MINISTÈRE
DE L'ÉDUCATION
NATIONALE,
DE LA JEUNESSE
ET DES SPORTS
*Liberté*
*Égalité*
*Fraternité*

**EAI MAT 2**

**SESSION 2022**

# AGRÉGATION
# CONCOURS INTERNE
# ET CAER

Section
**MATHÉMATIQUES**

## DEUXIÈME ÉPREUVE

Durée : 6 heures

*L'usage de tout ouvrage de référence, de tout dictionnaire et de tout matériel électronique (y compris la calculatrice) est rigoureusement interdit.*

*Si vous repérez ce qui vous semble être une erreur d'énoncé, vous devez le signaler très lisiblement sur votre copie, en proposer la correction et poursuivre l'épreuve en conséquence. De même, si cela vous conduit à formuler une ou plusieurs hypothèses, vous devez la (ou les) mentionner explicitement.*

**NB : Conformément au principe d'anonymat, votre copie ne doit comporter aucun signe distinctif, tel que nom, signature, origine, etc. Si le travail qui vous est demandé consiste notamment en la rédaction d'un projet ou d'une note, vous devrez impérativement vous abstenir de la signer ou de l'identifier.**

**Tournez la page S.V.P.**

A

**INFORMATION AUX CANDIDATS**

Vous trouverez ci-après les codes nécessaires vous permettant de compléter les rubriques figurant en en-tête de votre copie.

Ces codes doivent être reportés sur chacune des copies que vous remettrez.

► **Concours interne de l'Agrégation de l'enseignement public** :

| Concours | Section/option | Epreuve | Matière |
|---|---|---|---|
| EAI | 1300A | 102 | 0530 |

► **Concours interne du CAER / Agrégation de l'enseignement privé** :

| Concours | Section/option | Epreuve | Matière |
|---|---|---|---|
| EAH | 1300A | 102 | 0530 |

**Notations**.

*Dans tout le sujet,* $\mathbb{R}$ *désignera le corps des nombres réels et* $\mathbb{N}$ *l'ensemble des entiers naturels. De plus,* $n$ *désignera un entier naturel strictement supérieur à 1.*

*On considère un espace probabilisé* $(\Omega, \mathcal{A}, \mathbf{P})$*. Pour* $X$ *est une variable aléatoire réelle définie sur* $\Omega$*, on note* $\mathbf{E}(X)$ *son espérance (lorsqu'elle existe) et* $\mathbf{V}(X)$ *sa variance (lorsqu'elle existe).*
*On rappelle que, sous réserve d'existence, la covariance de deux variables aléatoires* $X$ *et* $Y$ *définies sur* $\Omega$ *est le nombre réel noté* $\mathbf{Cov}(X,Y)$ *et défini par*

$$\mathbf{Cov}(X,Y) = \mathbf{E}\big((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))\big) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y).$$

*En conséquence, sous réserve de l'existence de* $\mathbf{V}(X)$, $\mathbf{V}(Y)$ *et* $\mathbf{V}(X+Y)$ *:*

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

**Objectifs du problème.**

*Soit* $X$ *une variable aléatoire définie sur* $\Omega$*. On se place dans le contexte où la loi de* $X$ *n'est pas complètement spécifiée et où cette loi dépend d'un paramètre* $\theta$ *inconnu. Le but de l'estimation consiste à approcher la valeur de ce paramètre* $\theta$*, ou éventuellement la valeur de son image* $g(\theta)$ *par une fonction réelle* $g$*.*
*La première partie revient sur quelques résultats au sujet des séries entières. La deuxième partie étudie les familles sommables. Dans les deux dernières parties, on se place dans le cas où* $X$ *est une variable aléatoire définie sur* $\Omega$ *suivant une loi de Poisson de paramètre* $\theta$ *inconnu et on développe quelques outils pour l'estimation de ce paramètre.*

