



PRÉPASIMPLE

prepasimple.fr

HEC — 2022

Corrigé de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Corrigé

Corrigé

Rappel utilisé plusieurs fois. Si deux familles finies $(A_1, \dots, A_n)$ et $(A'_1, \dots, A'_n)$ de variables aléatoires discrètes sont chacune mutuellement indépendantes et si A_k et A'_k ont même loi pour tout k , alors, pour toute fonction ψ de n variables, $\psi(A_1, \dots, A_n)$ et $\psi(A'_1, \dots, A'_n)$ ont même loi. En effet, pour toute valeur v , $P([\psi(A_1, \dots, A_n) = v])$ est la somme, sur les n -uplets $(a_1, \dots, a_n)$ tels que $\psi(a_1, \dots, a_n) = v$, des produits $P([A_1 = a_1]) \cdots P([A_n = a_n])$, et cette expression est la même pour la seconde famille.

Rappel sur les lois géométriques. Pour $p \in]0, 1[$, une variable X de loi $\mathcal{G}(p)$ vérifie $X(\Omega) = \mathbb{N}^*$, $P(X = k) = p(1 - p)^{k-1}$, $E(X) = \frac{1}{p}$ et $\text{Var}(X) = \frac{1-p}{p^2}$. Pour $p = 1$, on convient que $\mathcal{G}(1)$ est la loi certaine égale à 1 ; les formules précédentes restent valables ($E(X) = 1$, $\text{Var}(X) = 0$).

Questions préliminaires

- 1) a) Par définition, $X_i = C_i - C_{i-1}$ pour tout $i \in \llbracket 1, n \rrbracket$, avec $C_0 = 0$. La somme est télescopique :

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

Conclusion

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

- b) Pour $i = 1$, le premier paquet acheté fournit toujours une vignette d'un type nouveau, donc $X_1 = 1$: X_1 suit la loi certaine égale à 1, c'est-à-dire $\mathcal{G}(1) = \mathcal{G}\left(\frac{n-1+1}{n}\right)$.

Soit $i \in \llbracket 2, n \rrbracket$. Une fois obtenues $i - 1$ vignettes de types différents, chaque nouveau paquet contient une vignette dont le type est uniformément réparti parmi les n types, indépendamment des paquets précédents. Le paquet apporte un type nouveau lorsque sa vignette est de l'un des $n - (i - 1) = n - i + 1$ types non encore obtenus, ce qui arrive

avec probabilité $\frac{n-i+1}{n}$. Les achats successifs forment donc une suite d'épreuves de Bernoulli indépendantes de même paramètre $p_i = \frac{n-i+1}{n} \in]0, 1[$, et X_i est le rang du premier succès.

Conclusion

pour tout $i \in \llbracket 1, n \rrbracket$, X_i suit la loi géométrique $\mathcal{G}\left(\frac{n-i+1}{n}\right)$. Dans la suite, on note $p_i = \frac{n-i+1}{n}$.

Partie I

2) a) Soit $k \geq 2$. On a $\frac{1}{k} - \frac{1}{k+1} = \frac{1}{k(k+1)}$ et $\frac{1}{k-1} - \frac{1}{k} = \frac{1}{k(k-1)}$. Comme $0 < k(k-1) < k^2 < k(k+1)$, la décroissance de $t \mapsto \frac{1}{t}$ sur $]0, +\infty[$ donne :

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

Conclusion

pour tout $k \geq 2$, $\frac{1}{k} - \frac{1}{k+1} \leq \frac{1}{k^2} \leq \frac{1}{k-1} - \frac{1}{k}$.

b) Soit $n \geq 2$. On somme les inégalités de a) pour k allant de 2 à n ; les sommes extrêmes sont télescopiques :

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

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

Conclusion

$$\text{pour tout } n \geq 2, \quad \frac{3}{2} - \frac{1}{n+1} \leq S_n \leq 2 - \frac{1}{n}.$$

- c) La suite (S_n) est croissante, car $S_{n+1} - S_n = \frac{1}{(n+1)^2} > 0$, et majorée par 2 d'après b) (et $S_1 = 1 \leq 2$). Par le théorème de la limite monotone, elle converge. En passant à la limite dans l'encadrement de b), sachant que $\frac{1}{n+1} \rightarrow 0$ et $\frac{1}{n} \rightarrow 0$:

Conclusion

$$(S_n) \text{ converge vers un réel } S \text{ tel que } \frac{3}{2} \leq S \leq 2.$$

- d) Soit $n \geq 1$ et $N > n$. Pour $k \in \llbracket n+1, N \rrbracket$, on a $k \geq 2$ et l'encadrement de a) s'applique. En sommant pour k de $n+1$ à N :

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

On fait tendre N vers $+\infty$ (n fixé) : $S_N \rightarrow S$.

Conclusion

$$\text{pour tout } n \geq 1, \quad \frac{1}{n+1} \leq S - S_n \leq \frac{1}{n}.$$

- e) Code adapté en Python.

D'après d), pour tout $n \geq 1$, S appartient à l'intervalle $\left[S_n + \frac{1}{n+1} ; S_n + \frac{1}{n} \right]$, de longueur $\frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)}$.

Méthode 1 — Borne inférieure de l'encadrement. Tout point de l'intervalle précédent approche S à $\frac{1}{n(n+1)}$ près; en particulier $\left| S - \left(S_n + \frac{1}{n+1} \right) \right| \leq \frac{1}{n(n+1)}$. Il suffit donc de calculer S_n jusqu'au premier n tel que $\frac{1}{n(n+1)} \leq 10^{-7}$.

Python

```
n = 1
S = 1 # S contient S_n
while 1/(n*(n+1)) > 1e-7:
    n = n + 1
    S = S + 1/n**2
print(n, S + 1/(n+1))
```

La variable S contient S_n à chaque passage; la boucle s'arrête dès que $\frac{1}{n(n+1)} \leq 10^{-7}$ et le programme affiche $S_n + \frac{1}{n+1}$.

Code exécuté : le programme affiche 3162 1.644934016865801.

Méthode 2 — Milieu de l'encadrement. Le milieu $m_n = S_n + \frac{1}{2} \left(\frac{1}{n} + \frac{1}{n+1} \right)$ de l'intervalle vérifie $|S - m_n| \leq \frac{1}{2n(n+1)}$, ce qui permet de s'arrêter plus tôt.

Python

```
n = 1
S = 1 # S contient S_n
while 1/(2*n*(n+1)) > 1e-7:
    n = n + 1
    S = S + 1/n**2
print(n, S + (1/n + 1/(n+1))/2)
```

Code exécuté : le programme affiche 2236 1.6449340668780281.

Remarque : le choix direct de S_n avec $n = 10^7$ convient aussi d'après d) ($0 \leq S - S_n \leq \frac{1}{n} = 10^{-7}$), mais il demande 10^7 termes. Pour information (ce résultat n'est pas utilisé dans le problème), $S = \frac{\pi^2}{6} \approx 1,6449340668$, ce qui est cohérent avec les deux affichages (écarts d'environ 5×10^{-8} et 3×10^{-11}).

Conclusion

chacun des deux programmes fournit une valeur approchée de S à 10^{-7} près, par exemple $S \approx 1,6449341$.

3) a) Soit $n \geq 1$. On a $\ln\left(\frac{n+1}{n}\right) = \ln(n+1) - \ln(n) = \int_n^{n+1} \frac{dt}{t}$. Pour $t \in [n, n+1]$, $\frac{1}{n+1} \leq \frac{1}{t} \leq \frac{1}{n}$; par croissance de l'intégrale sur un segment de longueur 1 :

Conclusion

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

b) Soit $n \geq 1$. Par télescope,

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

(somme vide, nulle, si $n = 1$). En sommant les inégalités de a) pour k de 1 à $n - 1$:

$$\sum_{k=1}^{n-1} \frac{1}{k+1} \leq \ln(n) \leq \sum_{k=1}^{n-1} \frac{1}{k}, \quad \text{soit} \quad H_n - 1 \leq \ln(n) \leq H_n - \frac{1}{n}.$$

Donc $u_n = H_n - \ln(n) \geq \frac{1}{n} \geq 0$ et $u_n \leq 1$.

Conclusion

pour tout $n \geq 1$, $0 \leq u_n \leq 1$.

- c) Pour $n \geq 1$, $u_{n+1} - u_n = \frac{1}{n+1} - \ln\left(\frac{n+1}{n}\right) \leq 0$ d'après a). La suite (u_n) est décroissante et minorée par 0, donc converge par le théorème de la limite monotone. En passant à la limite dans $0 \leq u_n \leq 1$, sa limite γ vérifie $0 \leq \gamma \leq 1$.

Conclusion

(u_n) converge vers un réel $\gamma \in [0, 1]$.

