,1,'geom',(n-i+1)/n)
    end
    y = C/n-log(n)
endfunction
```

(b)

```
n = input('Entrer un entier n: ')
V = []
for k=1:1000
    V = [V, simulV(n)]
end
```

(c) Plus $n$ est grand, plus la hauteur des bâtons de l'histogramme (représentant la loi de $V_n$) est proche de la courbe (représentant la fonction $f$, densité de la loi de Gumbel).
On peut donc conjecturer que la suite de variables aléatoires $(V_n)$ converge en loi vers la loi de Gumbel.

12. Puisque $U$ est à valeurs positives et $\alpha > 0$, la variable aléatoire $V$ prend ses valeurs dans $\mathbb{N}^*$.
Pour tout $k \in \mathbb{N}^*$ :

$$\begin{aligned} \mathrm{P}(V=k) &= \mathrm{P}\left(\lfloor \alpha U\rfloor + 1 = k\right) = \mathrm{P}\left(\lfloor \alpha U\rfloor = k-1\right) = \mathrm{P}\left(k-1 \leq \alpha U < k\right) \\ &= \mathrm{P}\left(\frac{k-1}{\alpha} \leq U < \frac{k}{\alpha}\right) = F_U\left(\frac{k}{\alpha}\right) - F_U\left(\frac{k-1}{\alpha}\right) \end{aligned}$$

où $F_U$ est la fonction de répartition de $U$, variable aléatoire de loi $\mathcal{E}(p)$ :

$$\forall x \geq 0, \quad F_U(x) = 1-\mathrm{e}^{-px}$$

donc :

$$\begin{aligned} \mathrm{P}(V=k) &= 1-\mathrm{e}^{-p\frac{k}{\alpha}} - \left(1-\mathrm{e}^{-p\frac{k-1}{\alpha}}\right) = \mathrm{e}^{-p\frac{k-1}{\alpha}} - \mathrm{e}^{-p\frac{k}{\alpha}} \\ &= \mathrm{e}^{(k-1)\ln(1-p)} - \mathrm{e}^{k\ln(1-p)} \\ &= (1-p)^{k-1} - (1-p)^k \\ &= (1-p)^{k-1}\big(1-(1-p)\big) = p(1-p)^{k-1}. \end{aligned}$$

Ainsi, la variable aléatoire $V$ suit la loi géométrique $\mathcal{G}(p)$.

13. (a) Tout réel $x$ vérifie :

$$x-1 < \lfloor x\rfloor \leq x \quad \text{donc} \quad 0 < \lfloor x\rfloor - x + 1 \leq 1$$

donc la variable aléatoire

$$V - \alpha U = \lfloor \alpha U\rfloor - \alpha U - 1$$

est à valeurs dans $]0,1]$.

Notons, pour cette question uniquement, $X = V - \alpha U$.

Comme $X$ est à valeurs dans $[0,1]$, $X^2$ également.
