DS8 - Concours blanc

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.
Les candidat·es sont invité·es à ***encadrer*** *dans la mesure du possible les résultats de leurs calculs.*
Aucun document n'est autorisé. ***L'utilisation de toute calculatrice et de tout matériel électronique*** *est* ***interdite****. Seule l'utilisation d'une règle graduée est autorisée.*
Si au cours de l'épreuve, un candidat ou une candidate repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre.

On suppose, et c'est valable pour toute l'épreuve, que les librairies suivantes sont importées sous leur alias habituel :

- `import numpy as np`
- `import numpy.linalg as al`
- `import numpy.random as rd`
- `import matplotlib.pyplot as plt`
- `import pandas as pd`

Notations

- Tout au long du sujet $(\Omega, \mathcal{F}, \mathbb{P})$ désignera un espace probabilisé et les variables aléatoires utilisées seront toutes définies sur cet espace probabilisé. Sous réserve d'existence, l'espérance mathématique d'une variable aléatoire réelle X sera notée $\mathbb{E}(X)$ et sa variance sera notée $\mathbb{V}(X)$.
- Pour un événement A, on notera $\mathbb{P}_B(A)$ la probabilité conditionnelle de A sachant B où B est un événement non négligeable.

Si T est un estimateur de θ (respectivement de $g(\theta)$), on définit :

- le biais de T par
$$b_\theta(T) = \mathbb{E}(T) - \theta \qquad \big(\text{respectivement,} \qquad b_{g(\theta)}(T) = \mathbb{E}(T) - g(\theta)\big)$$
sous réserve de l'existence de $\mathbb{E}(T)$,
- et le risque quadratique de T par
$$r_\theta(T) = \mathbb{E}\left((T-\theta)^2\right) \qquad \big(\text{respectivement,} \qquad r_{g(\theta)}(T) = \mathbb{E}\left((T-g(\theta))^2\right)\big)$$
sous réserve de l'existence de $\mathbb{V}(T)$.

On dit que T est *sans biais* lorsque son biais est nul.
On considérera qu'un estimateur T est meilleur (au sens large) qu'un autre estimateur T' en termes de risque quadratique si le risque quadratique de T est inférieur ou égal à celui de T'.

Le sujet est composé de quatre parties. Les parties I, II, III et IV.1 sont **indépendantes**.
Il s'agit de variations autour de la notion de risque quadratique en théorie de l'estimation.

I. Premier problème d'estimation

Dans ce premier problème d'estimation, on dispose d'une seule observation notée X.
On suppose que X admet pour densité la fonction f_θ définie sur $\mathbb{R}$ par :
$$f_\theta : x \mapsto \begin{cases} \dfrac{k+1}{\theta^{k+1}}\, x^k & \text{si } x \in [0,\theta] \\ 0 & \text{sinon} \end{cases}$$
où k est un entier naturel non nul et θ un paramètre réel inconnu strictement positif que l'on souhaite estimer.

1. Montrer que f_θ est bien une densité de probabilité.

2. Calculer $\mathbb{E}(X)$.

3. Déterminer λ_0 un réel dépendant uniquement de k tel que $\lambda_0 X$ soit un estimateur de θ sans biais.

4. Calculer $\mathbb{V}(X)$.

5. Démontrer que pour tout T estimateur de θ admettant une variance :
$$r_\theta(T) = \big(\mathbb{E}(T) - \theta\big)^2 + \mathbb{V}(T)$$

6. Donner la valeur de $r_\theta(\lambda_0 X)$.
Le but de la fin de cette partie I est de déterminer un estimateur de θ ayant un plus petit risque quadratique que celui de $\lambda_0 X$.

7. En utilisant I.5 montrer que pour tout λ réel
$$r_\theta(\lambda X) = \theta^2\, Q(\lambda)$$
où Q est un polynôme de degré 2 dont les coefficients ne dépendent que de k.

8. Montrer que la fonction $\lambda \mapsto Q(\lambda)$ atteint son minimum en un unique réel noté λ^* que l'on exprimera en fonction de k.