- 4) D'après 1) b), chaque X_i suit la loi $\mathcal{G}(p_i)$ et admet donc une espérance, égale à $\frac{1}{p_i} = \frac{n}{n-i+1}$. Comme

$$C_n = \sum_{i=1}^n X_i$$

(question 1) a)) est une somme finie de variables admettant une espérance, C_n admet une espérance et, par linéarité :

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

Conclusion

C_n admet une espérance et $E(C_n) = nH_n$.

- 5) Chaque X_i admet une variance, $\text{Var}(X_i) = \frac{1-p_i}{p_i^2}$ (nulle pour $i = 1$). Les X_i étant mutuellement indépendantes, C_n admet une variance égale à la somme des variances. Pour $i \in \llbracket 1, n \rrbracket$:

$$\mathbf{Var}(X_i) = \frac{\frac{i-1}{n}}{\left(\frac{n-i+1}{n}\right)^2} = \frac{n(i-1)}{(n-i+1)^2}.$$

Avec le changement d'indice $k = n - i + 1$ (donc $i - 1 = n - k$) :

$$\mathbf{Var}(C_n) = \sum_{k=1}^n \frac{n(n-k)}{k^2} = n^2 \sum_{k=1}^n \frac{1}{k^2} - n \sum_{k=1}^n \frac{1}{k}.$$

Conclusion

C_n admet une variance et $\mathbf{Var}(C_n) = n^2 S_n - n H_n$.

- 6) a) Soit $a > 0$. On applique l'inégalité de Bienaymé-Tchebychev à C_n , qui admet une variance, avec $\varepsilon = an > 0$ et $E(C_n) = n H_n$:

$$P(|C_n - n H_n| \geq an) \leq \frac{\mathbf{Var}(C_n)}{a^2 n^2} = \frac{n^2 S_n - n H_n}{a^2 n^2} \leq \frac{S_n}{a^2}.$$

Enfin $S_n \leq S$ car (S_n) est croissante de limite S .

Conclusion

pour tout $a > 0$, $P(|C_n - n H_n| \geq an) \leq \frac{S}{a^2}$.

- b) Soit $c > 1$. Comme $\ln(n) = H_n - u_n$, l'inégalité triangulaire donne :

$$\left| \frac{C_n}{n} - \ln(n) \right| = \left| \frac{C_n}{n} - H_n + u_n \right| \leq \left| \frac{C_n}{n} - H_n \right| + u_n \leq \left| \frac{C_n}{n} - H_n \right| + 1,$$

d'après 3) b). Donc, si $\left| \frac{C_n}{n} - \ln(n) \right| \geq c$, alors $\left| \frac{C_n}{n} - H_n \right| \geq c - 1$, c'est-à-dire $|C_n - n H_n| \geq (c - 1)n$. D'où l'inclusion d'événements :

$$\left[\left| \frac{C_n}{n} - \ln(n) \right| \geq c \right] \subset [|C_n - nH_n| \geq (c-1)n].$$

Par croissance de P et a) appliqué avec $a = c - 1 > 0$:

Conclusion

$$\text{pour tout } c > 1, \quad P\left(\left|\frac{C_n}{n} - \ln(n)\right| \geq c\right) \leq \frac{S}{(c-1)^2}.$$

c) Ici $n = 10^6$ et l'approximation donnée assure $13,815 \leq \ln(n) \leq 13,817$. On utilise $S \leq 2$ (question 2) c)).

Méthode 1 — À l'aide de 6) b). Avec $c = 6$: $P\left(\left|\frac{C_n}{n} - \ln(n)\right| \geq 6\right) \leq \frac{S}{25} \leq \frac{2}{25} = 0,08$.

Par passage à l'événement contraire, avec probabilité au moins 0,92, on a

$\ln(n) - 6 < \frac{C_n}{n} < \ln(n) + 6$. Or $\ln(n) - 6 \geq 7,815 \geq 7,81$ et $\ln(n) + 6 \leq 19,817 \leq 19,92$, donc :

$$\left[\left| \frac{C_n}{n} - \ln(n) \right| < 6 \right] \subset \left[\frac{C_n}{n} \in [7,81; 19,92] \right].$$

Méthode 2 — À l'aide de 6) a) et de l'encadrement de H_n . Avec $a = 5$:

$P(|C_n - nH_n| \geq 5n) \leq \frac{S}{25} \leq 0,08$. Sur l'événement contraire, $H_n - 5 < \frac{C_n}{n} < H_n + 5$.

D'après 3) b), $\ln(n) \leq H_n \leq \ln(n) + 1$, d'où $\ln(n) - 5 < \frac{C_n}{n} < \ln(n) + 6$, avec

$\ln(n) - 5 \geq 8,815 \geq 7,81$ et $\ln(n) + 6 \leq 19,92$. On obtient la même inclusion, avec un intervalle même un peu plus étroit.

Dans les deux cas, par croissance de P :

$$P\left(\frac{C_n}{n} \in [7,81; 19,92]\right) \geq 1 - 0,08 = 0,92.$$

Conclusion

$$\text{pour } n = 10^6, \quad P\left(\frac{C_n}{n} \in [7,81; 19,92]\right) \geq 0,92.$$

Partie II

- 7) Soit F la fonction de répartition commune des $T_i : F(x) = 1 - e^{-x}$ si $x \geq 0$ et $F(x) = 0$ si $x < 0$. Pour tout réel x , $[M_n \leq x] = \bigcap_{i=1}^n [T_i \leq x]$, donc, par indépendance mutuelle :

$$F_{M_n}(x) = \prod_{i=1}^n P(T_i \leq x) = F(x)^n = \begin{cases} (1 - e^{-x})^n & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$$

La fonction F_{M_n} est continue sur $\mathbb{R}$ (les deux expressions valent 0 en 0) et de classe C^1 sur $\mathbb{R}^*$. Donc M_n est à densité, et une densité s'obtient en dérivant là où F_{M_n} est dérivable : pour $x > 0$, $F'_{M_n}(x) = n e^{-x} (1 - e^{-x})^{n-1}$; pour $x < 0$, $F'_{M_n}(x) = 0$; on choisit la valeur 0 en 0.

Conclusion

M_n admet pour densité la fonction f_n de l'énoncé.

- 8) a) Comme $Z \sim \mathcal{E}(n+1)$, $g(y) = (n+1)e^{-(n+1)y}$ pour $y > 0$ et $g(y) = 0$ pour $y \leq 0$. Soit $x > 0$ et $t \in]0, x[$; alors $t > 0$ et $x - t > 0$. On écrit $1 - e^{-t} = e^{-t}(e^t - 1)$, d'où :

$$f_n(t) = n e^{-t} e^{-(n-1)t} (e^t - 1)^{n-1} = n e^{-nt} (e^t - 1)^{n-1}.$$

Par conséquent :

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

Conclusion

$$\text{pour } x > 0 \text{ et } t \in]0, x[, \quad f_n(t) g(x-t) = (n+1) e^{-(n+1)x} n e^t (e^t - 1)^{n-1}.$$

b) On raisonne par récurrence sur $n \geq 1$, en notant

$$\Sigma_n = \sum_{i=1}^n Z_i.$$

Initialisation : $\Sigma_1 = Z_1$ suit la loi $\mathcal{E}(1)$, de densité $x \mapsto e^{-x}$ sur $]0, +\infty[$ et nulle ailleurs, qui est f_1 .

Hérédité : supposons que Σ_n admette la densité f_n . On a $\Sigma_{n+1} = \Sigma_n + Z_{n+1}$, où $Z_{n+1} \sim \mathcal{E}(n+1)$ a pour densité g (celle de a)) et, par le lemme des coalitions, Z_{n+1} est indépendante de Σ_n . La densité g est bornée (par $n+1$) ; d'après le théorème de convolution, Σ_{n+1} admet pour densité :

$$h(x) = \int_{-\infty}^{+\infty} f_n(t) g(x-t) dt.$$

Le produit $f_n(t) g(x-t)$ n'est non nul que si $t > 0$ et $x-t > 0$.

- Si $x \leq 0$, aucun t ne vérifie ces deux conditions, donc $h(x) = 0 = f_{n+1}(x)$.
- Si $x > 0$, d'après a) :

$$\begin{aligned} h(x) &= (n+1) e^{-(n+1)x} \int_0^x n e^t (e^t - 1)^{n-1} dt = (n+1) e^{-(n+1)x} [(e^t - 1)^n]_0^x \\ &= (n+1) e^{-(n+1)x} (e^x - 1)^n. \end{aligned}$$

Comme $e^{-nx}(e^x - 1)^n = (1 - e^{-x})^n$, on obtient $h(x) = (n+1)e^{-x}(1 - e^{-x})^n = f_{n+1}(x)$.