Les espérance de ces deux variables aléatoires sont donc comprises entre 0 et 1 :

$$0 \leq \mathrm{E}(X) \leq 1 \quad \text{et} \quad 0 \leq \mathrm{E}\left(X^2\right) \leq 1$$

donc :

$$\mathrm{Var}(X) = \mathrm{E}\left(X^2\right) - \mathrm{E}(X) \leq 1 - 0 = 1.$$

(b) On somme membre à membre les deux égalités :

$$\begin{aligned} \mathrm{Var}(X+Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y) \\ \mathrm{Var}(X-Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\mathrm{Cov}(X,Y) \end{aligned}$$

pour obtenir :

$$\mathrm{Var}(X+Y) + \mathrm{Var}(X-Y) = 2\big(\mathrm{Var}(X) + \mathrm{Var}(Y)\big)$$

et comme $\mathrm{Var}(X-Y) \geq 0$ :

$$\mathrm{Var}(X+Y) \leq 2\big(\mathrm{Var}(X) + \mathrm{Var}(Y)\big).$$

(c) On applique la question précédente avec les variables aléatoires

$$X = V - \alpha U \quad \text{et} \quad Y = (\alpha+1)U$$

qui admettent chacune un moment d'ordre 2 :

$$\begin{aligned} \mathrm{Var}(V-U) &\leq 2\Big(\mathrm{Var}(V-\alpha U) + \mathrm{Var}\big((\alpha+1)U\big)\Big) \\ &\leq 2\Big(\underbrace{\mathrm{Var}(V-\alpha U)}_{\substack{\leq 1 \\ \text{d'après (a)}}} + (\alpha+1)^2 \underbrace{\mathrm{Var}(U)}_{\substack{=\frac{1}{p^2} \\ \text{car } U \hookrightarrow \mathcal{E}(p)}}\Big) \\ &\leq 2 + 2 \times \frac{(\alpha+1)^2}{p^2}. \end{aligned}$$

14. (a) On a :

$$V_n' - W_n' = \frac{X_1' + \cdots + X_n'}{n} - \frac{Y_1 + \cdots + Y_n}{n} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i' - Y_i\right).$$

Les variables aléatoires

$$X_1' - Y_1,\ X_2' - Y_2,\ \ldots,\ X_n' - Y_n$$

sont mutuellement indépendantes, car les variables aléatoires $Y_1, \ldots, Y_n$ le sont et, pour tout $i$, $X_i'$ ne dépend pas des $Y_j$ pour $j \neq i$.
Par propriété de la variance :

$$\mathrm{Var}\left(V_n' - W_n'\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left(X_i' - Y_i\right).$$

— Pour $i = 1$, sachant que $X_1' = 1$ et $Y_1 \hookrightarrow \mathcal{E}\left(\underbrace{\frac{n-1+1}{n}}_{=1}\right)$ :

$$\mathrm{Var}\left(X_1' - Y_1\right) = \mathrm{Var}(1 - Y_1) = \mathrm{Var}(Y_1) = \frac{1}{1^2} = 1.$$

— Pour $i \geq 2$, on applique la question 13. à :

$$p = \frac{n-i+1}{n} \in \left]0;1\right[, \quad U = Y_i \hookrightarrow \mathcal{E}(p),$$

$$\alpha = -\frac{p}{\ln(1-p)} = -\frac{\frac{n-i+1}{n}}{\ln\left(1 - \frac{n-i+1}{n}\right)} = -\frac{n-i+1}{n} \times \frac{1}{\ln\left(\frac{i-1}{n}\right)} = \alpha_i,$$

$$V = \lfloor \alpha U \rfloor + 1 = \lfloor \alpha_i Y_i \rfloor + 1 = X_i',$$

et on obtient :

$$\mathrm{Var}\left(X_i' - Y_i\right) \leq 2 + 2\frac{(1-\alpha)^2}{p^2} = 2 + 2\left(\frac{n}{n-i+1}\right)^2 (1-\alpha_i)^2.$$

Par conséquent :

$$\mathrm{Var}\left(V_n' - W_n'\right) \leq \frac{1}{n^2} + \frac{2}{n^2}\sum_{i=2}^{n}\left[1 + \left(\frac{n}{n-i+1}\right)^2 (1-\alpha_i)^2\right].$$

(b) i. La fonction $\Phi$ est clairement continue sur l'intervalle ouvert $]0;1[$.

Calculons les limites de $\Phi$ aux bornes.

— En $0^+$, aucune difficulté :

$$\frac{1-z}{\ln(z)} \underset{z\to 0^+}{\longrightarrow} 0 \quad \left(\text{« } \frac{1}{-\infty} \text{ »}\right) \quad \text{donc} \quad \Phi(z) \underset{z\to 0^+}{\longrightarrow} \frac{1}{(1-0)^2}(1+0)^2 = 1.$$

— En $1^-$, connaître la limite de $\dfrac{1-z}{\ln(z)}$ ne suffit pas...
