

DM 5

THÈME : DEUX CLASSIQUES EN PROBABILITÉ

Sujet A : faire exercice 1.

Sujet * : exercices 1 et 2.

Exercice 1

Inégalité de concentration : Bienaymé-Tchebychev ou Chernoff?

Pour tout n de $\mathbb{N}^*$, S_n désigne une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbf{P})$ de loi binomiale $\mathcal{B}(n, p)$, avec $0 < p < 1$, et x un réel strictement positif fixé. On pose $q = 1 - p$.

1. À l'aide de l'inégalité de Bienaymé-Tchebychev, majorer la quantité $\mathbf{P}(|S_n - np| \geq x\sqrt{n})$ indépendamment de n et p .

L'objectif de la suite est d'établir une deuxième majoration.

2. Montrer que pour tout $\lambda \in \mathbb{R}_*^+$

$$\mathbf{P}(S_n - np \geq nx) \leq \frac{\mathbf{E}(e^{\lambda(S_n - np)})}{e^{n\lambda x}}.$$

3. Établir la formule $\mathbf{E}(e^{\lambda(S_n - np)}) = e^{-n\lambda p} (p \cdot e^\lambda + q)^n$.

4. Pour tout $\lambda \geq 0$, on pose : $\psi(\lambda) = \ln(p \cdot e^\lambda + q) - \lambda p$.

- a) Montrer que : $\sup_{\lambda \geq 0} \psi''(\lambda) \leq \frac{1}{4}$.

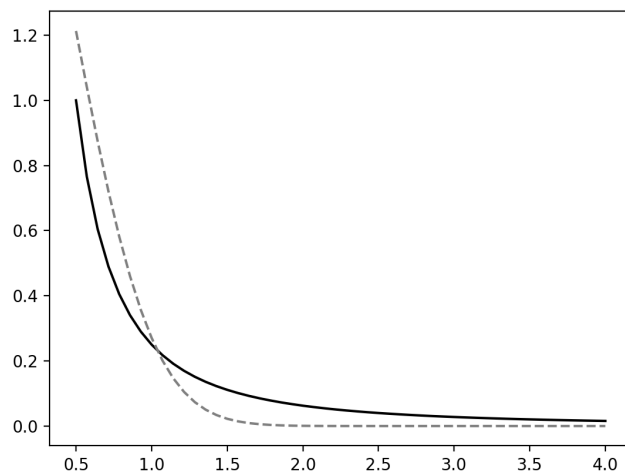
- b) En déduire que pour tout $\lambda \in \mathbb{R}^+$: $\psi(\lambda) \leq \frac{\lambda^2}{8}$.

5. Montrer les inégalités :

$$\mathbf{P}(S_n - np \geq nx) \leq e^{-2nx^2} \quad \text{et} \quad \mathbf{P}(|S_n - np| \geq x\sqrt{n}) \leq 2e^{-2x^2}.$$

6. Commenter le résultat suivant.

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 x=np.linspace(0.5,4,50)
5 y=1/(4*x**2)
6 z=2*np.exp(-2*x**2)
7 plt.plot(x,y,color='black')
8 plt.plot(x,z,'--',color='gray')
9 plt.show()
```



Exercice 2

Problème du collectionneur

On suppose qu'un enfant collectionne des images de joueurs de football qu'il trouve en cadeau dans des tablettes de chocolat. Une tablette contient une seule image et il y a $N \geq 2$ images de joueurs différentes numérotées de 1 à N , que l'entreprise de chocolat a réparties avec la même fréquence $1/N$. Notons X_n le numéro de l'image trouvée dans la n -ième tablette ouverte par l'enfant. Alors, les hypothèses sur la répartition des images impliquent que $(X_n)_{n \in \mathbb{N}^*}$ est une suite de variables uniformes sur $[[1, N]]$.

Pour tout $k \in [[1, N]]$, on note

$$\tau_k^N = \inf \left\{ n \in \mathbb{N}^* \mid \text{Card}(\{X_1, \dots, X_n\}) = k \right\}.$$

Pour tout $N \in \mathbb{N}^*$, on définit aussi

$$H_N = \sum_{k=1}^N \frac{1}{k} \quad \text{et} \quad S_N = \sum_{k=1}^{N-1} \frac{1}{k^2}.$$

1. Simulation

- a) Écrire une fonction python `testcoeff()` qui prend en argument une matrice ligne et compte le nombre de termes non nuls de la matrice.
- b) Compléter le programme suivant qui permet de simuler τ_k^N .