Conclusion

pour tout $n \geq 1$,

$$\sum_{i=1}^n Z_i$$

admet pour densité f_n ; elle a donc la même loi que M_n . La démonstration n'utilise que l'indépendance mutuelle de $Z_1, \dots, Z_n$ et le fait que $Z_j \sim \mathcal{E}(j)$ pour $j \in \llbracket 1, n \rrbracket$; elle s'applique donc à toute famille finie de ce type.

- 9) La fonction f est continue et positive sur $\mathbb{R}$. Posons $G(x) = e^{-e^{-x}}$; G est de classe C^1 sur $\mathbb{R}$ et $G'(x) = e^{-x} e^{-e^{-x}} = f(x)$. Pour $A < B$:

$$\int_A^B f(x) dx = G(B) - G(A).$$

Quand $B \rightarrow +\infty$, $e^{-B} \rightarrow 0$ donc $G(B) \rightarrow e^0 = 1$; quand $A \rightarrow -\infty$, $e^{-A} \rightarrow +\infty$ donc $G(A) \rightarrow 0$. L'intégrale $\int_{-\infty}^{+\infty} f$ converge et vaut 1.

Conclusion

f est une densité de probabilité sur $\mathbb{R}$, et la fonction de répartition de la loi de Gumbel est $G : x \mapsto e^{-e^{-x}}$.

10) a) D'après 8) b),

$$\Sigma_n = \sum_{i=1}^n Z_i$$

a la même densité, donc la même fonction de répartition, que M_n ; celle-ci a été calculée en 7). Pour tout réel x :

$$F_{W_n}(x) = P(\Sigma_n \leq x + \ln(n)) = F_{M_n}(x + \ln(n)).$$

Si $x + \ln(n) \leq 0$, c'est-à-dire $x \leq -\ln(n)$, cette quantité est nulle. Sinon, elle vaut

$$(1 - e^{-x-\ln(n)})^n = \left(1 - \frac{e^{-x}}{n}\right)^n.$$

Conclusion

$$F_{W_n}(x) = \left(1 - \frac{e^{-x}}{n}\right)^n \text{ si } x > -\ln(n) \text{ et } F_{W_n}(x) = 0 \text{ sinon.}$$

b) Soit $x \in \mathbb{R}$ fixé. Comme $-\ln(n) \rightarrow -\infty$, on a $x > -\ln(n)$ pour n assez grand, et alors $0 < \frac{e^{-x}}{n} < 1$ et :

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

Comme $\frac{e^{-x}}{n} \rightarrow 0$, l'équivalent usuel $\ln(1+v) \sim v$ quand $v \rightarrow 0$ donne

$n \ln\left(1 - \frac{e^{-x}}{n}\right) \sim n \times \left(-\frac{e^{-x}}{n}\right) = -e^{-x}$, donc cette quantité tend vers $-e^{-x}$. Par continuité de l'exponentielle, $F_{W_n}(x) \rightarrow e^{-e^{-x}} = G(x)$.

Cette convergence a lieu en tout réel x , donc en particulier en tout point de continuité de G .

Conclusion

(W_n) converge en loi vers une variable aléatoire de loi de Gumbel.

11) a) Code adapté en Python.

D'après 1) a) et 1) b),

$$C_n = \sum_{i=1}^n X_i$$

avec $X_i \sim \mathcal{G}\left(\frac{n-i+1}{n}\right)$, les X_i étant mutuellement indépendantes. On simule donc successivement $X_1, \dots, X_n$ par des appels indépendants à `np.random.geometric`, on les additionne, puis on renvoie $\frac{C_n}{n} - \ln(n)$.

Python

```
import numpy as np

def simulV(n):
    C = 0
    for i in range(1, n + 1):
        C = C + np.random.geometric((n - i + 1) / n)
    return C / n - np.log(n)
```

La boucle parcourt $i = 1, \dots, n$ (la borne supérieure de range est exclue, d'où $n + 1$); pour $i = 1$, le paramètre vaut 1 et `np.random.geometric(1.0)` renvoie 1, conformément à $X_1 = 1$. La variable C contient à la fin une simulation de C_n .

Code exécuté : par exemple, pour $n = 5$, la plus petite des 1000 valeurs simulées (voir b)) est $1 - \ln(5) \approx -0,609$, valeur minimale théorique de V_5 puisque $C_5 \geq 5$.

Conclusion

la fonction `simulV` ci-dessus renvoie une simulation de V_n .

b) Code adapté en Python.

Python

```
n = int(input("n = "))
V = np.zeros(1000)
for k in range(1000):
    V[k] = simulV(n)
```

L'utilisateur saisit n ; le tableau V est initialisé avec 1000 zéros, puis chaque coefficient $V[k]$, pour k de 0 à 999, reçoit une simulation de V_n . Les appels successifs à `simulV` utilisent des tirages aléatoires distincts, de sorte que les 1000 simulations sont indépendantes.

Code exécuté (avec n fourni au programme) : pour $n = 5, 10$ et 50 , la moyenne des valeurs obtenues est respectivement d'environ 0,62, 0,64 et 0,58, cohérente avec $E(V_n) = H_n - \ln(n) = u_n$, qui vaut environ 0,674, 0,626 et 0,587.

Conclusion

ce programme construit le tableau V de 1000 simulations indépendantes de V_n .

c) Code adapté en Python.

Le code complet (avec les parties a) et b)) a été exécuté pour $n = 5, 10$ et 50 : pour $n = 10$ et $n = 50$, il produit des figures de même allure que celles de l'énoncé ; pour $n = 5$, l'histogramme obtenu est plus irrégulier que celui de l'énoncé, car V_5 prend ses valeurs dans $\left\{ \frac{k}{5} - \ln(5), k \geq 5 \right\}$, ensemble de pas $\frac{1}{5}$, et les 20 classes de même largeur choisies par `plt.hist` ne sont pas alignées sur ces valeurs (certaines classes en contiennent deux, d'autres une seule).

Le code trace l'histogramme normalisé (aire totale 1) des 1000 simulations de V_n et, sur la même figure, la densité f de la loi de Gumbel. Un histogramme normalisé d'un grand nombre de simulations indépendantes approche la densité de la loi simulée lorsqu'elle

existe ; plus précisément, l'aire d'une barre est la fréquence des simulations tombant dans la classe correspondante.

Observations :

- pour $n = 5$, l'histogramme diffère nettement de la courbe : il commence vers $-0,6$ (ce qui correspond à la valeur minimale $1 - \ln(5) \approx -0,61$ de V_5) et ses barres centrales atteignent environ $0,53$, bien au-dessus du maximum $f(0) = e^{-1} \approx 0,37$ de la courbe ;
- pour $n = 10$, l'histogramme commence vers $-1,3$ (valeur minimale $1 - \ln(10) \approx -1,30$) et suit déjà mieux la courbe, avec encore des écarts (barre maximale vers $0,45$) ;
- pour $n = 50$, l'histogramme épouse presque la courbe.

L'ajustement entre l'histogramme des simulations de V_n et la densité de Gumbel s'améliore donc lorsque n augmente.

Conclusion

on peut conjecturer que la suite $(V_n)_{n \geq 1}$ converge en loi vers une variable aléatoire de loi de Gumbel. Il s'agit d'une conjecture issue d'une observation expérimentale, qui sera démontrée en 14) d) puis, par une autre voie, en 20) b).

12) Comme $p \in]0, 1[$, $\ln(1 - p) < 0$, donc $\alpha > 0$. La variable U prend ses valeurs dans $\mathbb{R}_+$ (presque sûrement), donc $\lfloor \alpha U \rfloor \in \mathbb{N}$ et V prend ses valeurs dans $\mathbb{N}^*$.

Soit $k \in \mathbb{N}^*$. Par définition de la partie entière, et comme $\alpha > 0$:

$$[V = k] = [k - 1 \leq \alpha U < k] = \left[\frac{k-1}{\alpha} \leq U < \frac{k}{\alpha} \right].$$

Pour $u \geq 0$, $P(U \geq u) = e^{-pu}$ et U est à densité. Or $\frac{p}{\alpha} = -\ln(1 - p)$, d'où

$e^{-\frac{pj}{\alpha}} = e^{j \ln(1-p)} = (1-p)^j$ pour tout $j \in \mathbb{N}$. Ainsi :

$$P(V = k) = e^{-\frac{p(k-1)}{\alpha}} - e^{-\frac{pk}{\alpha}} = (1-p)^{k-1} - (1-p)^k = p(1-p)^{k-1}.$$

Conclusion

$V = \lfloor \alpha U \rfloor + 1$ suit la loi géométrique $\mathcal{G}(p)$.

