

HEC Paris – ESCP Europe

Mathématiques II

Mathématiques approfondies

Sujet adapté – 2017

Les parties du sujet contenant des éléments n'étant plus au programme depuis 2022 ont été adaptées pour convenir aux nouveaux programmes de la filière Mathématiques approfondies en vigueur depuis 2023.

Table des matières

1	Partie I – Une démonstration probabiliste de la formule de Stirling	2
	<i>Thèmes abordés : fonctions d'une variable ; intégrales généralisées ; convergences de variables.</i>	2
2	Partie II – Quelques propriétés de la loi de Cauchy	3
	<i>Thèmes abordés : fonctions d'une variable ; variables à densité.</i>	3
3	Partie III – La loi de la moyenne empirique	4
	<i>Thèmes abordés : variables à densité ; convergences de variables ; estimations ; Python.</i>	4
4	Partie IV – La loi de la médiane empirique	5
	<i>Thèmes abordés : variables à densité ; convergences de variables ; estimations ; Python.</i>	5

Problème – Formule de Stirling et loi de Cauchy

Dans tout le problème, toutes les variables aléatoires introduites sont supposées définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et θ désigne un paramètre réel.

1 Partie I – Une démonstration probabiliste de la formule de Stirling

Thèmes abordés : fonctions d'une variable ; intégrales généralisées ; convergences de variables.

Pour tout $n \in \mathbb{N}^*$, soit h_n la fonction définie par :

$$\forall x \in [0, 1], \quad h_n(x) = ((1-x)e^x)^n.$$

Pour tout $n \in \mathbb{N}^*$, on pose $I_n = \int_0^1 h_n(x) dx$.

1. À l'aide du changement de variable $u = n(1-x)$, montrer que, pour tout $n \in \mathbb{N}^*$,

$$I_n = \frac{e^n}{n^{n+1}} \int_0^n u^n e^{-u} du.$$

- a) Montrer que, pour tout $x \in [0, 1[$, $x + \ln(1-x) \leq -\frac{x^2}{2}$.
 b) En se référant à une densité de la loi normale centrée réduite, en déduire que, pour tout $n \in \mathbb{N}^*$,

$$0 \leq I_n \leq \sqrt{\frac{\pi}{2n}}.$$

2. On note h_n^* la restriction à l'intervalle $]0, 1[$ de la fonction h_n . On pose, pour tout $x \in]0, 1[$,

$$h_n^*(x) = \exp\left(-\frac{nx^2}{2}H(x)\right) \quad \text{et} \quad g(x) = (1-x)\ln(1-x) + x - \frac{x^2}{2}.$$

- a) Montrer que H est prolongeable par continuité en 0. On note encore H la fonction ainsi prolongée.
 b) Montrer que la fonction g est convexe et strictement positive sur $]0, 1[$.
 c) En déduire que la fonction H réalise une bijection strictement croissante de $[0, 1[$ sur $[1, +\infty[$.
 3. Soit $(u_n)_{n \in \mathbb{N}^*}$ une suite convergente de limite nulle telle que

$$\lim_{n \rightarrow +\infty} u_n \sqrt{n} = +\infty \quad \text{et} \quad \forall n \in \mathbb{N}^*, \quad 0 < u_n < 1.$$

- a) Donner un exemple d'une telle suite $(u_n)_{n \in \mathbb{N}^*}$.
 b) Soit $(v_n)_{n \in \mathbb{N}^*}$ la suite définie par : $\forall n \in \mathbb{N}^*, v_n = H(u_n)$. Montrer que la suite $(v_n)_{n \in \mathbb{N}^*}$ est convergente et préciser sa limite.
 c) Établir, pour tout $n \in \mathbb{N}^*$, l'encadrement

$$I_n \geq \int_0^{u_n} h_n(x) dx \geq \frac{1}{\sqrt{nv_n}} \int_0^{u_n \sqrt{nv_n}} \exp\left(-\frac{y^2}{2}\right) dy.$$