# I. Séries entières.

**1.** Les affirmations suivantes sont-elles vraies ou fausses ? On justifiera soigneusement les réponses.

- **(a)** Affirmation : « soient $(u_k)_{k\in\mathbb{N}}$ et $(v_k)_{k\in\mathbb{N}}$ deux suite de nombre réels positifs, telles que la série de terme général $v_k$ converge et que pour tout $k \in \mathbb{N}$, $u_k \leqslant v_k$. Alors la série de terme général $u_k$ converge également ».
- **(b)** Affirmation : « soient $(u_k)_{k\in\mathbb{N}}$ et $(v_k)_{k\in\mathbb{N}}$ deux suite de nombre réels, telles que la série de terme général $v_k$ converge et que pour tout $k \in \mathbb{N}$, $u_k \leqslant v_k$. Alors la série de terme général $u_k$ converge également ».
- **(c)** Affirmation : « soit $(u_k)_{k\in\mathbb{N}}$ une suite de nombres réels tous non nuls. On suppose que

$$\lim_{k\longrightarrow+\infty} \left|\frac{u_{k+1}}{u_k}\right| = \ell,$$

avec $\ell < 1$. Alors la série de terme général $u_k$ converge absolument ».
- **(d)** Affirmation : « soit $(u_k)_{k\in\mathbb{N}}$ une suite de fonctions réelles continues définies sur un même intervalle $I$, telle que pour tout $x \in I$, la suite $(u_k(x))_{k\in\mathbb{N}}$ converge vers un réel $U(x)$. La fonction $U$ ainsi définie sur $I$ est continue ».

2. *Question de cours*. Soit $(u_k)_{k\in\mathbb{N}}$ une suite de nombres réels. On rappelle que le rayon de convergence de la série entière de terme général $u_k x^k$ est défini par

$$R = \sup\{x \in \mathbb{R} \mid (|u_k x^k|)_{k\in\mathbb{N}} \text{ est majorée}\} \in \mathbb{R}^+ \cup \{+\infty\}.$$

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

    i. Montrer que si $|x| < R$, alors la série de terme général $u_k x^k$ converge absolument.

    ii. Montrer que si $|x| > R$, alors la série de terme général $u_k x^k$ diverge.

  On considère alors la fonction $S : ]-R, R[\longrightarrow \mathbb{R}$ définie par

$$S(x) = \sum_{k=0}^{+\infty} u_k x^k.$$

  **(b)** Soit $R' \in ]0, R[$. Montrer que la série de terme général $u_k x^k$ converge uniformément sur $[-R', R']$. Que peut-on en déduire sur la régularité de $S$ ?

  **(c)** Montrer que le rayon de convergence de la série de terme général $(k+1)u_{k+1}x^k$ est égal à $R$.

  **(d)** Montrer que $S$ est indéfiniment dérivable sur $]-R, R[$.

3. Soit $r$ un entier naturel non nul. Déterminer le rayon de convergence de la série entière de terme général $\dfrac{k^r x^k}{k!}$. Déterminer sa somme lorsque $r = 1$ et lorsque $r = 2$.

4. On note, pour tout entier $N \geqslant 1$, $\alpha_N$ le nombre de partitions de l'ensemble $[\![1, N]\!]$ et on convient que $\alpha_0 = 1$.

  **(a)** Calculer $\alpha_1$, $\alpha_2$ et $\alpha_3$.

  **(b)** Montrer que pour tout entier $N \geqslant 0$,

$$\alpha_{N+1} = \sum_{k=0}^{N} \binom{N}{k} \alpha_k.$$

  **(c)** Montrer que, pour tout $N \geqslant 0$, $\alpha_N \leqslant N!$.

  **(d)** En déduire que la série entière de terme général $\dfrac{\alpha_N x^N}{N!}$ converge pour tout $x$ réel tel que $|x| < 1$. On note $f(x)$ sa somme.

  **(e)** Montrer que $f$ est dérivable sur $]-1, 1[$ et que pour tout $x \in ]-1, 1[$, $f'(x) = \mathrm{e}^x f(x)$.
  En déduire que pour tout $x \in ]-1, 1[$, $f(x) = \mathrm{e}^{\mathrm{e}^x - 1}$.