```
1 import numpy as rd
2 import numpy.random as rd
3
4 def tnk(N,k):
5     C=0
6     A=np.zeros(N)
7     while ...
8         i=rd.randint(0,N)
9         # tire au hasard un entier entre 0 et N-1
10        A[i]+=1
11        C= ...
12    return C
```

- c) Que renvoie `tnk(5,7)` ?

2. a) Calculer $\mathbf{P}(\tau_2^N - \tau_1^N = m)$ pour tout $m \in \mathbb{N}^*$. Plus généralement, calculer pour tout $m \in \mathbb{N}^*$,

$$\mathbf{P}(\tau_{k+1}^N - \tau_k^N = m).$$

- b) Calculer également $\mathbf{E}(\tau_{k+1}^N - \tau_k^N)$.

3. Posons $T_N = \tau_N^N$. Calculer $H_N^{-1} \cdot \mathbf{E}(T_N)$.
4. Exprimer la variance de T_N en fonction de H_N et S_N .
5. Déterminer une constante C tel que pour tout $\varepsilon \in \mathbb{R}_*^+$, tout $N \in \mathbb{N}$,

$$\mathbf{P}(|H_N^{-1} T_N - N| \geq N\varepsilon) \leq \frac{C}{\ln(N)^2 \varepsilon^2}.$$

DM 5 - éléments de solution

1. L'inégalité de Bienaymé-Tchebychev donne

$$\mathbf{P}(|S_n - \mathbf{E}[S_n]| \geq x\sqrt{n}) \leq \frac{p(1-p)}{x^2}.$$

Sachant que $p(1-p) \leq 1/4$, $\mathbf{E}[S_n] = np$, on obtient :

$$\mathbf{P}(|S_n - np| \geq x\sqrt{n}) \leq \frac{1}{4x^2}.$$

2. Par stricte croissance de la fonction exponentielle et comme $\lambda \geq 0$, on a :

$$\mathbf{P}(S_n - np \geq nx) = \mathbf{P}(e^{\lambda(S_n - np)} \geq e^{\lambda nx}).$$

On applique alors l'inégalité de Markov à $e^{\lambda(S_n - np)}$ (positive et avec une espérance). Il vient :

$$\mathbf{P}(S_n - np \geq nx) \leq \frac{\mathbf{E}(e^{\lambda(S_n - np)})}{e^{\lambda nx}}.$$

3. Comme S_n suit la loi binomiale $\mathcal{B}(n, p)$, en appliquant le théorème du transfert, il vient :

$$\begin{aligned} \mathbf{E}(e^{\lambda(S_n - np)}) &= \sum_{k=0}^n e^{\lambda(k - np)} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n e^{-\lambda np} \binom{n}{k} (e^{\lambda p})^k (1-p)^{n-k} \\ &= e^{-n\lambda p} (pe^{\lambda} + q)^n. \end{aligned}$$

4.a) La fonction ψ est de classe $\mathcal{C}^2$ et

$$\psi'(\lambda) = -p + \frac{pe^{\lambda}}{pe^{\lambda} + q} \quad \text{et} \quad \psi''(\lambda) = \frac{pqe^{\lambda}}{(pe^{\lambda} + q)^2}.$$