En effet, cette limite vaut $-1$ car $\ln(z) = \ln(1 + (z-1)) \underset{z\to 1}{\sim} z - 1$ ; on a donc une forme indéterminée :

$$\underbrace{\frac{1}{(1-z)^2}}_{\underset{z\to 1}{\longrightarrow} +\infty} \times \underbrace{\left(1 + \frac{1-z}{\ln(z)}\right)^2}_{\underset{z\to 1}{\longrightarrow} 0}$$

On effectue le changement de variable $h = z - 1$ ($\underset{z\to 1^-}{\longrightarrow} 0^+$) :

$$\Phi(1-h) = \frac{1}{h^2}\left(1 + \frac{h}{\ln(1-h)}\right)^2 = \frac{1}{h^2} \times \left(\frac{\ln(1-h) + h}{\ln(1-h)}\right)^2$$

et on utilise

∗ le développement limité en 0 à l'ordre 2 de $\ln(1-h)$ au numérateur :

$$\ln(1-h) = -h - \frac{h^2}{2} + \underset{h\to 0}{\mathrm{o}}\left(h^2\right) \qquad \text{donc} \qquad \ln(1-h) + h \underset{h\to 0}{\sim} -\frac{h^2}{2};$$

∗ l'équivalent de $\ln(1-h)$ au dénominateur :

$$\ln(1-h) \underset{h\to 0}{\sim} -h.$$

On obtient :

$$\Phi(1-h) \underset{h\to 0}{\sim} \frac{1}{h^2} \times \left(\frac{-\dfrac{h^2}{2}}{-h}\right)^2 = \frac{1}{4}$$

et cette constante non nulle est donc la limite de $\Phi$ en 1.

En conclusion, $\Phi$ peut être prolongée en une fonction continue sur $[0;1]$, en posant :

$$\Phi(0) = 1 \quad \text{et} \quad \Phi(1) = \frac{1}{4}.$$

ii. Soit $i \in [\![2, n]\!]$. Avec $z = \dfrac{i-1}{n} \in ]0;1[$, on a $1 - z = \dfrac{n-i+1}{n}$, donc :

$$\begin{aligned}\left(\frac{n}{n-i+1}\right)^2 (1-\alpha_i)^2 &= \frac{1}{(1-z)^2}\left(1 + \frac{n-i+1}{n} \times \frac{1}{\ln\left(\dfrac{i-1}{n}\right)}\right)^2 \\ &= \frac{1}{(1-z)^2}\left(1 + \frac{1-z}{\ln(z)}\right)^2 = \Phi(z).\end{aligned}$$

On en déduit :

$$\frac{1}{n}\sum_{i=2}^{n}\left(\frac{n}{n-i+1}\right)^2 (1-\alpha_i)^2 = \frac{1}{n}\sum_{i=2}^{n}\Phi\left(\frac{i-1}{n}\right) = \frac{1}{n}\sum_{i=1}^{n-1}\Phi\left(\frac{i}{n}\right).$$

On « ajoute et soustrait » le terme manquant pour reconnaître une somme de Riemann, sachant que $\Phi(1) = \dfrac{1}{4}$ :

$$\frac{1}{n}\sum_{i=2}^{n}\left(\frac{n}{n-i+1}\right)^2 (1-\alpha_i)^2 = \frac{1}{n}\sum_{i=1}^{n}\Phi\left(\frac{i}{n}\right) - \frac{1}{4n}.$$

La fonction $\Phi$ étant continue sur $[0;1]$, son intégrale sur ce segment est convergente et :

$$\frac{1}{n}\sum_{i=1}^{n}\Phi\left(\frac{i}{n}\right) \underset{n\to+\infty}{\longrightarrow} \int_0^1 \Phi(t)\mathrm{d}t.$$

On en conclut que :

$$\frac{1}{n}\sum_{i=2}^{n}\left(\frac{n}{n-i+1}\right)^2 (1-\alpha_i)^2 \underset{n\to+\infty}{\longrightarrow} A = \int_0^1 \Phi(t)\mathrm{d}t$$

et $A > 0$ car la fonction $\Phi$ est continue et strictement positive.

