



PRÉPASIMPLE

prepasimple.fr

HEC — 2017

Sujet de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Sujet

Dans tout le problème, les variables aléatoires introduites sont définies sur le même espace probabilisé. On note θ un paramètre réel.

Partie I. Une démonstration probabiliste de la formule de Stirling

Pour $n \in \mathbb{N}^*$, on définit h_n sur $[0, 1]$ par :

$$h_n(x) = ((1-x)e^x)^n, \quad I_n = \int_0^1 h_n(x) dx.$$

1. (a) À l'aide du changement de variable $u = n(1-x)$, montrer que :

$$\forall n \in \mathbb{N}^*, \quad I_n = \frac{e^n}{n^{n+1}} \int_0^n u^n e^{-u} du.$$

1. (b) Montrer que, pour $x \in [0, 1[$:

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

1. (c) En se référant à une densité de la loi normale centrée réduite, en déduire que :

$$\forall n \in \mathbb{N}^*, \quad 0 \leq I_n \leq \sqrt{\frac{\pi}{2n}}.$$

2. On note h_n^* la restriction de h_n à $]0, 1[$. On définit H et g par :

$$\forall x \in]0, 1[, \quad h_n^*(x) = \exp\left(-\frac{nx^2}{2}H(x)\right), \quad g(x) = (1-x)\ln(1-x) + x - \frac{x^2}{2}.$$

2. (a) Montrer que H est prolongeable par continuité en 0. On note encore H la fonction prolongée.
2. (b) Montrer que g est convexe et strictement positive sur $]0, 1[$.
2. (c) En déduire que H réalise une bijection strictement croissante de $[0, 1[$ sur $[1, +\infty[$.
3. Soit $(u_n)_{n \geq 1}$ une suite convergeant vers 0 telle que $u_n \sqrt{n} \rightarrow +\infty$ et $0 < u_n < 1$ pour tout $n \geq 1$.
3. (a) Donner un exemple d'une telle suite.
3. (b) On pose $v_n = H(u_n)$. Montrer que (v_n) converge et préciser sa limite.
3. (c) Établir, pour tout $n \in \mathbb{N}^*$:

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

3. (d) Dédurre de 1. (c) et 3. (c) un équivalent de I_n lorsque $n \rightarrow +\infty$.
4. Soient $(T_n)_{n \geq 1}$ des variables aléatoires mutuellement indépendantes de même loi exponentielle de paramètre 1. On pose $S_n = T_1 + \dots + T_n$.
4. (a) Rappeler la loi de S_n et montrer que :

$$\lim_{n \rightarrow +\infty} P(S_n \leq n) = \frac{1}{2}.$$

4. (b) On pose $U_n = T_{n+1}/\sqrt{n}$. Montrer que U_n converge en probabilité vers 0.

4. (c) En déduire que :

$$\lim_{n \rightarrow +\infty} P(S_{n+1} \leq n) = \frac{1}{2}.$$

5. Montrer que :

$$n! \underset{n \rightarrow +\infty}{\sim} n^n e^{-n} \sqrt{2\pi n}.$$

Partie II. Quelques propriétés de la loi de Cauchy

6. La fonction arctan est la réciproque de la restriction de $\tan$ à $] -\pi/2, \pi/2[$. Elle est de classe C^∞ sur $\mathbb{R}$, sa dérivée est $x \mapsto 1/(1+x^2)$ et elle réalise une bijection de $\mathbb{R}$ sur $] -\pi/2, \pi/2[$.

6. (a) Montrer que arctan est impaire.

6. (b) Justifier l'existence d'un développement limité de arctan à l'ordre 3 en 0 et le déterminer.

6. (c) Établir, pour tout $x \in \mathbb{R}_+$:

$$0 \leq \arctan(x) \leq x.$$

6. (d) Montrer que, pour tout $x > 0$:

$$\arctan(x) + \arctan\left(\frac{1}{x}\right) = \frac{\pi}{2}.$$

7. (a) Montrer que la fonction suivante est une densité sur $\mathbb{R}$:

$$x \mapsto \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}.$$

Dans toute la suite, X désigne une variable aléatoire de densité :

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

On dit que X suit la loi de Cauchy de paramètre θ , notée $\mathcal{C}_\theta$.

7. (b) La variable X admet-elle une espérance ?
7. (c) Pour $\theta = 0$, tracer la courbe représentative de f_X dans un repère orthonormal.
8. (a) Calculer la fonction de répartition F_X de X .
8. (b) Montrer que l'équation $F_X(x) = 1/2$ admet une unique solution et la déterminer. Cette solution est la médiane théorique de X .

Partie III. La loi de la moyenne empirique

9. Pour $n \in \mathbb{N}^*$ et $x \in \mathbb{R}$, on définit sur $\mathbb{R}$:

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

On pose $\sigma_{n,x} = (x^2 + (n + 1)^2)(x^2 + (n - 1)^2)$. Lorsque $(n, x) \neq (1, 0)$, on admet la décomposition :

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

avec les coefficients indépendants de t suivants :

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

Pour $(n, x) = (1, 0)$, on a directement $\varphi_{1,0}(t) = (1 + t^2)^{-2}$.