13) a) Par définition de la partie entière, $\lfloor \alpha U \rfloor \leq \alpha U < \lfloor \alpha U \rfloor + 1$, donc :

$$0 < V - \alpha U = \lfloor \alpha U \rfloor + 1 - \alpha U \leq 1.$$

La variable $Y = V - \alpha U$ est donc à valeurs dans $[0, 1]$. Elle est bornée, donc (par domination) elle admet un moment d'ordre 2 et $0 \leq Y^2 \leq 1$. Par la formule de Koenig-Huygens et la croissance de l'espérance :

$$\mathbf{Var}(Y) = E(Y^2) - E(Y)^2 \leq E(Y^2) \leq 1.$$

Conclusion

$V - \alpha U$ est à valeurs dans $[0, 1]$ et $\mathbf{Var}(V - \alpha U) \leq 1$.

b) Existence de la covariance. Pour tous réels x et y , $(|x| - |y|)^2 \geq 0$ donne

$|xy| \leq \frac{x^2 + y^2}{2}$. Donc $|XY| \leq \frac{X^2 + Y^2}{2}$, et le majorant admet une espérance puisque X et Y admettent un moment d'ordre 2. Par domination, XY admet une espérance; comme X et Y admettent aussi une espérance,

$\mathbf{Cov}(X, Y) = E(XY) - E(X)E(Y)$ existe. De plus $X + Y$ et $X - Y$ admettent un moment d'ordre 2.

Méthode 1 — Somme des deux variances. Par bilinéarité de la covariance :

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

En additionnant : $\mathbf{Var}(X + Y) + \mathbf{Var}(X - Y) = 2 (\mathbf{Var}(X) + \mathbf{Var}(Y))$. Comme $\mathbf{Var}(X - Y) \geq 0$, on obtient l'inégalité voulue.

Méthode 2 — Inégalité de Cauchy-Schwarz. L'inégalité de Cauchy-Schwarz pour la covariance donne $|\mathbf{Cov}(X, Y)| \leq \sigma(X)\sigma(Y)$, et $2\sigma(X)\sigma(Y) \leq \sigma(X)^2 + \sigma(Y)^2$ (car $(\sigma(X) - \sigma(Y))^2 \geq 0$). Donc :

$$\mathbf{Var}(X + Y) = \mathbf{Var}(X) + \mathbf{Var}(Y) + 2 \mathbf{Cov}(X, Y) \leq 2 (\mathbf{Var}(X) + \mathbf{Var}(Y)).$$

Conclusion

$\mathbf{Cov}(X, Y)$ existe et $\mathbf{Var}(X + Y) \leq 2 (\mathbf{Var}(X) + \mathbf{Var}(Y))$.

- c) On écrit $V - U = (V - \alpha U) + (\alpha - 1)U$. La variable $V - \alpha U$ admet un moment d'ordre 2 d'après a), et $(\alpha - 1)U$ aussi, puisque $U \sim \mathcal{E}(p)$ admet une variance égale à $\frac{1}{p^2}$. D'après b), puis a) :

$$\mathbf{Var}(V - U) \leq 2 (\mathbf{Var}(V - \alpha U) + (\alpha - 1)^2 \mathbf{Var}(U)) \leq 2 \left(1 + \frac{(1 - \alpha)^2}{p^2} \right).$$

Conclusion

$$\text{Var}(V - U) \leq 2 + 2 \frac{(1 - \alpha)^2}{p^2}.$$

14) Préliminaires. Pour $i \in \llbracket 1, n \rrbracket$, on note encore $p_i = \frac{n - i + 1}{n}$, de sorte que $Y_i \sim \mathcal{E}(p_i)$.

- Pour $i \in \llbracket 2, n \rrbracket$, $p_i \in]0, 1[$ et $1 - p_i = \frac{i - 1}{n}$, donc $\alpha_i = -\frac{p_i}{\ln(1 - p_i)}$ est exactement le réel α de la question 12) associé à $p = p_i$ et $U = Y_i$. D'après 12), $X'_i \sim \mathcal{G}(p_i)$, c'est-à-dire que X'_i a la même loi que X_i .

- Pour $i = 1$, $X'_1 = 1$ a la même loi que X_1 .

- Chaque X'_i est une fonction de Y_i ; les Y_i étant mutuellement indépendantes, les X'_i le sont aussi (lemme des coalitions).

D'après le rappel du début, appliqué à $\psi(a_1, \dots, a_n) = \frac{a_1 + \dots + a_n}{n} - \ln(n)$, V'_n a la même loi que V_n .

De même, pour $j \in \llbracket 1, n \rrbracket$, $Y_{n-j+1} \sim \mathcal{E}\left(\frac{j}{n}\right)$, donc $\frac{Y_{n-j+1}}{n} \sim \mathcal{E}(j)$ (transformation

affine d'une loi exponentielle : $P\left(\frac{Y_{n-j+1}}{n} > y\right) = e^{-\frac{j}{n}ny} = e^{-jy}$ pour $y \geq 0$). Les

variables $\frac{Y_n}{n}, \frac{Y_{n-1}}{n}, \dots, \frac{Y_1}{n}$ sont mutuellement indépendantes et suivent respectivement les lois $\mathcal{E}(1), \dots, \mathcal{E}(n)$. D'après 8) b) (dont la démonstration s'applique à une telle famille finie), leur somme $\frac{Y_1 + \dots + Y_n}{n}$ admet la densité f_n , comme

$$\sum_{i=1}^n Z_i.$$

Donc W'_n a la même loi que W_n et, d'après 10) b), $(W'_n)_{n \geq 2}$ converge en loi vers une variable aléatoire de loi de Gumbel.

a) On a

$$V'_n - W'_n = \frac{1}{n} \sum_{i=1}^n (X'_i - Y_i).$$

Les variables $X'_i - Y_i$, $i \in \llbracket 1, n \rrbracket$, sont des fonctions respectives de $Y_1, \dots, Y_n$, donc mutuellement indépendantes (lemme des coalitions) ; leurs covariances deux à deux sont nulles. Chacune admet une variance : $X'_1 - Y_1 = 1 - Y_1$, et pour $i \geq 2$, $X'_i - Y_i = (X'_i - \alpha_i Y_i) + (\alpha_i - 1)Y_i$ (voir 13) c)). La variance d'une somme de variables deux à deux de covariance nulle étant la somme des variances :

$$\mathbf{Var}(V'_n - W'_n) = \frac{1}{n^2} \sum_{i=1}^n \mathbf{Var}(X'_i - Y_i).$$

- Pour $i = 1$: $\mathbf{Var}(1 - Y_1) = \mathbf{Var}(Y_1) = \frac{1}{p_1^2} = 1$, car $p_1 = 1$.
- Pour $i \geq 2$: d'après 13) c) avec $p = p_i$, $U = Y_i$, $\alpha = \alpha_i$ et $V = X'_i$:

$$\mathbf{Var}(X'_i - Y_i) \leq 2 + 2 \frac{(1 - \alpha_i)^2}{p_i^2} = 2 \left[\left(\frac{n}{n - i + 1} \right)^2 (1 - \alpha_i)^2 + 1 \right].$$

En sommant :

Conclusion

$$\mathbf{Var}(V'_n - W'_n) \leq \frac{1}{n^2} + \frac{2}{n^2} \sum_{i=2}^n \left[\left(\frac{n}{n - i + 1} \right)^2 (1 - \alpha_i)^2 + 1 \right].$$

- b) (i) Continuité sur $]0, 1[$. Sur $]0, 1[$, $\ln(z) \neq 0$ et $1 - z \neq 0$: ϕ est continue comme composée, produit et quotient de fonctions continues.

Limite en 0. Quand $z \rightarrow 0^+$, $\ln(z) \rightarrow -\infty$ et $1 - z \rightarrow 1$, donc $\frac{1 - z}{\ln(z)} \rightarrow 0$ et

$$\phi(z) \rightarrow \frac{1}{1}(1 + 0)^2 = 1.$$

Limite en 1. Pour $z \in]0, 1[$, on réduit au même dénominateur :

$$\phi(z) = \left(\frac{\ln(z) + 1 - z}{(1 - z)\ln(z)} \right)^2.$$

On pose $h = 1 - z$, qui tend vers 0^+ quand $z \rightarrow 1^-$. La fonction $h \mapsto \ln(1 - h)$ est de classe C^2 au voisinage de 0, de dérivées $-\frac{1}{1 - h}$ et $-\frac{1}{(1 - h)^2}$, valant -1 et -1 en 0. La formule de Taylor-Young à l'ordre 2 donne :

$$\ln(1 - h) = -h - \frac{h^2}{2} + o(h^2).$$

D'où $\ln(z) + 1 - z = \ln(1 - h) + h = -\frac{h^2}{2} + o(h^2)$ et $(1 - z)\ln(z) = h\ln(1 - h) = -h^2 + o(h^2)$. Donc :

$$\frac{\ln(z) + 1 - z}{(1 - z)\ln(z)} = \frac{-\frac{1}{2} + o(1)}{-1 + o(1)} \rightarrow \frac{1}{2}, \quad \text{puis} \quad \phi(z) \rightarrow \frac{1}{4}.$$

En posant $\phi(0) = 1$ et $\phi(1) = \frac{1}{4}$, on obtient une fonction continue sur $[0, 1]$.