- d) Dédurre des questions 1.c) et 3.c) un équivalent de I_n lorsque n tend vers $+\infty$.
4. Soit $(T_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, mutuellement indépendantes et de même loi exponentielle de paramètre 1. Pour tout $n \in \mathbb{N}^*$, on pose $S_n = T_1 + T_2 + \dots + T_n$.
- a) Rappeler la loi suivie par la variable aléatoire S_n et montrer que $\lim_{n \rightarrow +\infty} \mathbb{P}([S_n \leq n]) = \frac{1}{2}$.
- b) Pour tout $n \in \mathbb{N}^*$, on pose $U_n = \frac{T_{n+1}}{\sqrt{n}}$. Montrer que la suite $(U_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers la constante 0.
- Rappel.** *Lemme de Slutsky.* Si X_n converge en loi vers X et si Y_n converge en probabilité vers une constante c , alors $X_n + Y_n$ converge en loi vers $X + c$.
- c) En déduire que $\lim_{n \rightarrow +\infty} \mathbb{P}([S_{n+1} \leq n]) = \frac{1}{2}$.
5. Montrer que $n! \underset{n \rightarrow +\infty}{\sim} n^n e^{-n} \sqrt{2\pi n}$ (formule de Stirling).

2 Partie II – Quelques propriétés de la loi de Cauchy

Thèmes abordés : fonctions d'une variable ; variables à densité.

6. On rappelle que la fonction Arctan est la fonction réciproque de la restriction à l'intervalle ouvert $] -\frac{\pi}{2}, \frac{\pi}{2}[$ de la fonction $\tan$, qu'elle est de classe $\mathcal{C}^\infty$ sur $\mathbb{R}$, admet pour dérivée la fonction $x \mapsto \frac{1}{1+x^2}$ pour tout $x \in \mathbb{R}$ et qu'elle réalise une bijection de $\mathbb{R}$ sur $] -\frac{\pi}{2}, \frac{\pi}{2}[$.
- a) Montrer que la fonction Arctan est impaire.
- b) Justifier l'existence d'un développement limité à l'ordre 3 de la fonction Arctan en 0 et le déterminer.
- c) Établir, pour tout $x \in \mathbb{R}_+$, l'encadrement $0 \leq \text{Arctan}(x) \leq x$.
- d) Montrer que, pour tout $x \in \mathbb{R}_+^*$, $\text{Arctan}(x) + \text{Arctan}\left(\frac{1}{x}\right) = \frac{\pi}{2}$.
7. a) Montrer que la fonction $x \mapsto \frac{1}{\pi} \times \frac{1}{1+(x-\theta)^2}$ est une densité de probabilité sur $\mathbb{R}$.

Dans toute la suite du problème, on note X une variable aléatoire à valeurs réelles, de densité f_X telle que :

$$\forall x \in \mathbb{R}, \quad f_X(x) = \frac{1}{\pi} \times \frac{1}{1+(x-\theta)^2}.$$

On dit que X suit une loi de Cauchy de paramètre θ et on note $X \hookrightarrow \mathcal{C}_\theta$.

- b) La variable aléatoire X admet-elle une espérance ?
- c) Pour $\theta = 0$, tracer la courbe représentative de f_X dans le plan rapporté à un repère orthogonal.
8. a) On note F_X la fonction de répartition de X . Pour tout $x \in \mathbb{R}$, calculer $F_X(x)$.
- b) Montrer que l'équation $F_X(x) = \frac{1}{2}$ d'inconnue x admet une unique solution que l'on déterminera. Cette solution est la *médiane théorique* de X .

3 Partie III – La loi de la moyenne empirique

Thèmes abordés : variables à densité ; convergences de variables ; estimations ; Python.