  **(f)** En déduire que pour tout entier naturel $N$,

$$\alpha_N = \frac{1}{\mathrm{e}} \sum_{k=0}^{+\infty} \frac{k^N}{k!}.$$

## II. Familles sommables.

*Soit $I \subset \mathbb{N}^n$. Les éléments de I seront notés sous la forme $\underline{i} = (i_1, i_2, \ldots, i_n)$.*

***Cas des familles sommables de réels positifs***.

*Soit $u = (u_{\underline{i}})_{\underline{i}\in I}$ une famille de réels positifs ou nuls indexée par I. On dit que u est sommable lorsque la borne supérieure suivante est finie :*

$$\sup \left\{ \sum_{\underline{i}\in J} u_{\underline{i}},\ J \subset I \text{ et } J \text{ fini} \right\} < +\infty.$$

*Dans ce cas, on note*

$$\sum_{\underline{i}\in I} u_{\underline{i}} = \sup \left\{ \sum_{\underline{i}\in J} u_{\underline{i}},\ J \subset I \text{ et } J \text{ fini} \right\}.$$

Soit $u = (u_{\underline{i}})_{\underline{i}\in I}$ et $(v_{\underline{i}})_{\underline{i}\in I}$ deux familles de réels positifs et $a$, $b$ deux réels positifs.

**5.** On suppose que pour tout $\underline{i} \in I$ ,

$$u_{\underline{i}} \leqslant v_{\underline{i}}.$$

Montrer que si $v$ est sommable, alors $u$ est sommable.

**6.** On suppose que $u$ et $v$ sont sommables. Montrer que la famille $au + bv$ définie par

$$(au + bv)_{\underline{i}} = au_{\underline{i}} + bv_{\underline{i}}$$

est sommable et que

$$\sum_{\underline{i}\in I}(au_{\underline{i}} + bv_{\underline{i}}) = a\left(\sum_{\underline{i}\in I} u_{\underline{i}}\right) + b\left(\sum_{\underline{i}\in I} v_{\underline{i}}\right).$$

**7.** On considère $(I_k)_{k\in\mathbb{N}}$ une famille de sous-ensembles de $I$ tels que

- ▷ $\bigcup_{k\in\mathbb{N}} I_k = I$.
- ▷ Pour tout $k, l \in \mathbb{N}$, distincts, $I_k \cap I_l = \emptyset$.

Par convention, si $I_k = \emptyset$, on pose

$$\sum_{\underline{i}\in I_k} u_{\underline{i}} = 0.$$

On suppose que $u$ est sommable.

**(a)** Montrer que la famille $(u_{\underline{i}})_{\underline{i}\in I_k}$ est sommable pour tout $k \in \mathbb{N}$.

**(b)** Montrer que pour tout $p \in \mathbb{N}$,

$$\sum_{k=0}^{p}\left(\sum_{\underline{i}\in I_k} u_{\underline{i}}\right) \leqslant \sum_{\underline{i}\in I} u_{\underline{i}}.$$

**(c)** Montrer que la série de terme général $\sum_{\underline{i}\in I_k} u_{\underline{i}}$ converge. Sa somme est notée $\sum_{k=0}^{+\infty}\sum_{\underline{i}\in I_k} u_{\underline{i}}$.



**Tournez la page S.V.P.**

**(d)** Montrer que

$$\sum_{k=0}^{+\infty} \sum_{\underline{i} \in I_k} u_{\underline{i}} \leqslant \sum_{\underline{i} \in I} u_{\underline{i}}.$$

**8.** Réciproquement, montrer que si la famille $(u_{\underline{i}})_{\underline{i} \in I_k}$ est sommable pour tout $k \in \mathbb{N}$ et si la série de terme général $\displaystyle\sum_{\underline{i} \in I_k} u_{\underline{i}}$ converge, alors $(u_{\underline{i}})_{\underline{i} \in I}$ est sommable et

$$\sum_{\underline{i} \in I} u_{\underline{i}} \leqslant \sum_{k=0}^{+\infty} \sum_{\underline{i} \in I_k} u_{\underline{i}}.$$

**9.** Déduire des questions précédentes le résultat suivant, appelé *théorème de sommation par paquets* :
$(u_{\underline{i}})_{\underline{i} \in I_k}$ est sommable pour tout $k \in \mathbb{N}$ et la série de terme général $\displaystyle\sum_{\underline{i} \in I_k} u_{\underline{i}}$ converge si et seulement si $(u_{\underline{i}})_{\underline{i} \in I}$ est sommable. De plus, dans ce cas,

$$\sum_{\underline{i} \in I} u_{\underline{i}} = \sum_{k=0}^{+\infty} \sum_{\underline{i} \in I_k} u_{\underline{i}}.$$