Conclusion

ϕ se prolonge en une fonction continue sur $[0, 1]$, avec $\phi(0) = 1$ et $\phi(1) = \frac{1}{4}$.

(ii) Soit $i \in \llbracket 2, n \rrbracket$ et $z_i = \frac{i - 1}{n} \in]0, 1[$. Alors $1 - z_i = \frac{n - i + 1}{n}$ et $\alpha_i = -\frac{1 - z_i}{\ln(z_i)}$,
donc $1 - \alpha_i = 1 + \frac{1 - z_i}{\ln(z_i)}$ et :

$$\left(\frac{n}{n-i+1}\right)^2 (1-\alpha_i)^2 = \frac{1}{(1-z_i)^2} \left(1 + \frac{1-z_i}{\ln(z_i)}\right)^2 = \phi\left(\frac{i-1}{n}\right).$$

Avec $k = i - 1$:

$$\frac{1}{n} \sum_{i=2}^n \left(\frac{n}{n-i+1}\right)^2 (1-\alpha_i)^2 = \frac{1}{n} \sum_{k=1}^{n-1} \phi\left(\frac{k}{n}\right) = \frac{1}{n} \sum_{k=0}^{n-1} \phi\left(\frac{k}{n}\right) - \frac{\phi(0)}{n}.$$

Comme ϕ (prolongée) est continue sur $[0, 1]$, la somme de Riemann

$$\frac{1}{n} \sum_{k=0}^{n-1} \phi\left(\frac{k}{n}\right)$$

tend vers $\int_0^1 \phi(z) dz$, et $\frac{\phi(0)}{n} \rightarrow 0$. La limite cherchée existe et vaut $A = \int_0^1 \phi(z) dz$.

Positivité stricte : ϕ est positive et continue sur $[0, 1]$, avec $\phi(0) = 1$. Par continuité en 0, il existe $\delta \in]0, 1]$ tel que $\phi(z) \geq \frac{1}{2}$ pour $z \in [0, \delta]$. Donc $A \geq \int_0^\delta \phi \geq \frac{\delta}{2} > 0$.

Conclusion

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=2}^n \left(\frac{n}{n-i+1}\right)^2 (1-\alpha_i)^2 = A$$

avec $A = \int_0^1 \phi(z) dz > 0$. Une valeur approchée numérique, non nécessaire ici, est $A \approx 0,338$.

c) Espérance. Pour tout i , $X'_i \sim \mathcal{G}(p_i)$ et $Y_i \sim \mathcal{E}(p_i)$ ont la même espérance $\frac{1}{p_i}$. Donc

$$E(V'_n - W'_n) = \frac{1}{n} \sum_{i=1}^n (E(X'_i) - E(Y_i)) = 0.$$

Variance. On note

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

La majoration de a) s'écrit :

$$\mathbf{Var}(V'_n - W'_n) \leq \frac{1}{n^2} + \frac{2}{n} B_n + \frac{2(n-1)}{n^2}.$$

D'après b) (ii), $B_n \rightarrow A$, donc $\frac{2}{n} B_n \rightarrow 0$; les deux autres termes tendent vers 0. Ainsi $\mathbf{Var}(V'_n - W'_n) \rightarrow 0$.

Soit $\varepsilon > 0$. Par l'inégalité de Bienaymé-Tchebychev :

$$P(|V'_n - W'_n| \geq \varepsilon) \leq \frac{\mathbf{Var}(V'_n - W'_n)}{\varepsilon^2} \rightarrow 0.$$

Conclusion

$(V'_n - W'_n)_{n \geq 2}$ converge en probabilité vers 0.

d) On note $D'_n = V'_n - W'_n$, de sorte que $V'_n = W'_n + D'_n$, et G la fonction de répartition de la loi de Gumbel (question 9)), continue sur $\mathbb{R}$. On démontre directement la convergence des fonctions de répartition, par encadrement.

Encadrement. Soit $x \in \mathbb{R}$ et $\varepsilon > 0$.

- Si $W'_n \leq x - \varepsilon$ et $|D'_n| \leq \varepsilon$, alors $V'_n \leq x$. Donc $[W'_n \leq x - \varepsilon] \subset [V'_n \leq x] \cup [|D'_n| > \varepsilon]$,
et $P(W'_n \leq x - \varepsilon) \leq P(V'_n \leq x) + P(|D'_n| > \varepsilon)$.
- Si $V'_n \leq x$ et $|D'_n| \leq \varepsilon$, alors $W'_n = V'_n - D'_n \leq x + \varepsilon$. Donc $[V'_n \leq x] \subset [W'_n \leq x + \varepsilon] \cup [|D'_n| > \varepsilon]$, et $P(V'_n \leq x) \leq P(W'_n \leq x + \varepsilon) + P(|D'_n| > \varepsilon)$.

D'où, pour tout $n \geq 2$:

$$F_{W'_n}(x - \varepsilon) - P(|D'_n| > \varepsilon) \leq F_{V'_n}(x) \leq F_{W'_n}(x + \varepsilon) + P(|D'_n| > \varepsilon).$$

Passage à la limite. Soit $\eta > 0$. Par continuité de G en x , on choisit $\varepsilon > 0$ tel que $G(x + \varepsilon) \leq G(x) + \eta$ et $G(x - \varepsilon) \geq G(x) - \eta$. D'après les préliminaires, $F_{W'_n}(x \pm \varepsilon) \rightarrow G(x \pm \varepsilon)$; d'après c), $P(|D'_n| > \varepsilon) \rightarrow 0$. Il existe donc un rang n_0 tel que pour tout $n \geq n_0$:

$$F_{W'_n}(x + \varepsilon) \leq G(x + \varepsilon) + \eta, \quad F_{W'_n}(x - \varepsilon) \geq G(x - \varepsilon) - \eta, \quad P(|D'_n| > \varepsilon) \leq \eta.$$

Pour $n \geq n_0$, l'encadrement donne alors $G(x) - 3\eta \leq F_{V'_n}(x) \leq G(x) + 3\eta$. Ceci étant vrai pour tout $\eta > 0$, $F_{V'_n}(x) \rightarrow G(x)$.

Conclusion

pour tout réel x , $F_{V'_n}(x) \rightarrow G(x)$, donc $(V'_n)_{n \geq 2}$ converge en loi vers une variable aléatoire de loi de Gumbel. Comme V_n a la même loi que V'_n (préliminaires), $F_{V'_n} = F_{V_n}$ et (V_n) converge aussi en loi vers la loi de Gumbel, ce qui démontre la conjecture de 11) c).

e) Pour $n = 10^6$, $\ln(n) \approx 13,816$, donc :

$$\left[\frac{C_n}{n} \in [12,65; 17,02] \right] = [V_n \in [12,65 - \ln(n); 17,02 - \ln(n)]] \quad \text{avec} \quad 12,65 - \ln(n) \approx -1,17, \quad 17,02 - \ln(n) \approx 3,20.$$

D'après d), pour n grand, $F_{V_n}(x) \approx G(x)$ en tout réel x . On a

$$P(a \leq V_n \leq b) = F_{V_n}(b) - F_{V_n}(a) + P(V_n = a), \text{ et pour tout } \eta > 0,$$

$P(V_n = a) \leq F_{V_n}(a) - F_{V_n}(a - \eta)$, quantité proche de $G(a) - G(a - \eta)$, aussi petite que l'on veut; on néglige donc ce terme. En prenant $n = 10^6$ comme « grand » :

$$P\left(\frac{C_n}{n} \in [12,65; 17,02]\right) \approx G(3,20) - G(-1,17) \approx 0,96 - 0,04 = 0,92.$$

Comparaison et commentaire.

- La partie I donne, pour $n = 10^6$, une minoration certaine : $\frac{C_n}{n} \in [7,81 ; 19,92]$ avec probabilité au moins 0,92. Ce résultat est exact et non asymptotique, mais l'intervalle a une longueur d'environ 12,1.
- L'approximation de Gumbel fournit, pour la même probabilité (environ 0,92), un intervalle de longueur environ 4,37, près de trois fois plus court. En revanche, ce n'est qu'une approximation asymptotique : la question 14) d) ne donne aucun contrôle de l'erreur commise pour $n = 10^6$.
- L'intervalle obtenu par Gumbel n'est pas centré en $\ln(n)$: il reflète la dissymétrie de la loi de Gumbel, tandis que l'inégalité de Bienaymé-Tchebychev, qui n'utilise que l'espérance et la variance, conduit à un intervalle symétrique et à une estimation pessimiste.