On vérifie que $\psi''(\lambda) \leq \frac{1}{4}$ car

$$(pe^{\lambda} - q)^2 = (pe^{\lambda} + q)^2 - 4pqe^{\lambda} \geq 0.$$

4.b) En intégrant et en vérifiant que $\psi'(0) = \psi(0) = 0$, il vient :

$$\psi'(\lambda) = \int_0^{\lambda} \psi''(u) du \leq \int_0^{\lambda} \frac{1}{4} du = \frac{\lambda}{4}$$

$$\text{et} \quad \psi(\lambda) = \int_0^{\lambda} \psi'(u) du \leq \int_0^{\lambda} \frac{u}{4} du = \frac{\lambda^2}{8}$$

Méthode 2. on peut aussi appliquer l'inégalité de Taylor-Lagrange entre 0 et λ .

5. On sait que pour tout $\lambda > 0$,

$$\mathbf{P}([S_n - np] \geq nx) \leq \frac{\mathbf{E}(e^{\lambda(S_n - np)})}{e^{\lambda nx}} \leq e^{-n(\lambda x - \psi(\lambda))}.$$

Pour résoudre cette question, il suffit donc de montrer l'existence de $\lambda > 0$ tel que $2x^2 \leq \lambda x - \psi(\lambda)$, et par la question précédente, l'existence de $\lambda > 0$ tel que $\frac{\lambda^2}{8} \leq \lambda x - 2x^2$. En résolvant cette inéquation en x (qui est un « carré parfait »), on voit que le réel $\lambda = 4x$ convient. D'où

$$\mathbf{P}\left(\underbrace{S_n - np}_{=A} \geq nx\right) \leq e^{-2nx^2}.$$

Appliquons maintenant le résultat précédent avec $\tilde{S}_n$ de loi binomiale (n, q) . On obtient

$$\mathbf{P}(\tilde{S}_n - nq \geq nx) \leq e^{-2nx^2}.$$

Comme $q = 1 - p$, on a

$$\mathbf{P}((\tilde{S}_n - n) + np \geq nx) \leq e^{-2nx^2}$$

ou encore $\mathbf{P}((n - \tilde{S}_n) - np \leq -nx) \leq e^{-2nx^2}$.

Or par symétrie $n - \tilde{S}_n$ et S_n ont même loi, donc

$$\mathbf{P}\left(\underbrace{S_n - np}_{=B} \leq -nx\right) \leq e^{-2nx^2}.$$

Enfin la formule du crible

$$\begin{aligned} \mathbf{P}(|S_n - np| \geq nx) &= \mathbf{P}(A \cup B) \\ &= \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B) \\ &\leq \mathbf{P}(A) + \mathbf{P}(B) = 2e^{-2nx^2}. \end{aligned}$$

6. Le programme donne les graphes des deux majorations obtenues :

$$x \mapsto \frac{1}{4x^2} \quad \text{et} \quad x \mapsto 2\exp(-2x^2).$$

Dès qu'on dépasse environ 1, la seconde majoration devient bien meilleure (de plus, il y a une forte décroissance de l'exponentielle par rapport aux polynômes).

Exercice 2

1.a,b)