(c) Les questions (a) et (b) donnent :

$$\begin{aligned}0 \le \mathrm{Var}\left(V_n' - W_n'\right) &\le \frac{1}{n^2} + \frac{2}{n^2}\sum_{i=2}^{n}\left[1 + \left(\frac{n}{n-i+1}\right)^2 (1-\alpha_i)^2\right] \\ &\le \underbrace{\frac{1}{n^2} + \frac{2(n-1)}{n^2} + \frac{2}{n}\underbrace{\sum_{i=2}^{n}\left(\frac{n}{n-i+1}\right)^2 (1-\alpha_i)^2}_{\underset{n\to+\infty}{\longrightarrow} A}}_{\underset{n\to+\infty}{\longrightarrow} 0}.\end{aligned}$$

Par le théorème d'encadrement :

$$\mathrm{Var}\left(V_n' - W_n'\right) \underset{n\to+\infty}{\longrightarrow} 0.$$

Par ailleurs, par linéarité de l'espérance :

$$\mathrm{E}\left(V_n' - W_n'\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{E}\left(X_i'\right) - \mathrm{E}\left(Y_i\right)\right)$$

Comme détaillé en 14.(a) :

$$\forall i \in [\![2, n]\!], \quad p = \frac{n-i+1}{n} \implies -\frac{p}{\ln(1-p)} = \alpha_i$$

et donc, d'après la question 13., $X_i' = \lfloor \alpha_i Y_i \rfloor + 1$ suit la loi $\mathcal{G}\left(\dfrac{n-i+1}{n}\right)$.
Ceci est vrai aussi pour $n = 1$ : $X_1' = 1 \hookrightarrow \mathcal{G}(1)$.
On a aussi $Y_i \hookrightarrow \mathcal{E}\left(\dfrac{n-i+1}{n}\right)$.
Ainsi :

$$\mathrm{E}\left(X_i'\right) = \mathrm{E}\left(Y_i\right) \quad \left(= \frac{n}{n-i+1}\right)$$

et la variable aléatoire $V_n' - W_n'$ est donc d'espérance nulle.

On applique alors l'inégalité de Bienaymé-Tchébychev :

$$\forall \varepsilon > 0, \quad 0 \le \mathrm{P}\left(\left|V_n' - W_n'\right| > \varepsilon\right) \le \frac{\mathrm{Var}\left(V_n' - W_n'\right)}{\varepsilon^2}$$

ce qui prouve, avec le théorème d'encadrement, que :

$$\forall \varepsilon > 0, \quad \mathrm{P}\left(|V'_n - W'_n| > \varepsilon\right) \underset{n\to+\infty}{\longrightarrow} 0$$

et on en déduit la convergence en probabilité vers 0 de la suite de variables aléatoires $(V'_n - W'_n)$.

(d) Pour tout $i \in [\![1, n]\!]$, la variable aléatoire $Y_i$ suit la loi $\mathcal{E}\left(\frac{n-i+1}{n}\right)$, donc $\frac{Y_i}{n}$ suit la loi $\mathcal{E}(n-i+1)$ qui est aussi la loi suivie par $Z_{n-i+1}$.

*(Le résultat classique*

$$X \hookrightarrow \mathcal{E}(\lambda) \implies \frac{X}{a} \hookrightarrow \mathcal{E}(a\lambda)$$

*avec $a, \lambda > 0$, se prouve facilement en écrivant les fonctions de répartition.)*

Ainsi, la variable aléatoire

$$W'_n = \sum_{i=1}^{n} \frac{Y_i}{n} - \ln(n)$$

suit la même loi que la variable aléatoire

$$\underbrace{\sum_{i=1}^{n} Z_{n-i+1} - \ln(n) = \sum_{j=1}^{n} Z_j - \ln(n)}_{\text{(changement d'indice } j=n-i+1)} = W_n$$

c'est-à-dire, d'après la question 10., la loi de Gumbel.