9. (a) Établir la convergence de :

$$\int_{-\infty}^{+\infty} \varphi_{n,x}(t) dt.$$

9. (b) À l'aide d'une primitive de :

$$\psi_{n,x}(t) = \frac{2t}{1 + t^2} - \frac{2n(nt - x)}{1 + (x - nt)^2},$$

montrer que $\int_{-\infty}^{+\infty} \psi_{n,x}(t) dt = 0$.

9. (c) Établir, pour tous $n \in \mathbb{N}^*$ et $x \in \mathbb{R}$:

$$\int_{-\infty}^{+\infty} \varphi_{n,x}(t) dt = \frac{(n+1)\pi}{x^2 + (n+1)^2}.$$

10. On pose $Y = X - \theta$. Pour $n \in \mathbb{N}^*$, soient $Y_1, \dots, Y_n$ indépendantes de même loi que Y . On pose :

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

La notation S_n désigne désormais cette somme, différente de celle de la partie I.

10. (a) Déterminer la fonction de répartition F_Y . Quelle est la loi de Y ?
10. (b) Quelle est la fonction de répartition de S_2 ? En déduire la loi de $\bar{Y}_2$.
10. (c) Déterminer, pour tout $n \in \mathbb{N}^*$, la loi de $\bar{Y}_n$.
10. (d) La loi faible des grands nombres s'applique-t-elle à $(\bar{Y}_n)_{n \geq 1}$? Pourquoi ?
11. Soient $N, n \in \mathbb{N}^*$. On souhaite simuler N réalisations de $\bar{Y}_n$. On dispose d'une fonction Python `cauchy(N, n)` qui renvoie un tableau de forme (N, n) dont les coefficients sont des simulations indépendantes de loi $\mathcal{C}_0$. Cette fonction peut être définie par :

Python

```
import numpy as np

def cauchy(N, n):
    return np.random.standard_cauchy((N, n))
```

La commande `np.random.standard_cauchy((N, n))` simule Nn variables indépendantes de loi $\mathcal{C}_0$ et les range dans un tableau à N lignes et n colonnes.

Pour un tableau M de forme (N, n) , `np.sum(M)` calcule la somme de tous les coefficients, `np.sum(M, axis=0)` les sommes par colonne et `np.sum(M, axis=1)` les sommes par ligne. La commande `np.linspace(a, b, m)` fournit m valeurs régulièrement espacées de a à b , extrémités comprises.

Dans `plt.hist(data, bins=x, weights=poids)`, x donne les bornes des classes et `poids` attribue un poids à chaque observation. Pour des classes de largeur 0.5, les poids

$1/(0.5N)$ donnent une hauteur égale à l'effectif de la classe divisé par $0.5N$. Les observations hors des classes affichées ne sont pas représentées. Les paramètres de couleurs fournis donnent des barres blanches bordées de noir.

11. (a) Compléter le programme pour que MoyEmp contienne 12000 réalisations de $\bar{Y}_{200}$:

Python

```
import numpy as np

import matplotlib.pyplot as plt

def cauchy(N, n):

    return np.random.standard_cauchy((N, n))

N = 12000

n = 200

A = cauchy(N, n)

MoyEmp = ...

x = np.linspace(-8, 8, 33)

poids = np.ones(N) / (0.5 * N)

plt.hist(MoyEmp, bins=x, weights=poids,

         facecolor="white", edgecolor="black")

plt.title("histogramme 1")

plt.show()

plt.hist(A[:, 0], bins=x, weights=poids,

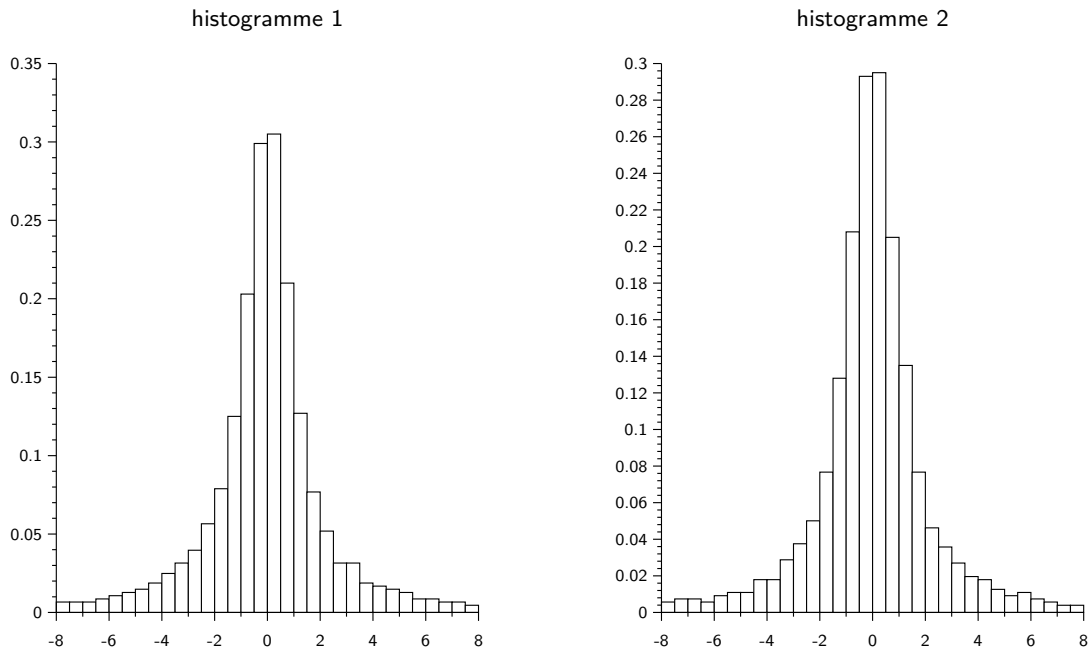
         facecolor="white", edgecolor="black")

plt.title("histogramme 2")

plt.show()
```