9. Pour tout $n \in \mathbb{N}^*$ et pour tout $x \in \mathbb{R}$, soit $\varphi_{n,x}$ la fonction définie sur $\mathbb{R}$ par :

$$\forall t \in \mathbb{R}, \quad \varphi_{n,x}(t) = \frac{1}{(1+t^2)(1+(x-nt)^2)}.$$

On admet l'existence d'un unique quadruplet $(\alpha, \beta, \gamma, \delta)$ de réels indépendants de t pour lesquels on a :

$$\forall t \in \mathbb{R}, \quad \varphi_{n,x}(t) = \frac{\alpha t + \beta}{1+t^2} + \frac{\gamma t + \delta}{1+(x-nt)^2}.$$

Pour tout $n \in \mathbb{N}^*$ et pour tout $x \in \mathbb{R}$, on pose $\sigma_{n,x} = (x^2 + (n+1)^2)(x^2 + (n-1)^2)$. On admet sans démonstration que :

$$\alpha = \frac{2nx}{\sigma_{n,x}}, \quad \beta = \frac{1+x^2-n^2}{\sigma_{n,x}}, \quad \gamma = -\frac{2n^2x}{\sigma_{n,x}}, \quad \delta = \frac{n^2(3x^2+n^2-1)}{\sigma_{n,x}}.$$

- a) Établir la convergence de l'intégrale $\int_{-\infty}^{+\infty} \varphi_{n,x}(t) dt$.
 - b) À l'aide d'une primitive de la fonction $\psi_{n,x} : t \mapsto \frac{2t}{1+t^2} - \frac{2n(nt-x)}{1+(x-nt)^2}$, montrer que $\int_{-\infty}^{+\infty} \psi_{n,x}(t) dt = 0$.
 - c) Établir la relation : $\forall n \in \mathbb{N}^*, \forall x \in \mathbb{R}, \int_{-\infty}^{+\infty} \varphi_{n,x}(t) dt = \frac{(n+1)\pi}{x^2 + (n+1)^2}$.
10. On pose $Y = X - \theta$. Pour n entier de $\mathbb{N}^*$, soit $(Y_1, Y_2, \dots, Y_n)$ un n -échantillon de variables aléatoires indépendantes et de même loi que Y . Pour tout $n \in \mathbb{N}^*$, on pose

$$S_n = \sum_{i=1}^n Y_i \quad \text{et} \quad \bar{Y}_n = \frac{S_n}{n},$$

où $\bar{Y}_n$ est la moyenne empirique de l'échantillon $(Y_1, Y_2, \dots, Y_n)$.

- a) Déterminer la fonction de répartition F_Y de Y . Quelle est la loi de Y ?
 - b) Quelle est la fonction de répartition de S_2 ? En déduire la loi de $\bar{Y}_2$.
 - c) Déterminer, pour tout $n \in \mathbb{N}^*$, la loi de la variable aléatoire $\bar{Y}_n$.
 - d) La loi faible des grands nombres s'applique-t-elle à la suite $(\bar{Y}_n)_{n \in \mathbb{N}^*}$? Pourquoi ?
11. Soit $(N, n) \in \mathbb{N}^{*2}$. On veut simuler N réalisations de la moyenne empirique $\bar{Y}_n$. On suppose que l'on connaît une fonction Python `cauchy` telle que l'appel `cauchy(N,n)` renvoie un tableau **A** de taille (N,n) , réalisation d'une famille $(Y_{i,j})_{\substack{1 \leq i \leq N \\ 1 \leq j \leq n}}$ de variables aléatoires indépendantes de loi $\mathcal{C}_0$.

Soit **M** un tableau réel de taille (N,n) . Dans le langage Python (avec `numpy` importé sous le nom `np` et `matplotlib.pyplot` sous le nom `plt`) :

- `np.sum(M)` renvoie la somme de tous les éléments de **M** ;
- `np.sum(M, axis=0)` renvoie un vecteur contenant les sommes des éléments de **M** calculées colonne par colonne ;