***Cas des familles sommables de réels quelconques***.

*Soit $u = (u_{\underline{i}})_{\underline{i} \in I}$ une famille de réels indexée par I. On dit que $u$ est sommable lorsque la famille de réels positifs ou nuls $|u| = (|u_{\underline{i}}|)_{\underline{i} \in I}$ est sommable.*

*Soit $u = (u_{\underline{i}})_{\underline{i} \in I}$ une famille de réels indexée par I. Pour tout $\underline{i} \in I$ , on pose*

$$u_{\underline{i}}^{+} = \max\left(u_{\underline{i}}, 0\right) \qquad \textit{et} \qquad u_{\underline{i}}^{-} = \max\left(-u_{\underline{i}}, 0\right).$$

*Ceci définit deux familles $u^+ = (u_{\underline{i}}^+)_{\underline{i} \in I}$ et $u^- = (u_{\underline{i}}^-)_{\underline{i} \in I}$ de réels positifs ou nuls.*

**10.** Soit $u = (u_{\underline{i}})_{\underline{i} \in I}$ une famille de réels indexée par $I$. Montrer que la famille $u$ est sommable si et seulement si les familles $u^+$ et $u^-$ sont sommables.

*Pour $(u_{\underline{i}})_{\underline{i} \in I}$ une famille sommable de réels, on définit alors sa somme par*

$$\sum_{\underline{i} \in I} u_{\underline{i}} = \left(\sum_{\underline{i} \in I} u_{\underline{i}}^{+}\right) - \left(\sum_{\underline{i} \in I} u_{\underline{i}}^{-}\right).$$

**11.** Soient $u = (u_{\underline{i}})_{\underline{i} \in I}$ et $v = (v_{\underline{i}})_{\underline{i} \in I}$ deux familles sommables de réels.

**(a)** Montrer que la famille $u + v$ définie par

$$(u+v)_{\underline{i}} = u_{\underline{i}} + v_{\underline{i}}$$

est sommable et

$$\sum_{\underline{i} \in I} (u_{\underline{i}} + v_{\underline{i}}) = \left(\sum_{\underline{i} \in I} u_{\underline{i}}\right) + \left(\sum_{\underline{i} \in I} v_{\underline{i}}\right).$$

**(b)** Soient $a$ et $b$ deux réels. Montrer que la famille $au+bv$ définie par

$$(au+bv)_{\underline{i}} = au_{\underline{i}} + bv_{\underline{i}}$$

est sommable et déterminer sa somme en fonction des sommes de $u$ et $v$.

**12.** On considère $(I_k)_{k\in\mathbb{N}}$ une famille de sous-ensembles de $I$ tels que

▷ $\displaystyle\bigcup_{k\in\mathbb{N}} I_k = I$.

▷ Pour tout $k, l \in \mathbb{N}$, distincts, $I_k \cap I_l = \emptyset$.

Par convention, si $I_k = \emptyset$, on pose

$$\sum_{\underline{i}\in I_k} u_{\underline{i}} = 0.$$

Montrer que $(u_{\underline{i}})_{\underline{i}\in I_k}$ est sommable pour tout $k \in \mathbb{N}$ et la série de terme général $\displaystyle\sum_{\underline{i}\in I_k} |u_{\underline{i}}|$ converge si et seulement si $(u_{\underline{i}})_{\underline{i}\in I}$ est sommable. De plus dans ce cas, vérifier que

$$\sum_{\underline{i}\in I} u_{\underline{i}} = \sum_{k=0}^{+\infty} \sum_{\underline{i}\in I_k} u_{\underline{i}}.$$

***Application : Théorème du transfert***.

**13.** Soit $X_1, X_2, \ldots, X_k$ des variables aléatoires réelles discrètes définies sur $\Omega$ et $\varphi$ une application définie sur $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ telle que $T = \varphi(X_1, X_2, \ldots, X_k)$ soit encore une variable aléatoire réelle discrète définie sur $\Omega$. Pour tout $i \in \{1, \ldots, k\}$, l'univers-image de $X_i$ est noté

$$X_i(\Omega) = \{x_{i,j} \mid j \in I_i\},$$

où $I_i$ est une partie de $\mathbb{N}$. On pose $I = I_1 \times \ldots \times I_n$.