La commande `plt.title(texte)` donne son titre au graphique. L'expression `A[:, 0]` extrait la première colonne de A.

11. (b) Les histogrammes 1 et 2 ci-dessous ont été obtenus à l'aide du programme. Expliquer en quoi ils illustrent le résultat de 10. (c).



Partie IV. La loi de la médiane empirique

Dans les questions 13, 14 et 15, on suppose θ inconnu. On rappelle $X \sim \mathcal{C}_\theta$. Pour $n \in \mathbb{N}^*$, soient $X_1, \dots, X_{2n+1}$ des variables indépendantes de même loi que X .

On admet l'existence de fonctions $g_1, \dots, g_{2n+1}$ continues de $\mathbb{R}^{2n+1}$ dans $\mathbb{R}$, telles que $\widehat{X}_k = g_k(X_1, \dots, X_{2n+1})$ soient des variables à densité et que, pour chaque issue, $\widehat{X}_1, \dots, \widehat{X}_{2n+1}$ soient les valeurs de l'échantillon rangées par ordre croissant :

$$\widehat{X}_1 \leq \widehat{X}_2 \leq \dots \leq \widehat{X}_{2n+1}.$$

En particulier, $\widehat{X}_{n+1}$ est la médiane empirique de l'échantillon.

12. Pour $h > 0$ et $x \in \mathbb{R}$, on définit :

$$Z = \text{Card}\{i \in \{1, \dots, 2n+1\} : x < X_i \leq x+h\}.$$

On pose $A = [x < \widehat{X}_{n+1} \leq x + h]$ et $B = A \cap [Z = 1]$.

12. (a) Établir :

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

12. (b) À x fixé, montrer qu'il existe un réel K indépendant de h tel que :

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

12. (c) Montrer que $\widehat{X}_{n+1}$ admet la densité :

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

13. (a) Établir :

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

13. (b) En déduire l'existence de l'espérance de $\widehat{X}_{n+1}$.

13. (c) Justifier que $\widehat{X}_{n+1}$ est un estimateur de θ . Calculer $E(\widehat{X}_{n+1} - \theta)$ et conclure.

13. (d) À quelle condition nécessaire et suffisante sur n la variable $\widehat{X}_{n+1}$ admet-elle une variance ?

14. On note $F_{\widehat{X}_{n+1}}$ la fonction de répartition de $\widehat{X}_{n+1}$. Soit $\varepsilon > 0$.

14. (a) Établir, pour tout $t \in \mathbb{R}$:

$$f_{\widehat{X}_{n+1}}(2\theta - t) = f_{\widehat{X}_{n+1}}(t).$$

En déduire que :

$$P(|\widehat{X}_{n+1} - \theta| \geq \varepsilon) = 2F_{\widehat{X}_{n+1}}(\theta - \varepsilon).$$

14. (b) Montrer que $(\widehat{X}_{n+1})_{n \geq 1}$ converge en probabilité vers θ .

14. (c) Cette suite converge-t-elle en loi vers la variable certaine θ ?

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

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

15. (a) On note $f_{W_{n+1}}$ la densité continue de W_{n+1} . Montrer que :

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

15. (b) Montrer que, pour tout $x \in \mathbb{R}$:

$$\lim_{n \rightarrow +\infty} f_{W_{n+1}}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

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

15. (c) On note Φ la fonction de répartition de T . Pour $0 < \alpha \leq 1$, on pose $t_\alpha = \Phi^{-1}(1 - \alpha/2)$. Déterminer un intervalle de confiance asymptotique pour θ , centré en $\widehat{X}_{n+1}$, de niveau $1 - \alpha$.
16. En Python, `np.sort(A, axis=0)` renvoie une copie de A dont chaque colonne est triée par ordre croissant; `np.sort(A, axis=1)` trie chaque ligne séparément par ordre croissant.

On suppose $\theta = 0$ et $n \geq 2$, et on considère $p \geq 10^4$ réalisations d'un échantillon de taille $2n + 1$. Recopier et compléter le programme pour que `MedianeEmp` contienne p réalisations de $\widehat{X}_{n+1}$ et `W` contienne p réalisations de W_{n+1} :

Python

```
import numpy as np

def cauchy(N, n):
    return np.random.standard_cauchy((N, n))

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