On écrit alors :

$$V'_n = (V'_n - W'_n) + W'_n.$$

Comme $(V'_n - W'_n)$ converge en probabilité vers 0 et $(W'_n)$ converge en loi vers la loi de Gumbel, le théorème de Slutsky donne la convergence en loi vers la loi de Gumbel de la suite de variables aléatoires $(V'_n)$.

(e) — D'une part, on a :

$$V_n = \frac{X_1 + \cdots + X_n}{n} - \ln(n) \quad \text{et} \quad \forall i \in [\![1, n]\!], \quad X_i \hookrightarrow \mathcal{G}\left(\frac{n-i+1}{n}\right).$$

— D'autre part, comme détaillé en question c. :

$$V'_n = \frac{X'_1 + \cdots + X'_n}{n} - \ln(n) \quad \text{et} \quad \forall i \in [\![1, n]\!], \quad X'_i \hookrightarrow \mathcal{G}\left(\frac{n-i+1}{n}\right).$$

Par conséquent, $V_n$ suit la même loi que $V'_n$, donc converge en loi, elle aussi, vers la loi de Gumbel (ce qui confirme la conjecture faite en question 11.).

La fonction de répartition de la loi de Gumbel est définie sur $\mathbb{R}$ par $x \mapsto \mathrm{e}^{-\mathrm{e}^{-x}}$.
La convergence en loi de la suite de variables aléatoires

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

vers la loi de Gumbel s'écrit donc :

$$\forall a, b \in \mathbb{R} \text{ avec } a < b, \quad \mathrm{P}\left(a \leq \frac{C_n}{n} - \ln(n) \leq b\right) \underset{n\to+\infty}{\longrightarrow} \mathrm{e}^{-\mathrm{e}^{-b}} - \mathrm{e}^{-\mathrm{e}^{-a}}.$$

Pour $n = 10^6$ (« grande » valeur de $n$), cette limite peut être considérée comme une valeur approchée de cette probabilité.
Avec $b = 3.20$ et $a = -1.17$ :

$$\mathrm{P}\left(-1.17 \leq \frac{C_n}{n} - \ln(n) \leq 3.20\right) \approx \mathrm{e}^{-\mathrm{e}^{-3.20}} - \mathrm{e}^{-\mathrm{e}^{1.17}}$$

soit, avec la valeur approchée de $\ln(n) \approx 13.82$ donnée en fin de partie I :

$$\mathrm{P}\left(12.65 \leq \frac{C_n}{n} \leq 17.02\right) \approx 0.96 - 0.04 = 0.92$$

À niveau de confiance égal (0.92), l'intervalle de confiance obtenu en partie II (avec une convergence en loi) est meilleur que celui obtenu en partie I (avec une inégalité de concentration, moins précise).

*Néanmoins, on aurait obtenu un meilleur intervalle de confiance en partie I si l'énoncé nous avait donné l'approximation $S \approx 1.65$, au lieu d'utiliser la majoration peu fine : $S \leq 2$.*

## Partie III

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

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

(b) On utilise les propriétés suivantes de la covariance :

— linéarité à droite,

— la covariance de deux variables aléatoires, dont l'une est certaine, est nulle,

— la covariance de deux variables aléatoires égales correspond à la variance.

Ainsi :

$$\operatorname{Cov}(A_{1,m},\ m - A_{1,m}) = \underbrace{\operatorname{Cov}(A_{1,m}, m)}_{=0} - \underbrace{\operatorname{Cov}(A_{1,m}, A_{i,m})}_{=\operatorname{Var}(A_{1,m})} = -\frac{m}{n}\left(1-\frac{1}{n}\right)^2.$$

(c) La somme des $A_{i,m}$ pour $i \in [\![1,n]\!]$ est égale au nombre total de vignettes achetées (pas nécessairement différentes) :

$$A_{1,m} + \cdots + A_{n,m} = m \quad \text{donc} \quad m - A_{1,m} = A_{2,m} + \cdots + A_{n,m}.$$

Si les variables aléatoires $A_{1,m}, \ldots, A_{n,m}$ étaient mutuellement indépendantes, alors notamment, par le lemme des coalitions, les deux variables aléatoires

$$A_{1,m} \quad \text{et} \quad A_{2,m} + \cdots + A_{n,m}$$

seraient indépendantes, ce qui n'est pas le cas car leur covariance, calculée en question précédente, est non nulle.