Conclusion

$P\left(\frac{C_n}{n} \in [12,65 ; 17,02]\right) \approx 0,92$; l'approximation par la loi de Gumbel est beaucoup plus précise que la majoration de Bienaymé-Tchebychev, mais elle n'est qu'approchée, alors que le résultat de la partie I est une garantie.

Partie III

- 15) a) Soit $i \in \llbracket 1, n \rrbracket$. Les m premiers paquets contiennent chacun une vignette dont le numéro est uniformément réparti sur $\llbracket 1, n \rrbracket$, indépendamment des autres paquets. L'événement « la vignette du paquet est numérotée i » a donc la probabilité $\frac{1}{n}$ à chaque achat, et ces m épreuves de Bernoulli sont indépendantes. $A_{i,m}$ compte le nombre de succès.

Conclusion

$A_{i,m}$ suit la loi binomiale $\mathcal{B}\left(m, \frac{1}{n}\right)$.

- b) Par bilinéarité de la covariance, et comme la covariance d'une variable avec une constante est nulle :

$$\mathbf{Cov}(A_{1,m}, m - A_{1,m}) = -\mathbf{Cov}(A_{1,m}, A_{1,m}) = -\mathbf{Var}(A_{1,m}) = -m \frac{1}{n} \left(1 - \frac{1}{n}\right).$$

Conclusion

$$\mathbf{Cov}(A_{1,m}, m - A_{1,m}) = -\frac{m(n-1)}{n^2}.$$

- c) Chacun des m premiers paquets contient exactement une vignette, donc

$$\sum_{i=1}^n A_{i,m} = m,$$

soit

$$m - A_{1,m} = \sum_{i=2}^n A_{i,m}.$$

Supposons par l'absurde que $A_{1,m}, \dots, A_{n,m}$ soient mutuellement indépendantes. Par le lemme des coalitions, $A_{1,m}$ et

$$\sum_{i=2}^n A_{i,m} = m - A_{1,m}$$

sont indépendantes ; leur covariance est donc nulle. Or, d'après b), elle vaut

$$-\frac{m(n-1)}{n^2} \neq 0 \text{ puisque } n \geq 2 \text{ et } m \geq 1. \text{ Contradiction.}$$

Conclusion

les variables $A_{1,m}, \dots, A_{n,m}$ ne sont pas mutuellement indépendantes.

- 16) a) On raisonne par récurrence sur $r \geq 1$. Pour $r = 1$, il y a égalité. Si l'inégalité est vraie au rang r , alors, avec $A = \bigcup_{i=1}^r E_i$ et $B = E_{r+1}$:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B) \leq \sum_{i=1}^r P(E_i) + P(E_{r+1}).$$

Conclusion

pour tout $r \geq 1$,

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

- b) Soit $m \geq 1$. L'événement $[C_n > m]$ signifie que la collection n'est pas complète à l'issue des m premiers achats, c'est-à-dire qu'au moins un numéro i n'a pas été obtenu :

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

D'après 15) a), $P(A_{i,m} = 0) = \left(1 - \frac{1}{n}\right)^m$. Par a) :

$$P(C_n > m) \leq \sum_{i=1}^n \left(1 - \frac{1}{n}\right)^m = n \left(1 - \frac{1}{n}\right)^m.$$

Enfin, par convexité de l'exponentielle, $1 + u \leq e^u$ pour tout réel u ; avec $u = -\frac{1}{n}$, on a

$$0 \leq 1 - \frac{1}{n} \leq e^{-\frac{1}{n}}, \text{ et la croissance de } t \mapsto t^m \text{ sur } \mathbb{R}_+ \text{ donne } \left(1 - \frac{1}{n}\right)^m \leq e^{-\frac{m}{n}}.$$

Conclusion

$$\text{pour tout } m \geq 1, \quad P(C_n > m) \leq n \left(1 - \frac{1}{n}\right)^m \leq n e^{-\frac{m}{n}}.$$

- c) Rectification de l'énoncé : le sujet original demandait $P(C_n > c n \ln(n)) \leq n^{1-c}$, majoration fautive en général ; le sujet adapté demande

$P(C_n > c n \ln(n)) \leq \frac{1}{e^n} n^{1-c}$. Contre-exemple à la version originale : pour $n = 2$, on a toujours $C_2 \geq 2$; avec $c = \frac{1,9}{2 \ln(2)} \approx 1,37$, on a $c n \ln(n) = 1,9$, donc $P(C_2 > 1,9) = 1$, alors que $n^{1-c} = 2^{1-c} \approx 0,77$. La difficulté vient de ce que $c n \ln(n)$ n'est pas nécessairement un entier.

Soit $c > 0$ et $m = \lfloor c n \ln(n) \rfloor \in \mathbb{N}$. Comme C_n est à valeurs entières, $C_n > c n \ln(n)$ équivaut à $C_n > m$ (un entier est strictement supérieur à un réel y si et seulement s'il est strictement supérieur à $\lfloor y \rfloor$).

- Si $m \geq 1$: par b), $P(C_n > c n \ln(n)) = P(C_n > m) \leq n e^{-\frac{m}{n}}$. Comme $m > c n \ln(n) - 1$, on a $-\frac{m}{n} < -c \ln(n) + \frac{1}{n}$, d'où $n e^{-\frac{m}{n}} \leq n e^{\frac{1}{n}} e^{-c \ln(n)} = \frac{1}{e^n} n^{1-c}$.
- Si $m = 0$: alors $c n \ln(n) < 1$, soit $c \ln(n) < \frac{1}{n}$, et $\frac{1}{e^n} n^{1-c} = n e^{\frac{1}{n} - c \ln(n)} \geq n \geq 1 \geq P(C_n > c n \ln(n))$.

Conclusion

$$\text{pour tout réel } c > 0, \quad P(C_n > c n \ln(n)) \leq \frac{1}{e^n} n^{1-c}.$$

- d) Rectification de l'énoncé : le sujet original demandait $P(V_n > x) \leq e^{-x}$; cette majoration est fautive en général (pour $n = 2$ et $x = 0,95 - \ln(2) \approx 0,26$, on a

$P(V_2 > x) = P(C_2 > 1,9) = 1 > e^{-x} \approx 0,77$). Le sujet adapté demande

$P(V_n > x) \leq \frac{1}{e^n} e^{-x}$, conséquence de c) rectifiée. Ce résultat n'est pas utilisé dans la

suite.

Soit $x > -\ln(n)$. Comme $n \geq 2$, $\ln(n) > 0$ et l'on pose $c = 1 + \frac{x}{\ln(n)}$; alors $c > 0$ et $c n \ln(n) = n \ln(n) + nx$. Donc :

$$[V_n > x] = [C_n > n \ln(n) + nx] = [C_n > c n \ln(n)].$$

$$\text{D'après c), } P(V_n > x) \leq e^{\frac{1}{n}} n^{1-c} = e^{\frac{1}{n}} e^{(1-c)\ln(n)} = e^{\frac{1}{n}} e^{-x}.$$

Conclusion

$$\text{pour tout } x > -\ln(n), \quad P(V_n > x) \leq e^{\frac{1}{n}} e^{-x}.$$

- 17) a) Rectification de l'énoncé : le sujet original parlait de «la loi conditionnelle du couple $(\tilde{A}_i, N)$ conditionnée par l'événement $[N = p]$ », qui ne peut pas être une loi binomiale (N est constante égale à p sachant $[N = p]$); il s'agit de la loi conditionnelle de $\tilde{A}_i$ sachant $[N = p]$, ce que précise le sujet adapté.

Loi conditionnelle. Soit $p \in \mathbb{N}^*$. Sachant $[N = p]$, on achète p paquets; d'après le modèle, leurs vignettes ont des numéros indépendants et uniformément répartis sur $\llbracket 1, n \rrbracket$.

Comme en 15) a), le nombre de vignettes numérotées i parmi ces p paquets suit, sous la probabilité $P_{[N=p]}$, la loi $\mathcal{B}\left(p, \frac{1}{n}\right)$:

$$\forall k \in \llbracket 0, p \rrbracket, \quad P_{[N=p]}([\tilde{A}_i = k]) = \binom{p}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{p-k},$$

et $P_{[N=p]}([\tilde{A}_i = k]) = 0$ si $k > p$. Sachant $[N = 0]$, aucune vignette n'est obtenue, donc $\tilde{A}_i = 0$: la même formule reste vraie pour $p = 0$ (avec $\binom{0}{0} = 1$).