9. Conclure sur le but recherché.

II. Second problème d'estimation

Dans ce second problème d'estimation, on dispose de n observations indépendantes $(n \geqslant 2)$ notées $X_1, \ldots, X_n$ de même loi de Poisson de paramètre θ inconnu ($\theta \in]0, +\infty[$). On souhaite estimer le paramètre $\exp(-\theta)$. On définit pour tout i élément de $[\![1, n]\!]$ la variable aléatoire Y_i par :

$$Y_i \ : \ \omega \ \mapsto \ \begin{cases} 1 & \text{si } X_i(\omega) = 0 \\ 0 & \text{sinon} \end{cases}$$

Puis on note :

$$\overline{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

10. Pour tout i élément de $[\![1, n]\!]$, donner la loi de Y_i.

11. Donner la loi de $\sum_{i=1}^{n} Y_i$, puis montrer que $\mathbb{E}\big(\overline{Y}_n\big) = \exp(-\theta)$.

On dira dans ce cas que $\overline{Y}_n$ est un estimateur sans biais de $\exp(-\theta)$.

12. Calculer $\mathbb{V}\left(\overline{Y}_n\right)$.

Pour tout k élément de $[\![1, n]\!]$ on définit $S_k = \sum_{i=1}^{k} X_i$.

13. Rappeler sans démonstration la loi de S_k pour tout k élément de $[\![1, n]\!]$.
On définit jusqu'à la fin de cette partie II pour tout j entier naturel :

$$\varphi(j) = \mathbb{P}_{[S_n = j]}\big([X_1 = 0] \big)$$

14. Montrer que pour tout j entier naturel :

$$\varphi(j) = \left(1 - \frac{1}{n}\right)^j$$

On a donc $\varphi(j)$ indépendant du paramètre θ inconnu.
D'après la question II.14, on peut définir l'estimateur :

$$\varphi(S_n) = \left(1 - \frac{1}{n}\right)^{S_n}$$

15. Montrer que $\varphi(S_n)$ admet une espérance et que $\mathbb{E}\big(\varphi(S_n)\big) = \exp(-\theta)$.

On dira dans ce cas que $\varphi(S_n)$ est un estimateur sans biais de $\exp(-\theta)$.

16. Montrer que $\varphi(S_n)$ admet une variance vérifiant :

$$\mathbb{V}\big(\varphi(S_n)\big) = \exp(-2\theta) \ \left(\exp\left(\frac{\theta}{n}\right) - 1\right)$$

17. On souhaite comparer les performances de $\overline{Y}_n$ et $\varphi(S_n)$ en tant qu'estimateurs de $\exp(-\theta)$.

a) Démontrer :

$$1 \leqslant \frac{\exp(\theta) - 1}{\theta} \leqslant \exp(\theta)$$

b) Soit la fonction $h : [0, 1] \to \mathbb{R}$ définie par :

$$h(t) = t \exp(\theta) + (1 - t) - \exp(t\theta)$$

pour tout $t \in [0, 1]$. Étudier les variations de h.

c) En déduire :

$$\exp\left(\frac{\theta}{n}\right) \leqslant \frac{\exp(\theta)}{n} + \frac{n-1}{n}$$

puis l'inégalité :

$$\mathbb{V}\big(\varphi(S_n)\big) \leqslant \mathbb{V}\big(\overline{Y}_n\big)$$

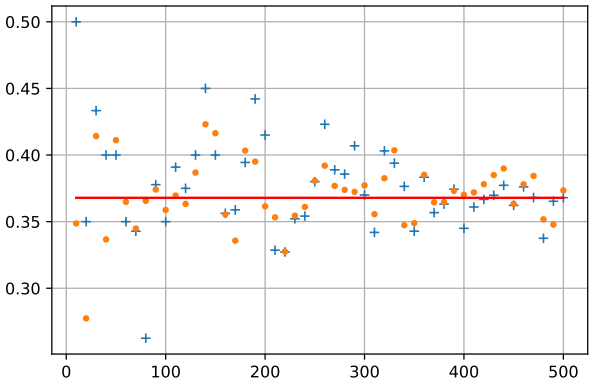
d) Comparer les risques quadratiques de $\overline{Y}_n$ et $\varphi(S_n)$ en tant qu'estimateurs de $\exp(-\theta)$.

18. a) Compléter la fonction **Python** suivante pour qu'elle simule les estimateurs $\overline{Y}_n$ et $\varphi(S_n)$.

```
def simulEstimateurs(theta, n):
    Xobs = ____________________________
    Y = np.zeros(n)
    for k in range(n):
        if _________________:
            Y[k] = 1
    Ybar = np.mean(Y)
    phi = ____________________________
    return [Ybar,phi]
```

b) On exécute le script **Python** ci-dessous. Commenter en lien avec la question ***17***.

```
YbarListe=[]; phiListe=[]; theta=1
nListe = [10*i for i in range(1, 51)]
for n in nListe:
  [y,p] = simulEstimateurs(theta, n)
  YbarListe.append(y)
  phiListe.append(p)
plt.plot(nListe, YbarListe, '+')
plt.plot(nListe, phiListe, '.')
plt.plot(nListe,
    [np.exp(-theta) for k in nListe])
plt.show()
```



On reprendra à la fin de la partie IV l'étude de $\varphi(S_n)$.