```
def test_coef(A):
    n=len(A)
    return n-np.sum(A==0)

def tnk(N,k):
    C=0
    A=np.zeros(N)
    while test_coef(A)<k:
        i=rd.randint(0,N)
        # tire au hasard un entier entre 0 et N-1
        A[i]+=1
        C+=1
    return C
```

1.c) Dans ce cas $N < k$, c'est impossible et la boucle while ne va jamais se terminer. Il n'y a donc aucun renvoi!

2.a) La variable $\tau_2^N - \tau_1^N$ correspond au nombre de tirages pour avoir une deuxième image différente de la première. À chaque tirage, il y a une probabilité $(N-1)/N$ que cela arrive. Par indépendance des tirages

$$\tau_2^N - \tau_1^N \hookrightarrow \mathcal{G}\left(\frac{N-1}{N}\right).$$

C'est-à-dire pour $m \in \mathbb{N}^*$

$$\mathbf{P}\left(\tau_2^N - \tau_1^N = m\right) = \frac{N-1}{N} \cdot \left(\frac{1}{N}\right)^{m-1}.$$

De même $\tau_{k+1}^N - \tau_k^N \hookrightarrow \mathcal{G}\left(\frac{N-k}{N}\right).$

On sait alors que

$$\mathbf{E}\left(\tau_{k+1}^N - \tau_k^N\right) = \frac{1}{(N-k)/N} = \frac{N}{N-k}.$$

3. On a par télescope et linéarité de l'espérance

$$\mathbf{E}\left(\tau_N^N - \tau_1^N\right) = \sum_{k=1}^{N-1} \mathbf{E}\left(\tau_{k+1}^N - \tau_k^N\right).$$

Avec $\tau_1^N = 1$ et le résultat précédent

$$\begin{aligned} \mathbf{E}\left(\tau_N^N\right) &= 1 + \sum_{k=1}^{N-1} \frac{N}{N-k} \\ &= 1 + \sum_{i=1}^{N-1} \frac{N}{i} \quad (i = N-k) \end{aligned}$$

$$\mathbf{E}\left(\tau_N^N\right) = NH_N.$$

D'où $H_N^{-1} \mathbf{E}\left(\tau_N^N\right) = N.$

4. Par indépendance des variables $\left(\tau_{k+1}^N - \tau_k^N\right)_k$, on a

$$\begin{aligned} \mathbf{V}\left(\tau_N^N\right) &= \mathbf{V}\left(\tau_N^N - \tau_1^N\right) = \mathbf{V}\left(\sum_{k=1}^{N-1} \tau_{k+1}^N - \tau_k^N\right) \\ &= \sum_{k=1}^{N-1} \mathbf{V}\left(\tau_{k+1}^N - \tau_k^N\right) = \sum_{k=1}^{N-1} \frac{k}{N} \cdot \frac{N^2}{(N-k)^2} \\ \mathbf{V}\left(T_N\right) &= N \sum_{k=1}^{N-1} \frac{k}{(N-k)^2}. \end{aligned}$$

On effectue le changement d'indice $i = N - k$

$$\mathbf{V}\left(T_N\right) = N \sum_{i=1}^{N-1} \frac{N-i}{i^2} = N^2 S_N - NH_N.$$

5. À l'aide de l'inégalité de Bienaymé-Tchebychev à la variable $H_N^{-1} T_N$ donne pour tout $\varepsilon \in \mathbb{R}_+^*$

$$\mathbf{P}\left(\left|H_N^{-1} T_N - \mathbf{E}\left(H_N^{-1} T_N\right)\right| \geq N\varepsilon\right) \leq \frac{\mathbf{V}\left(H_N^{-1} T_N\right)}{N^2 \varepsilon^2}.$$

Or

$$\mathbf{V}\left(H_N^{-1} T_N\right) = H_N^{-2} \mathbf{V}\left(T_N\right) = \frac{N^2 S_N}{H_N^2} - \frac{N}{H_N} \leq N^2 \frac{S_N}{H_N^2}.$$

Et $S_N \leq \sum_{k=1}^{+\infty} 1/k^2 = C = \pi^2/6$. Montrons que

$$H_N \geq \ln(N)$$

pour conclure. Puisque la fonction $t \mapsto \frac{1}{t}$ est décroissante, on a, pour tout $k \geq 2$,

$$\int_k^{k+1} \frac{dt}{t} \leq \frac{1}{k} \left(\leq \int_{k-1}^k \frac{dt}{t} \right).$$

Sommant cette relation pour k allant de 2 à $N-1$, puis ajoutant 1, on obtient :

$$1 + \int_2^N \frac{dt}{t} \leq H_N.$$

On calcule l'intégrale :

$$1 + \ln(N) - \ln(2) \leq H_N.$$