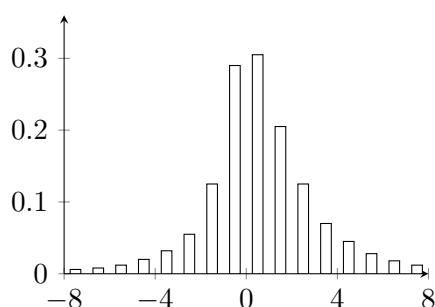
- `np.sum(M, axis=1)` renvoie un vecteur contenant les sommes des éléments de `M` calculées ligne par ligne ;
 - `np.linspace(a,b,m)` renvoie un vecteur de m valeurs régulièrement espacées entre a et b ;
 - `plt.hist(data, bins=x, density=True)` représente les éléments du vecteur `data` sous la forme d'un histogramme dont les classes sont définies par le vecteur strictement croissant `x`.
- a) Compléter le programme suivant afin que le tableau `MoyEmp` contienne 12 000 réalisations de la moyenne empirique $\bar{Y}_{200}$.

```

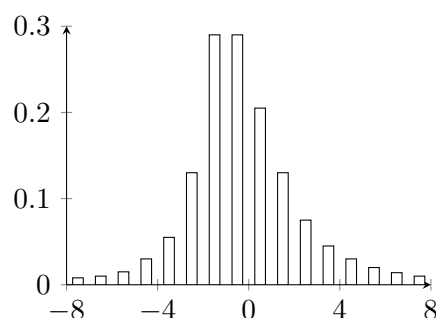
N = 12000 ; n = 200 ;
A = cauchy(N,n)
MoyEmp = .....
x = np.linspace(-8,8,33)
plt.hist(MoyEmp, bins=x, density=True)
# histogramme 1
plt.hist(A[:,0], bins=x, density=True)
# histogramme 2

```

histogramme 1



histogramme 2



- b) Les histogrammes 1 et 2 ont été obtenus à l'aide de ce programme. Expliquer en quoi ce couple d'histogrammes illustre le résultat de la question 10.c).

4 Partie IV – La loi de la médiane empirique

Thèmes abordés : variables à densité ; convergences de variables ; estimations ; Python.

Dans les questions 13, 14 et 15, on suppose que le paramètre θ est inconnu.

On rappelle que $X \hookrightarrow \mathcal{C}_\theta$. Pour n entier de $\mathbb{N}^*$, on note $(X_1, X_2, \dots, X_{2n+1})$ un $(2n+1)$ -échantillon de variables aléatoires indépendantes et de même loi que X .

On admet l'existence de $(2n+1)$ fonctions $g_1, g_2, \dots, g_{2n+1}$ continues sur $\mathbb{R}^{2n+1}$ à valeurs réelles. Les variables aléatoires réelles

$$\hat{X}_k = g_k(X_1, X_2, \dots, X_{2n+1}), \quad k \in \{1, \dots, 2n+1\},$$

sont supposées admettre une densité. Pour tout $\omega \in \Omega$, les réels $\hat{X}_1(\omega), \hat{X}_2(\omega), \dots, \hat{X}_{2n+1}(\omega)$ sont un réarrangement par ordre croissant de $X_1(\omega), X_2(\omega), \dots, X_{2n+1}(\omega)$:

$$\begin{aligned} \forall \omega \in \Omega, \quad \hat{X}_1(\omega) &\leq \hat{X}_2(\omega) \leq \dots \\ &\leq \hat{X}_{2n+1}(\omega). \end{aligned}$$

La variable aléatoire $\hat{X}_{n+1}$ est la *médiane empirique*.

Elle est associée à l'échantillon $(X_1, X_2, \dots, X_{2n+1})$.

12. Pour tout $h \in \mathbb{R}_+^*$ et pour tout $x \in \mathbb{R}$, on note Z la variable aléatoire discrète définie par :

$$\forall \omega \in \Omega, \quad Z(\omega) = \text{Card} \{i \in \{1, \dots, 2n+1\}; \\ x < X_i(\omega) \leq x+h\}.$$

On note A et B les deux événements suivants : $A = [x < \hat{X}_{n+1} \leq x+h]$ et $B = A \cap [Z = 1]$.

a) Établir la relation :

$$\mathbb{P}(B) = (n+1) \binom{2n+1}{n} (F_X(x))^n (F_X(x+h) - F_X(x)) (1 - F_X(x+h))^n.$$

b) On suppose que le réel x est fixé. Montrer qu'il existe un réel K indépendant de h pour lequel on a :

$$0 \leq \mathbb{P}(A) - \mathbb{P}(B) \leq K (F_X(x+h) - F_X(x))^2.$$

c) Montrer que $\hat{X}_{n+1}$ admet une densité $f_{\hat{X}_{n+1}}$ donnée par :

$$\forall x \in \mathbb{R}, \quad f_{\hat{X}_{n+1}}(x) = (n+1) \binom{2n+1}{n} (F_X(x))^n (1 - F_X(x))^n f_X(x).$$