III. Information de Fisher

III.1. Cas discret

Dans cette section III.1, on considère I un intervalle de $\mathbb{R}$, θ un paramètre inconnu appartenant à I et X une variable aléatoire à valeurs dans $\mathbb{N}$ ($X(\Omega) \subset \mathbb{N}$). On suppose qu'il existe une fonction p définie sur $I \times X(\Omega)$ telle que pour tout k élément de $X(\Omega)$:

$$\mathbb{P}\big([X=k]\big) = p(\theta,k)$$

et vérifiant, pour tout k de $X(\Omega)$, $\theta \mapsto p(\theta,k)$ est dérivable sur I.
On note de plus $h : (\theta,k) \mapsto \ln\big(p(\theta,k)\big)$.
On définit enfin, sous réserve d'existence l'**information de Fisher** de X par :

$$I_X(\theta) = \sum_{k\in X(\Omega)} \big(\partial_1(\ln\circ p)(\theta,k)\big)^2 p(\theta,k)$$

19. Dans cette question ***19***, on considère X une variable aléatoire qui suit la loi de Bernoulli de paramètre θ (où $\theta \in]0,1[$).
On a alors $X(\Omega) = \{0,1\}$, $\mathbb{P}\big([X=1]\big) = p(\theta,1) = \theta$, $\mathbb{P}\big([X=0]\big) = p(\theta,0) = 1-\theta$ et :

$$I_X(\theta) = \big(\partial_1(h)(\theta,1)\big)^2 p(\theta,1) + \big(\partial_1(h)(\theta,0)\big)^2 p(\theta,0)$$

Montrer :

$$I_X(\theta) = \frac{1}{\theta\,(1-\theta)}$$

20. Dans cette question ***20***, on considère X une variable aléatoire qui suit la loi binomiale de paramètres N et θ ($N \in \mathbb{N}^*$, $\theta \in]0,1[$).

a) Montrer :

$$I_X(\theta) = \frac{1}{\big(\theta\,(1-\theta)\big)^2} \sum_{k=0}^{N} (k - N\theta)^2 \binom{N}{k} \theta^k (1-\theta)^{N-k}$$

b) En déduire :

$$I_X(\theta) = \frac{\mathbb{V}(X)}{\big(\theta\,(1-\theta)\big)^2}$$

puis donner la valeur de $I_X(\theta)$.

21. Dans cette question ***21***, on considère X une variable aléatoire qui suit la loi de Poisson de paramètre θ ($\theta \in]0,+\infty[$). Puisque $X(\Omega) = \mathbb{N}$, on a sous réserve de convergence :

$$I_X(\theta) = \sum_{k=0}^{+\infty} \big(\partial_1(h)(\theta,k)\big)^2 p(\theta,k)$$

a) Montrer que la série de terme général $\big(\partial_1(h)(\theta,k)\big)^2 p(\theta,k)$ converge et calculer sa somme $I_X(\theta)$.

b) Justifier :

$$I_X(\theta) = \mathbb{E}\left(\big(\partial_1(h)(\theta,X)\big)^2\right)$$

III.2. Cas d'une variable gaussienne

Soit X une variable aléatoire qui suit la loi normale de moyenne θ ($\theta \in \mathbb{R}$) et de variance 1 dont la densité est notée $x \mapsto f(\theta,x)$. On définit sous réserve de convergence l'**information de Fisher** de X par :

$$I_X(\theta) = \int_{-\infty}^{+\infty} \big(\partial_1(\ln \circ f)(\theta,x)\big)^2 f(\theta,x)\; dx$$

22. Montrer que sous réserve de convergence :

$$I_X(\theta) = \int_{-\infty}^{+\infty} (x-\theta)^2 f(\theta,x)\; dx$$

23. En déduire l'existence et la valeur de $I_X(\theta)$.

24. Justifier :

$$I_X(\theta) = \mathbb{E}\left(\big(\partial_1(\ln \circ f)(\theta,X)\big)^2\right)$$

IV. Minoration du risque quadratique

IV.1. Inégalité de Cramer-Rao

Dans cette section IV.1, on considère I un intervalle de $\mathbb{R}$, θ un paramètre inconnu appartenant à I et X une variable aléatoire telle que $X(\Omega) = [\![0, N]\!]$ ($N \in \mathbb{N}$). On suppose qu'il existe une fonction p définie sur $I \times X(\Omega)$ telle que pour tout $k \in [\![0, N]\!]$:

$$\mathbb{P}\big([X = k]\big) = p(\theta, k)$$

et vérifiant :

- pour tout $k \in [\![0, N]\!]$, $\theta \mapsto p(\theta, k)$ est dérivable sur I,
- l'information de Fisher de X notée $I_X(\theta)$ définie dans la partie III est non nulle pour tout $\theta \in I$.