Montrer que les deux assertions $(A_1)$ et $(A_2)$ suivantes sont équivalentes :

$(A_1)$ $T$ admet une espérance.

$(A_2)$ La famille $\displaystyle\left(\varphi(x_{1,i_1}, x_{2,i_2}, \ldots, x_{n,i_n})\mathbf{P}\left(\bigcap_{k=1}^{n}[X_k = x_{k,i_k}]\right)\right)_{\underline{i}\in I}$ est sommable.

De plus dans ce cas, vérifier que

$$\mathbf{E}(T) = \sum_{(i_1 i_2,\ldots,x_k)\in I} \varphi(x_{1,i_1}, x_{2,i_2}, \ldots, x_{n,i_n})\mathbf{P}\left(\bigcap_{k=1}^{n}[X_k = x_{k,i_k}]\right).$$

# III. Estimateurs.

*Soit $I$ un intervalle ouvert, non vide, inclus dans $]0, +\infty[$.*

*On considère une variable aléatoire $X$ définie sur $\Omega$ suivant une loi de Poisson dépendante d'un paramètre réel $\theta \in I$, inconnu : autrement dit, $X(\Omega) = \mathbb{N}$ et, pour tout entier naturel $k$,*

$$\mathbf{P}([X = k]) = \frac{\theta^k}{k!}e^{-\theta}.$$



**Tournez la page S.V.P.**

*On cherche à approcher la valeur de ce paramètre $\theta$, ou éventuellement la valeur de son image $g(\theta)$ par une fonction réelle $g$ définie sur I.*

**14. (a)** À l'aide des résultats de la première partie, montrer que $X$ admet des moments d'ordre $r$ pour tout $r$ entier naturel non nul (c'est-à-dire que $X^r$ admet une espérance pour tout $r$ entier naturel non nul) puis que l'espérance et la variance de $X$ sont égales à $\theta$.

**(b)** Donner une interprétation combinatoire des moments d'ordre $r$ de $X$ lorsque $\theta = 1$.

**15.** Soit $n$ variables aléatoires $Y_1, Y_2, \ldots, Y_n$ définies sur $\Omega$, mutuellement indépendantes, suivant des lois de Poisson de paramètre respectif $\theta_1, \ldots, \theta_n$. Montrer que la variable aléatoire $\sum_{i=1}^{n} Y_i$ suit la loi de Poisson de paramètre $\sum_{i=1}^{n} \theta_i$. *Indication* : on pourra démontrer d'abord le cas $n = 2$.

*Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires réelles définies sur $\Omega$ , mutuellement indépendantes, et de même loi que X. Soit $\varphi$ une application définie sur $\mathbb{R}^n$ et à valeurs dans $\mathbb{R}$ telle que $\varphi(X_1, X_2, \ldots, X_n)$ soit encore une variable aléatoire définie sur $\Omega$.*

- ▷ *Le n-uplet $(X_1, X_2, \ldots, X_n)$ est appelé un n-échantillon de la loi de X et la variable aléatoire $T_n = \varphi(X_1, X_2, \ldots, X_n)$ est appelée un estimateur de $g(\theta)$.*
- ▷ *Après une réalisation de l'expérience aléatoire associée à $\Omega$, on note $x_1, x_2, \ldots, x_n$ les valeurs prises par $X_1, X_2, \ldots, X_n$ : ce sont les* observations de l'échantillon. *Le réel $t = \varphi(x_1, x_2, \ldots, x_n)$ est appelé une réalisation de l'estimateur $T_n$ de $g(\theta)$ et est choisie comme estimation ponctuelle de la valeur de $g(\theta)$.*

*Autrement dit, estimer ponctuellement $g(\theta)$, c'est décider d'accorder à $g(\theta)$ la valeur expérimentale $\varphi(x_1, x_2, \ldots, x_n)$.*

*Les outils suivants vont nous permettre d'étudier les estimateurs et de choisir le plus pertinent : soit $(X_1, X_2, \ldots, X_n)$ un n-échantillon de la loi de X et $T_n$ un estimateur de $g(\theta)$ admettant une espérance $\mathbf{E}(T_n)$, dépendant éventuellement de $\theta$.*