13. a) Établir l'équivalence suivante :

$$x f_{\hat{X}_{n+1}}(x) \underset{x \rightarrow +\infty}{\sim} (n+1) \binom{2n+1}{n} \frac{1}{\pi^{n+1}} \times \frac{1}{x^{n+1}}.$$

b) En déduire l'existence de l'espérance $\mathbb{E}(\hat{X}_{n+1})$ de la variable aléatoire $\hat{X}_{n+1}$.

c) Justifier que $\hat{X}_{n+1}$ est un estimateur du paramètre θ . Calculer $\mathbb{E}(\hat{X}_{n+1} - \theta)$. Conclure.

d) À quelle condition nécessaire et suffisante portant sur n , la variable aléatoire $\hat{X}_{n+1}$ admet-elle une variance ?

14. On note $F_{\hat{X}_{n+1}}$ la fonction de répartition de $\hat{X}_{n+1}$. Soit ε un réel strictement positif.

a) Établir la relation : $\forall t \in \mathbb{R}, f_{\hat{X}_{n+1}}(2\theta - t) = f_{\hat{X}_{n+1}}(t)$. En déduire que $\mathbb{P}(|\hat{X}_{n+1} - \theta| \geq \varepsilon) = 2F_{\hat{X}_{n+1}}(\theta - \varepsilon)$.

b) Montrer que la suite d'estimateurs $(\hat{X}_{n+1})_{n \in \mathbb{N}^*}$ converge en probabilité vers θ .

c) La suite $(\hat{X}_{n+1})_{n \in \mathbb{N}^*}$ converge-t-elle en loi vers la variable certaine θ ?

15. Pour tout entier $n \geq 2$, on pose :

$$W_{n+1} = \frac{2\sqrt{2n+1}}{\pi} (\hat{X}_{n+1} - \theta).$$

a) On note $f_{W_{n+1}}$ la densité continue sur $\mathbb{R}$ de W_{n+1} . Montrer que :

$$\forall x \in \mathbb{R}, \quad f_{W_{n+1}}(x) = \frac{n+1}{2\sqrt{2n+1}} \binom{2n+1}{n} \left[\frac{1}{4} - \frac{1}{\pi^2} \left(\text{Arctan} \left(\frac{\pi x}{2\sqrt{2n+1}} \right) \right)^2 \right]^n \\ \times \left(1 + \frac{\pi^2 x^2}{4(2n+1)} \right)^{-1}.$$

b) Montrer que, pour tout $x \in \mathbb{R}$, $\lim_{n \rightarrow +\infty} f_{W_{n+1}}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$.

On admet que ce résultat implique la convergence en loi de la suite de variables aléatoires $(W_{n+1})_{n \geq 2}$ vers une variable aléatoire T qui suit la loi normale centrée réduite.

c) On note Φ la fonction de répartition de T . Soit α un réel vérifiant $0 < \alpha < 1$; on pose : $t_\alpha = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$. Déterminer un intervalle de confiance asymptotique pour θ , centré sur $\hat{X}_{n+1}$, au niveau de confiance $1 - \alpha$.

16. En Python, la fonction `np.sort(A, axis=1)[: , :-1]` permet de trier, par ordre décroissant, les éléments de chaque ligne d'un tableau réel **A**. On suppose que $\theta = 0$ et on considère p réalisations ($p \geq 10^4$) du $(2n + 1)$ -échantillon $(X_1, X_2, \dots, X_{2n+1})$. Recopier et compléter le code suivant afin que son exécution renvoie un vecteur **MedianeEmp** de p réalisations de la médiane empirique $\hat{X}_{n+1}$, puis un vecteur **W** de p réalisations de W_{n+1} .

```
A = cauchy(p, 2*n+1)
S = .....
MedianeEmp = .....
W = .....
```