Le but de la section IV.1 est de démontrer l'inégalité suivante dûe à Cramer et Rao.

Théorème 1. *(de Cramer-Rao)*
Soit $f(X)$ un estimateur sans biais de $g(\theta)$ à savoir tel que $\mathbb{E}\big(f(X)\big) = g(\theta)$ où g est dérivable sur I. On a alors :

$$\mathbb{V}\left(f(X)\right) \geqslant \frac{\big(g'(\theta)\big)^2}{I_X(\theta)}$$

25. Montrer que pour tout θ élément de I :

$$\sum_{k=0}^{N} \partial_1(p)(\theta, k) = 0$$

26. En déduire que pour tout θ élément de I :

$$\mathbb{E}\left(\partial_1(h)(\theta, X)\right) = 0 \qquad (E)$$

27. En dérivant partiellement par rapport à θ les deux membres de l'égalité (E), montrer que pour tout θ élément de I :

$$\mathbb{E}\left(\partial_{1,1}^2(h)(\theta, X)\right) = -\mathbb{E}\left(\big(\partial_1(h)(\theta, X)\big)^2\right)$$

28. Montrer que pour tout θ élément de I :

$$g'(\theta) = \sum_{k=0}^{N} f(k)\ \big(\partial_1(h)(\theta, k)\big)\ p(\theta, k)$$

puis :

$$g'(\theta) = \mathbb{E}\left(\big(f(X) - g(\theta)\big)\big(\partial_1(h)(\theta, X)\big)\right)$$

29. On pose pour tout t réel :

$$L(t) = \mathbb{E}\left(\big(\big(f(X) - g(\theta)\big) + t\big(\partial_1(h)(\theta, X)\big)\big)^2\right)$$

a) Développer le polynôme L suivant les puissances décroissantes de t.

b) Calculer le discriminant Δ de L et justifier : $\Delta \leqslant 0$.

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

IV.2. Extension du théorème de Cramer-Rao

On reprend dans cette section IV.2 les notations et hypothèses de la partie II. On admet que, dans ce contexte, le théorème de Cramer-Rao peut se généraliser comme suit :

Théorème 2. *(de Cramer-Rao)*
Soit $T_n = f(X_1, \dots, X_n)$ un estimateur sans biais de $g(\theta)$ à savoir tel que $\mathbb{E}\big(f(X_1, \dots, X_n)\big) = g(\theta)$ où g est dérivable sur $]0, +\infty[$. On a alors :

$$\mathbb{V}(T_n) \geqslant \frac{\big(g'(\theta)\big)^2}{n\, I_{X_1}(\theta)}$$

où $I_{X_1}(\theta)$ est l'information de Fisher d'une variable aléatoire de loi de Poisson de paramètre θ définie et calculée à la partie III.

Il s'agit dans cette section d'exploiter cette nouvelle inégalité de Cramer-Rao. On note :

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

30. Calculer $\mathbb{E}\big(\overline{X}_n\big)$ et $\mathbb{V}\big(\overline{X}_n\big)$.

31. Déduire de la généralisation de Cramer-Rao, que $\overline{X}_n$ a le plus petit risque quadratique parmi les estimateurs sans biais de θ.

32. Montrer que pour $g(\theta) = \exp(-\theta)$ où $\theta \in]0, +\infty[$:

$$\mathbb{V}\big(\varphi(S_n)\big) \underset{n\to+\infty}{\sim} \frac{\big(g'(\theta)\big)^2}{n\, I_{X_1}(\theta)}$$

33. Que prouve ce résultat en terme d'optimalité de $\varphi(S_n)$ dans l'estimation de $\exp(-\theta)$?

34. À la lumière de la partie II, peut-on conclure que lorsque n est grand $\varphi(S_n)$ est le meilleur estimateur de $\exp(-\theta)$ en terme de risque quadratique ?