Les variables aléatoires $A_{1,m}, \ldots, A_{n,m}$ ne sont donc pas mutuellement indépendantes.

*La manière de procéder dans cette question est un peu « tordue ».*
*On pouvait très bien s'en sortir avec la définition de l'indépendance mutuelle. En effet, on remarque par exemple que l'événement*

$$(A_{1,m} = 0) \cap \cdots \cap (A_{n,m} = 0)$$

*est impossible (il signifie « n'obtenir aucune des vignettes à collectionner lors de $m$ achats »), alors que les événements $(A_{i,m} = 0)$ $(i \in [\![1,n]\!])$ ne le sont pas :*

$$0 = \mathrm{P}\big((A_{1,m} = 0)\cap\cdots\cap(A_{n,m} = 0)\big) \neq \mathrm{P}(A_{1,m} = 0)\times\cdots\times\mathrm{P}(A_{n,m} = 0) = \left(1-\frac{1}{n}\right)^{mn}.$$

16. (a) On procède par récurrence sur $r \geq 1$.

— *Initialisation.* On a égalité pour $r = 1$.

— *Hérédité.* Supposons, pour un entier $r \geq 1$ fixé, que $\mathrm{P}\left(\bigcup_{i=1}^{r} E_i\right) \leq \sum_{i=1}^{r} \mathrm{P}(E_i)$.

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

(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 :

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

L'inégalité classique (que l'on peut par exemple déduire de la convexité de l'exponentielle)

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

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

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

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

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

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 :

$$\mathrm{P}\big(C_n > cn\ln(n)\big) \leq \mathrm{P}\big(C_n > \lfloor cn\ln(n)\rfloor\big) \leq \underbrace{\mathrm{e}^{-\dfrac{\lfloor cn\ln(n)\rfloor}{n}} \cancel{\leq} \mathrm{e}^{-\dfrac{cn\ln(n)}{n}}}_{\text{(l'inégalité dans l'autre sens est vraie)}}$$

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

$$\mathrm{P}\big(C_n > cn\ln(n)\big) = \underbrace{\mathrm{P}(C_n \geq m) \cancel{\leq} \mathrm{P}(C_n > m)}_{\text{(l'inégalité dans l'autre sens est vraie)}}.$$

Dommage, car ça aurait bien fonctionné ensuite :

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

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).

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

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

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

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

17. (a) *Il faut comprendre dans la question : « la loi conditionnelle de $\tilde{A}_i$ » (et non du couple $\big(\tilde{A}_i, N\big)$).*
L'explication est la même qu'en 15.(a) :

$$\tilde{A}_i|_{(N=p)} \hookrightarrow \mathcal{B}\left(p, \frac{1}{n}\right).$$

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 système complet d'événements $\{(N = p)\}_{p\in\mathbb{N}}$ :

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

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

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

On reconnaît la somme d'une série exponentielle :

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

Par conséquent, la variable aléatoire $\tilde{A}_i$ 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

$$\Big(\big(\tilde{A}_1 = k_1\big) \cap \cdots \cap \big(\tilde{A}_n = k_n\big)\Big)$$

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

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

cas favorables, avec $n^p$ cas possibles.

(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 :

$$\Big(\big(\tilde{A}_1 = k_1\big) \cap \cdots \cap \big(\tilde{A}_n = k_n\big)\Big) \subset (N = p).$$

Donc :

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

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

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

Avec la question précédente, on a donc :

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

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

On remarque que :

— *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 on 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.