Loi de $\tilde{A}_i$. Soit $k \in \mathbb{N}$. La famille $([N = p])_{p \in \mathbb{N}}$ est un système complet d'événements, de probabilités $P(N = p) = e^{-\lambda} \frac{\lambda^p}{p!} > 0$. Par la formule des probabilités totales, en ne gardant que les termes $p \geq k$:

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

Avec $j = p - k$, la série exponentielle donne

$$\sum_{j=0}^{+\infty} \frac{1}{j!} \left(\lambda - \frac{\lambda}{n}\right)^j = e^{\lambda - \frac{\lambda}{n}}.$$

Donc :

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

Conclusion

sachant $[N = p]$, $\tilde{A}_i$ suit la loi $\mathcal{B}\left(p, \frac{1}{n}\right)$, et $\tilde{A}_i$ suit la loi de Poisson $\mathcal{P}\left(\frac{\lambda}{n}\right)$.

- b)** Soit $p \in \mathbb{N}^*$ et $(k_1, \dots, k_n) \in \mathbb{N}^n$ tel que $k_1 + \dots + k_n = p$. Sachant $[N = p]$, les numéros des p vignettes forment une suite $(v_1, \dots, v_p) \in \llbracket 1, n \rrbracket^p$ dont chaque valeur possible a la probabilité $\left(\frac{1}{n}\right)^p$ (indépendance et uniformité). L'événement $[\tilde{A}_1 = k_1] \cap \dots \cap [\tilde{A}_n = k_n]$ correspond aux suites contenant exactement k_j fois le numéro j pour chaque j . Pour les dénombrer, on choisit les k_1 positions du numéro 1 parmi p , puis les k_2 positions du numéro 2 parmi les $p - k_1$ restantes, et ainsi de suite :

$$\binom{p}{k_1} \binom{p-k_1}{k_2} \dots \binom{p-k_1-\dots-k_{n-1}}{k_n} = \frac{p!}{k_1! (p-k_1)!} \cdot \frac{(p-k_1)!}{k_2! (p-k_1-k_2)!} \dots = \frac{p!}{k_1! \dots k_n!},$$

le produit étant télescopique et le dernier facteur valant $\frac{k_n!}{k_n! 0!}$ puisque $p - k_1 - \dots - k_{n-1} = k_n$.

Conclusion

$$P_{[N=p]}([\tilde{A}_1 = k_1] \cap \dots \cap [\tilde{A}_n = k_n]) = \frac{p!}{k_1! \dots k_n!} \frac{1}{n^p}.$$

c) Soit $(k_1, \dots, k_n) \in \mathbb{N}^n$ et $p = k_1 + \dots + k_n$. Comme chaque paquet contient exactement une vignette, $\tilde{A}_1 + \dots + \tilde{A}_n = N$. On note $E = [\tilde{A}_1 = k_1] \cap \dots \cap [\tilde{A}_n = k_n]$; alors $E \subset [N = p]$, donc $P(E) = P(E \cap [N = p]) = P(N = p) P_{[N=p]}(E)$.

• Si $p \geq 1$, d'après b) :

$$P(E) = e^{-\lambda} \frac{\lambda^p}{p!} \cdot \frac{p!}{k_1! \dots k_n!} \frac{1}{n^p} = \prod_{i=1}^n \frac{e^{-\frac{\lambda}{n}} \left(\frac{\lambda}{n}\right)^{k_i}}{k_i!},$$

en écrivant

$$e^{-\lambda} = \prod_{i=1}^n e^{-\frac{\lambda}{n}}$$

et

$$\frac{\lambda^p}{n^p} = \prod_{i=1}^n \left(\frac{\lambda}{n}\right)^{k_i}.$$

• Si $p = 0$, tous les k_i sont nuls et $E = [N = 0]$ (aucune vignette n'est obtenue si et

seulement si aucun paquet n'est acheté), donc

$$P(E) = e^{-\lambda} = \prod_{i=1}^n e^{-\frac{\lambda}{n}},$$

ce qui est la formule annoncée.

Conclusion

pour tout $(k_1, \dots, k_n) \in \mathbb{N}^n$,

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

d) D'après a),

$$P(\tilde{A}_i = k_i) = \frac{e^{-\frac{\lambda}{n}} \left(\frac{\lambda}{n}\right)^{k_i}}{k_i!}.$$

La formule de c) s'écrit donc :

$$\forall (k_1, \dots, k_n) \in \mathbb{N}^n, \quad P([\tilde{A}_1 = k_1] \cap \dots \cap [\tilde{A}_n = k_n]) = \prod_{i=1}^n P(\tilde{A}_i = k_i),$$

ce qui caractérise l'indépendance mutuelle de variables aléatoires discrètes.

Commentaire : lorsque le nombre m d'achats est fixé (question 15) c)), les effectifs $A_{1,m}, \dots, A_{n,m}$ sont liés par la contrainte

$$\sum_i A_{i,m} = m :$$

un grand nombre de vignettes numérotées 1 laisse moins de place aux autres numéros, d'où la covariance négative et la dépendance. Lorsque le nombre d'achats N est lui-même aléatoire de loi de Poisson, cette contrainte disparaît et les effectifs deviennent indépendants, chacun de loi $\mathcal{P}\left(\frac{\lambda}{n}\right)$. C'est l'intérêt de ce modèle « poissonisé ».

Conclusion

$\tilde{A}_1, \dots, \tilde{A}_n$ sont mutuellement indépendantes, contrairement à $A_{1,m}, \dots, A_{n,m}$.

18) La collection est complète à l'issue des N achats si et seulement si chaque numéro a été obtenu au moins une fois : $D_n = \bigcap_{i=1}^n [\tilde{A}_i \geq 1]$. Les variables $\tilde{A}_i$ étant mutuellement indépendantes (question 17) d)), les événements $[\tilde{A}_i \geq 1]$ le sont aussi, et d'après 17) a) :

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

Conclusion

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

19) a) On applique la formule des probabilités totales avec le système complet d'événements $([N = p])_{p \in \mathbb{N}}$, de probabilités non nulles :

$$P(D_n) = \sum_{p=0}^{+\infty} P(N = p) P_{[N=p]}(D_n).$$

Pour $p \geq 1$, le résultat admis donne $P_{[N=p]}(D_n) = P(C_n \leq p)$. Pour $p = 0$, aucune vignette n'est obtenue et $n \geq 2$, donc $P_{[N=0]}(D_n) = 0$; de même $P(C_n \leq 0) = 0$ puisque $C_n \geq n$. La formule vaut donc aussi pour $p = 0$.

Conclusion

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

b) Comme $a < \lambda$, $\lambda - a > 0$ et $k_1 = \lfloor \lambda - a \rfloor \in \mathbb{N}$. Montrons l'inclusion

$$[|N - \lambda| < a] \subset [k_1 \leq N \leq k_2]. \text{ Si } |N - \lambda| < a :$$

- $N > \lambda - a \geq \lfloor \lambda - a \rfloor = k_1$, donc $N \geq k_1$;
- $N < \lambda + a < \lfloor \lambda + a \rfloor + 1 = k_2$, donc $N \leq k_2$.

La variable $N \sim \mathcal{P}(\lambda)$ admet une variance égale à λ et une espérance égale à λ ; par l'inégalité de Bienaymé-Tchebychev :

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

L'hypothèse $a > \sqrt{\lambda}$ assure que ce minorant est strictement positif.

Conclusion

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

c) La suite $p \mapsto P(C_n \leq p)$ est croissante (fonction de répartition) et à valeurs dans $[0, 1]$.

Minoration. Tous les termes de la série de 19) a) étant positifs :

$$\begin{aligned}
 P(D_n) &\geq \sum_{p=k_1}^{k_2} P(C_n \leq p) P(N = p) \geq P(C_n \leq k_1) \sum_{p=k_1}^{k_2} P(N = p) \\
 &\geq P(C_n \leq k_1) \left(1 - \frac{\lambda}{a^2}\right),
 \end{aligned}$$

en utilisant b) et $P(C_n \leq k_1) \geq 0$.

Majoration. On découpe la série en $p \leq k_2$ et $p > k_2$:

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

Or $k_2 > \lambda + a$, donc $[N > k_2] \subset [N - \lambda > a] \subset [|N - \lambda| \geq a]$, et par Bienaymé-Tchebychev $P(N > k_2) \leq \frac{\lambda}{a^2}$.

Conclusion

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

20) a) On pose $\lambda_n = n \ln(n) + c_n n$ et $a_n = n^{2/3}$. La suite (c_n) , convergente, est bornée : il existe $K > 0$ tel que $|c_n| \leq K$ pour tout n .

Vérification des hypothèses de 19) pour n assez grand.