- ▷ *Le réel $\mathbf{E}(T_n) - g(\theta)$ est appelé* biais *de $T_n$ et est noté $b_\theta(T_n)$.*
- ▷ *L'estimateur $T_n$ est dit* sans biais *lorsque $b_\theta(T_n) = 0$, c'est-à-dire lorsque $\mathbf{E}(T_n) = g(\theta)$.*
- ▷ *Si l'estimateur $T_n$ admet une variance $\mathbf{V}(T_n)$, dépendant éventuellement de $\theta$, on appelle* risque quadratique *de $T_n$ le réel*

$$r_\theta(T_n) = \mathbf{E}\left((T_n - g(\theta))^2\right) = b_\theta(T_n)^2 + \mathbf{V}(T_n).$$

*En particulier si $T_n$ est sans biais, alors $r_\theta(T_n) = \mathbf{V}(T_n)$.*

*Une suite $(T_n)_{n\in\mathbb{N}^*}$ d'estimateurs de $g(\theta)$ admettant une espérance est dite* asymptotiquement sans biais *lorsque* $\lim_{n\to+\infty} b_\theta(T_n) = 0$.

*On commence par étudier deux estimateurs de $\theta$ : la moyenne et la variance empiriques. Soit $(X_1, X_2, \ldots, X_n)$ un n-échantillon de la loi de X. On définit la* moyenne empirique *$\overline{X_n}$ de $X_1, X_2, \ldots, X_n$ et la* variance empirique *$\overline{S_n}$ de $X_1, X_2, \ldots, X_n$ par*

$$\overline{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_k \qquad \textit{et} \qquad \overline{S_n} = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \overline{X_n}\right)^2.$$

**16.** Montrer que $\overline{X_n}$ est un estimateur sans biais de $\theta$ et que son risque quadratique est $\dfrac{\theta}{n}$.

**17. (a)** Montrer que
$$n\overline{S_n} = \sum_{k=1}^{n} X_k^2 - n\overline{X_n}^2.$$
En déduire $\mathbf{E}(\overline{S_n})$.

**(b)** Montrer que $\overline{S_n}$ est un estimateur asymptotiquement sans biais de $\theta$.

**(c)** Montrer que $\widehat{S_n} = \dfrac{n}{n-1}\overline{S_n}$ est un estimateur sans biais de $\theta$. Cet estimateur est appelé *variance empirique corrigée*.

**18. (a)** Montrer que
$$n\overline{S_n} = \left(\sum_{k=1}^{n}(X_k-\theta)^2\right) - n(\overline{X_n}-\theta)^2.$$

**(b)** En déduire que $n\overline{S_n}$ admet une variance donnée par la formule suivante :
$$\mathbf{V}\left(n\overline{S_n}\right) = n\mathbf{V}\left((X-\theta)^2\right) - 2n^2\mathbf{Cov}\left((X_1-\theta)^2, (\overline{X_n}-\theta)^2\right) + n^2\mathbf{V}\left((\overline{X_n}-\theta)^2\right).$$

**(c)** Montrer que
$$n^4\mathbf{E}\left((\overline{X_n}-\theta)^4\right) = n\mathbf{E}\left((X-\theta)^4\right) + 3n(n-1)\theta^2.$$
*Indication* : on remarquera que lorsqu'on développe $\left(\sum_{k=1}^{n}(X_k-\theta)\right)^4$, on obtient des termes de la forme $(X_i-\theta)^4$ ou $(X_i-\theta)^2(X_j-\theta)^2$ ou $(X_i-\theta)^3(X_j-\theta)$ ou $(X_i-\theta)^2(X_j-\theta)(X_k-\theta)$ ou $(X_i-\theta)(X_j-\theta)(X_k-\theta)(X_l-\theta)$ avec $i, j, k, l$ distincts deux à deux, dont on précisera l'espérance.