18. Avoir une collection complète 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 :

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

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

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

(b) On a :

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

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 :

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

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

$$\sum_{p=k_1}^{k_2} \mathrm{P}(N = p) = 1 - \mathrm{P}\big(|N - \mathrm{E}(N)| > a\big) \geq 1 - \frac{\mathrm{Var}(N)}{a^2} = 1 - \frac{\lambda}{a^2}.$$

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

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

Or :

$$p \geq k_1 \implies (C_n \leq k_1) \subset (C_n \leq p).$$

D'où :

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

et donc, avec la question (b) :

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

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

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

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

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

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

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

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

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

On recolle les morceaux :

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

20. (a) 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 :

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

- 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 :

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

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

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

- 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 :

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

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

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

(b) Soit $x \in \mathbb{R}$.

— Avec la suite $(c_n)$ définie par :

$$\forall n \geq 1, \quad c_n = x + \frac{1}{n^{1/3}} + \frac{1}{n}$$

qui converge vers $x$, on a :

$$\mathrm{P}\left(V_n \leq c_n - \frac{1}{n^{1/3}} - \frac{1}{n}\right)\left(1 - \frac{\ln(n) + c_n}{n^{1/3}}\right) = \mathrm{P}(V_n \leq x)\left(1 - \frac{\ln(n) + c_n}{n^{1/3}}\right) \underset{n \to +\infty}{\sim} F_{V_n}(x)$$

(le facteur orange tend vers 1 par croissances comparées).

— De même, avec la suite $(c_n)$ définie par :

$$\forall n \geq 1, \quad c_n = x - \frac{1}{n^{1/3}} - \frac{1}{n}$$

qui converge elle aussi vers $x$ :

$$\mathrm{P}\left(V_n \leq c_n + \frac{1}{n^{1/3}} + \frac{1}{n}\right) + \frac{\ln(n) + c_n}{n^{1/3}} \underset{n\to+\infty}{\sim} F_{V_n}(x)$$

(le terme violet tend vers 0).

— Pour n'importe quelle suite $(c_n)$ qui converge vers $x$, donc notamment pour les deux suites précédentes, on a :

$$\left(1 - \frac{\mathrm{e}^{-c_n}}{n}\right)^n \underset{n\to+\infty}{\longrightarrow} \mathrm{e}^{-\mathrm{e}^{-x}}$$

(même explication qu'en 10.(b)).

On en déduit par théorème d'encadrement que :

$$\forall x \in \mathbb{R}, \quad F_{V_n}(x) \underset{n\to+\infty}{\longrightarrow} \mathrm{e}^{-\mathrm{e}^{-x}}$$

c'est-à-dire que la suite de variable aléatoires $(V_n)$ converge en loi vers la loi de Gumbel.

*

* *

Un problème intéressant, bien construit, avec des notations claires. Il comporte beaucoup de questions et de manipulations classiques, surtout en partie I ⋆ et début de partie II. Les outils du cours utilisés dans le problème sont assez variés.

⋆ Néanmoins, le début de la partie I est un peu trop proche du cours. Est-il vraiment intéressant, dans une telle épreuve, de redémontrer la convergence de la série de Riemann $\sum_{n\geq 1} \frac{1}{n^2}$ ?

Les parties I et II auraient pu être davantage reliées (une variable aléatoire suivant la loi de Gumbel admet pour espérance $\gamma$ et pour variance $S$).

La partie informatique est un peu légère et la question 2.(e) archaïque. On aurait apprécié par exemple, en toute première question, de nous demander de compléter une fonction `Scilab` qui simule la variable aléatoire $C_i$, en utilisant un vecteur ligne de taille $n$, dont chaque coordonnée numéro $j$ représente le nombre de vignettes numéro $j$ obtenues.

```
function C=simulC(i,n)
    A = zeros(1,n)       // Vignettes obtenues (aucune au départ)
    m = 0                // Nombre d'achats
    while sum(A <> 0) < i     // sum(A <> 0) est le nombre
                              // de coordonnees non nulles de A
        j = grand(1,1,'uin',1,n)
        A(j) = A(j) + 1
        m = m + 1
    end
    C = m
endfunction
```