- $\frac{\lambda_n}{n} = \ln(n) + c_n \geq \ln(n) - K \rightarrow +\infty$; donc, pour n assez grand, $\lambda_n \geq n > 1$ et $\lambda_n \geq n > n^{2/3} = a_n$.
- $\frac{\lambda_n}{a_n^2} = \frac{\ln(n) + c_n}{n^{1/3}} \rightarrow 0$ par croissances comparées (car $\frac{\ln(n)}{n^{1/3}} \rightarrow 0$ et $\frac{c_n}{n^{1/3}} \rightarrow 0$); donc, pour n assez grand, $\frac{\lambda_n}{a_n^2} < 1$, soit $a_n > \sqrt{\lambda_n}$, et l'on a aussi $1 - \frac{\lambda_n}{a_n^2} \geq 0$.

Pour un tel n , on applique 19) c) (le résultat de 19) vaut pour toute loi de Poisson de paramètre $\lambda > 1$; on l'applique au paramètre λ_n). Calculons les trois termes.

- Terme central : $\frac{\lambda_n}{n} = \ln(n) + c_n$, donc $e^{-\frac{\lambda_n}{n}} = \frac{e^{-c_n}}{n}$ et, d'après 18),

$$P(D_n) = \left(1 - \frac{e^{-c_n}}{n}\right)^n.$$
- $\frac{\lambda_n}{a_n^2} = \frac{\ln(n) + c_n}{n^{1/3}}.$
- Majorant : $k_2 = \lfloor \lambda_n + a_n \rfloor + 1 \leq \lambda_n + a_n + 1$, donc $[C_n \leq k_2] \subset [C_n \leq \lambda_n + a_n + 1]$. En divisant par n et en retranchant $\ln(n)$, avec $\frac{\lambda_n}{n} - \ln(n) = c_n$ et $\frac{a_n}{n} = \frac{1}{n^{1/3}}$:

$$P(C_n \leq k_2) \leq P\left(V_n \leq c_n + \frac{1}{n^{1/3}} + \frac{1}{n}\right).$$

- Minorant : $k_1 = \lfloor \lambda_n - a_n \rfloor + 1 > \lambda_n - a_n - 1$. Si $C_n \leq \lambda_n - a_n - 1$, alors $C_n < k_1$; donc $[C_n \leq \lambda_n - a_n - 1] \subset [C_n \leq k_1]$ et, de même :

$$P(C_n \leq k_1) \geq P\left(V_n \leq c_n - \frac{1}{n^{1/3}} - \frac{1}{n}\right).$$

En multipliant cette dernière inégalité par $1 - \frac{\lambda_n}{a_n^2} \geq 0$ et en reportant dans 19) c) :

Conclusion

pour tout n assez grand,

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

- b) Lemme.** Si (c_n) converge vers un réel ℓ , alors $\left(1 - \frac{e^{-c_n}}{n}\right)^n \rightarrow e^{-e^{-\ell}}$. En effet,
 $v_n = \frac{e^{-c_n}}{n} \rightarrow 0$ (numérateur borné), donc $0 < v_n < 1$ pour n assez grand et

$(1 - v_n)^n = \exp(n \ln(1 - v_n))$; l'équivalent usuel $\ln(1 - v_n) \sim -v_n$ donne $n \ln(1 - v_n) \sim -e^{-c_n} \rightarrow -e^{-\ell}$, et l'on conclut par continuité de l'exponentielle.

Soit $x \in \mathbb{R}$. On note $G(x) = e^{-e^{-x}}$ et $\varepsilon_n = \frac{1}{n^{1/3}} + \frac{1}{n} \rightarrow 0$.

Minoration de $F_{V_n}(x)$. On applique a) avec $c'_n = x - \varepsilon_n$, suite convergente de limite x . L'inégalité de droite donne, pour n assez grand :

$$F_{V_n}(x) = P(V_n \leq c'_n + \varepsilon_n) \geq \left(1 - \frac{e^{-c'_n}}{n}\right)^n - \frac{\ln(n) + c'_n}{n^{1/3}}.$$

Majoration de $F_{V_n}(x)$. On applique a) avec $c''_n = x + \varepsilon_n$, de limite x . L'inégalité de gauche donne, pour n assez grand (avec $1 - \frac{\ln(n) + c''_n}{n^{1/3}} > 0$) :

$$F_{V_n}(x) = P(V_n \leq c''_n - \varepsilon_n) \leq \frac{\left(1 - \frac{e^{-c''_n}}{n}\right)^n}{1 - \frac{\ln(n) + c''_n}{n^{1/3}}}.$$

Par le lemme, $\left(1 - \frac{e^{-c'_n}}{n}\right)^n \rightarrow G(x)$ et $\left(1 - \frac{e^{-c''_n}}{n}\right)^n \rightarrow G(x)$; par croissances comparées, $\frac{\ln(n) + c'_n}{n^{1/3}} \rightarrow 0$ et $\frac{\ln(n) + c''_n}{n^{1/3}} \rightarrow 0$. Le minorant et le majorant tendent donc tous deux vers $G(x)$ et, par le théorème d'encadrement, $F_{V_n}(x) \rightarrow G(x)$.

Conclusion

pour tout réel x , $F_{V_n}(x) \rightarrow e^{-e^{-x}}$: la suite $(V_n)_{n \geq 1}$ converge en loi vers une variable aléatoire de loi de Gumbel, ce qui redémontre le résultat de 14) d) par une méthode indépendante de la partie II.

Relevé des adaptations

- Question 2) e) : code adapté en Python. L'énoncé demande un programme Python au lieu d'un programme Scilab. Justification : seul le langage Python est au programme.
- Question 11) a) : code adapté en Python. Le rappel Scilab `grand(1,1,'geom',p)` est remplacé par `np.random.geometric(p)` (avec la précision que, pour $p = 1$, la commande renvoie 1, cas utile pour X_1) et le rappel de `np.log` est ajouté; l'en-tête `function y=simulV(n)` devient `def simulV(n):`. Justification : seul le langage Python est au programme; les commandes utilisées sont expliquées dans l'énoncé.
- Question 11) b) : code adapté en Python. Le « vecteur-ligne » Scilab devient un tableau numpy; l'énoncé rappelle la syntaxe de `input` (avec `int`) et de `np.zeros`, qui ne figurent pas dans la liste des commandes usuelles. Justification : seul le langage Python est au programme; les commandes supplémentaires sont expliquées.
- Question 11) c) : code adapté en Python. `histplot(20,V)` devient `plt.hist(V, 20, density=True)`, la fonction Scilab `f` devient une fonction Python, `fplot2d(absc,f)` devient `plt.plot(absc, f(absc))` suivi de `plt.show()`; l'énoncé explique ces commandes. Justification : seul le langage Python est au programme.
- Question 11) c), figure : les trois graphiques originaux sont reproduits dans le fichier compagnon `sujet_hec_escp_maths2_2022_figures.png` (extrait de la figure originale) et décrits dans l'énoncé. Justification : conservation des données nécessaires dans un format Markdown.
- Question 16) c) : rectification de l'énoncé. La majoration $P(C_n > c n \ln(n)) \leq n^{1-c}$ est remplacée par $P(C_n > c n \ln(n)) \leq \frac{1}{e^n} n^{1-c}$. Justification : la version originale est fautive lorsque $c n \ln(n)$ n'est pas entier (contre-exemple $n = 2, c n \ln(n) = 1,9$: probabilité 1 contre un majorant d'environ 0,77); la version rectifiée découle correctement de 16) b).

- Question 16) d) : rectification de l'énoncé. La majoration $P(V_n > x) \leq e^{-x}$ est remplacée par $P(V_n > x) \leq \frac{1}{e^n} e^{-x}$. Justification : conséquence de la rectification de 16) c) (même contre-exemple) ; ce résultat n'est pas utilisé dans la suite.
- Question 17) a) : rectification de l'énoncé. « La loi conditionnelle du couple $(\tilde{A}_i, N)$ conditionnée par l'événement $[N = p]$ » est remplacée par « la loi conditionnelle de la variable aléatoire $\tilde{A}_i$ sachant l'événement $[N = p]$ », explicitée par la formule de $P_{[N=p]}(\tilde{A}_i = k)$. Justification : une loi de couple ne peut pas être une loi binomiale ; la formule explicite rend la notion de loi conditionnelle accessible avec les probabilités conditionnelles du programme.
- Question 14) d) : énoncé conservé ; la convergence en loi est démontrée par encadrement des fonctions de répartition, sans recours au théorème de Slutsky ni à un résultat général hors programme.
- Question 14) b) (i) : énoncé conservé ; la limite en 1 est obtenue par la formule de Taylor-Young appliquée directement à $h \mapsto \ln(1 - h)$, sans composition de développements limités ni règle de L'Hôpital.