**(d)** En déduire que
$$n^2\mathbf{V}\left((\overline{X_n}-\theta)^2\right) = \frac{\mathbf{E}\left((X-\theta)^4\right) + (2n-3)\theta^2}{n}.$$

**(e)** Montrer que
$$n^2(\overline{X_n}-\theta)^2 = \sum_{k=1}^{n}(X_k-\theta)^2 + 2\sum_{1\leqslant i<j\leqslant n}(X_i-\theta)(X_j-\theta).$$

**(f)** En déduire que
$$n^2\mathbf{E}\left((X_1-\theta)^2(\overline{X_n}-\theta)^2\right) = \mathbf{E}\left((X-\theta)^4\right) + (n-1)\theta^2,$$
puis que
$$\mathbf{Cov}\left((X_1-\theta)^2, (\overline{X_n}-\theta)^2\right) = \frac{\mathbf{E}\left((X-\theta)^4\right) - \theta^2}{n^2}.$$

**(g)** Déduire des questions précédentes que
$$\mathbf{V}\left(\widehat{S_n}\right) = \frac{1}{n}\mathbf{E}\left((X-\theta)^4\right) - \frac{n-3}{n(n-1)}\theta^2.$$

# IV. Inégalité de Cramer-Rao.

*Le meilleur estimateur possible est un estimateur sans biais de variance minimale. Pour étudier cet extremum, nous allons établir l'inégalité de Cramer-Rao.*

*On se place toujours dans* $(\Omega, \mathcal{A}, \mathbf{P})$ *un espace probabilisé et on considère de nouveau* $X$ *une variable aléatoire définie sur* $\Omega$ *suivant une loi de Poisson dépendante d'un paramètre réel* $\theta \in I$ *inconnu avec* $I$ *intervalle non vide ouvert inclus dans* $]0, +\infty[$. *On considère les fonctions suivantes :*

$$f : \left\{ \begin{array}{ccc} I \times \mathbb{N} & \longrightarrow & \mathbb{R} \\ (\theta, k) & \longmapsto & \mathbf{P}([X = k]) = \dfrac{\theta^k}{k!} e^{-\theta}. \end{array} \right.$$

$$g : \left\{ \begin{array}{ccc} I \times \mathbb{N} & \longrightarrow & \mathbb{R} \\ (\theta, k) & \longmapsto & \ln\left(f(\theta, k)\right). \end{array} \right.$$

*On définit sous réserve d'existence* l'information de Fisher *de* $X$ *par*

$$F_X(\theta) = \mathbf{V}\left(\frac{\partial g}{\partial \theta}(\theta, X)\right).$$

**19.** Montrer que pour tout $\theta \in I$,

$$\mathbf{E}\left(\frac{\partial g}{\partial \theta}(\theta, X)\right) = 0,$$

puis que

$$F_X(\theta) = \frac{1}{\theta}.$$

*Soit* $\varphi$ *une fonction réelle telle que* $\varphi(X)$ *soit une variable aléatoire définie sur* $\Omega$*, admettant une espérance* $\mathbf{E}(\varphi(X))$ *et une variance* $\mathbf{V}(\varphi(X))$ *pour tout* $\theta \in I$.
*L'inégalité de Cramer-Rao affirme que*

$$\mathbf{V}(\varphi(X)) \geqslant \frac{\left(\dfrac{\partial}{\partial \theta}\mathbf{E}\left(\varphi(X)\right)\right)^2}{F_X(\theta)}.$$

*Nous allons démontrer cette inégalité.*

**20.** **(a)** Montrer que la fonction $\psi : \theta \longmapsto \sum_{k=0}^{+\infty} \varphi(k) f(\theta, k)$ est définie et dérivable sur $I$ et que pour tout $\theta \in I$,

$$\psi'(\theta) = \sum_{k=0}^{+\infty} \varphi(k) \frac{\partial f}{\partial \theta}(\theta, k).$$

**(b)** En déduire que

$$\frac{\partial}{\partial \theta}\mathbf{E}(\varphi(X)) = \mathbf{E}\left(\varphi(X)\frac{\partial g}{\partial \theta}(\theta, X)\right).$$

**21.** On fixe temporairement $\theta \in I$ et on considère la fonction $Q_\theta$ suivante dépendant d'une variable réelle $t$ :

$$Q_\theta(t) = \mathbf{E}\left(\left(\varphi(X) - \mathbf{E}(\varphi(X)) + t\frac{\partial g}{\partial \theta}(\theta, X)\right)^2\right).$$

**(a)** Montrer que $Q_\theta$ est un polynôme de degré 2 défini sur $\mathbb{R}$.

**(b)** Étudier le discriminant de $Q_\theta$.

**(c)** En déduire l'inégalité de Cramer-Rao.

*Nous allons généraliser cette inégalité à nos estimateurs.*
*Soit* $(X_1, X_2, \ldots, X_n)$ *un* $n$*-échantillon de la loi de* $X$. *On appelle* fonction de vraisemblance *pour les observations* $(k_1, k_2, \ldots, k_n)$ *de l'échantillon la fonction du paramètre* $\theta \in I$ *définie de la façon suivante :*

$$\mathcal{L}(\theta, k_1, k_2, \ldots, k_n) = \mathbf{P}\left(\bigcap_{j=1}^{n} [X_j = k_j]\right).$$

*On définit, sous réserve d'existence,* l'information de Fisher *de l'échantillon* $(X_1, X_2, \ldots, X_n)$ *par*

$$F(\theta) = V\left(\frac{\partial \ln(\mathcal{L})}{\partial \theta}(\theta, X_1, X_2, \ldots, X_n)\right).$$

**22.** Montrer que pour tout $\theta \in I$,

$$\mathbf{E}\left(\frac{\partial \ln(\mathcal{L})}{\partial \theta}(\theta, X_1, X_2, \ldots, X_n)\right) = 0,$$

puis que

$$F(\theta) = nF_X(\theta).$$

*Soit* $T_n = \varphi(X_1, X_2, \ldots, X_n)$ *un estimateur de* $g(\theta)$ *admettant une espérance pour tout* $\theta \in I$ *avec* $g$ *dérivable sur* $I$. *On dit que* $T_n$ *est* régulier *lorsque les deux conditions suivantes sont vérifiées :*

▷ *Pour tout* $\theta \in I$, *la famille*

$$\left(\varphi(k_1, k_2, \ldots, k_n)\frac{\partial \mathcal{L}}{\partial \theta}(\theta, k_1, k_2, \ldots, k_n)\right)_{(k_1, k_2, \ldots, k_n) \in \mathbb{N}^n}$$

*est sommable.*

▷ *La relation suivante est vérifiée :*

$$\frac{\partial}{\partial \theta}\left(\sum_{(k_1, k_2, \ldots, k_n) \in \mathbb{N}^n} \varphi(k_1, k_2, \ldots, k_n)\mathcal{L}(\theta, k_1, k_2, \ldots, k_n)\right)$$
$$= \sum_{(k_1, k_2, \ldots, k_n) \in \mathbb{N}^n} \varphi(k_1, k_2, \ldots, k_n)\frac{\partial \mathcal{L}}{\partial \theta}(\theta, k_1, k_2, \ldots, k_n).$$

**23. (a)** Montrer que tout estimateur $T_n$ de $g(\theta)$ admettant une espérance est régulier.

**(b)** En déduire que pour tout $\theta \in I$,

$$\frac{\partial}{\partial \theta}\mathbf{E}(T_n) = \mathbf{E}\left(T_n \frac{\partial \ln(\mathcal{L})}{\partial \theta}(\theta, X_1, X_2, \ldots, X_n)\right) = \mathbf{E}\left((T_n - g(\theta))\frac{\partial \ln(\mathcal{L})}{\partial \theta}(\theta, X_1, X_2, \ldots, X_n)\right).$$

**24.** Montrer que si $T_n$ est un estimateur sans biais de $g(\theta)$ admettant un moment d'ordre 2 pour tout $\theta \in I$, alors

$$\mathbf{V}(T_n) \geqslant \frac{g'(\theta)^2}{nF_X(\theta)}.$$

*Un estimateur sans biais de $g(\theta)$ admettant un moment d'ordre 2 est dit* efficace *s'il atteint la borne de Cramer-Rao, c'est-à-dire si*

$$\mathbf{V}(T_n) = \frac{g'(\theta)^2}{nF_X(\theta)}.$$

**25.** Montrer que $\overline{X_n}$ est un estimateur efficace de $\theta$.

**26.** Nous allons maintenant chercher les estimateurs réguliers sans biais efficaces de $\theta$.

**(a)** Soit $T_n$ un estimateur efficace et sans biais de $\theta$ admettant un moment d'ordre 2. Montrer que $T_n$ et $\overline{X_n}$ sont presque sûrement liées par une relation affine.

**(b)** Montrer qu'il existe deux estimateurs de $\theta$ sans biais et efficaces et admettant un moment d'ordre 2, dont l'un est $\overline{X_n}$. Que peut-on penser du second pour les applications pratiques ?

FIN DU SUJET