Dauphine | PSL
UNIVERSITÉ PARIS

MIDO - L3 Math. Appliquées 2021–2022

Statistical modelling

Examen final du 14 février 2022

Durée 2h00 – Documents et Calculatrice Non Autorisés

Ex. 1	Ex. 2	Total
...... / 8.5	...... / 11.5	...... / 20

NOM : ....

PRÉNOM : ....

(lisiblement)

PARTIE À RABATTRE

Important. Suivant les réglements en vigueur,

1. les enseignants présents lors de l'épreuve ne peuvent communiquer que sur les fautes d'énoncé potentielles. Toute autre question durant la composition ne sera pas acceptée.
2. les étudiants sont tenus de se lever au moment de l'annonce de fin de la composition. En cas de refus, le responsable de l'UE sera fondé à ne pas prendre en compte la copie incriminée.
3. l'identification de la copie de composition doit se faire au moment de la remise de la copie par les enseignants et surveillants. Il ne sera pas accordé de délai pour cette raison en fin d'épreuve.

Les exercices sont indépendants. Toutes les réponses sont à fournir sur la copie d'énoncé. L'espace blanc alloué à chaque question est amplement suffisant pour apporter une réponse correcte.

Sauf mention contraire, la mesure de référence est la mesure de Lebesgue.

Formulaire

Loi	Notation	Densité
Exponentielle	$\mathcal{E}(\lambda)$	$f(x \mid \lambda) = \lambda e^{-\lambda x} \mathbb{1}_{x>0}$
Gamma	$\mathcal{G}\mathrm{a}(a,b)$	$f(x \mid a,b) = \dfrac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \mathbb{1}_{x>0}$
Laplace	$\mathcal{L}(\mu,b)$	$f(x \mid \mu,b) = \dfrac{1}{2b} \exp\left[-\dfrac{\lvert x-\mu\rvert}{b}\right]$
Poisson	$\mathcal{P}oi(\lambda)$	$f(x \mid \lambda) = \dfrac{e^{-\lambda}}{x!} \lambda^x \mathbb{1}_{x\in\mathbb{N}}$

French – English Lexicon

- échantillon : *sample*
- famille exponentielle : *exponential family*
- espace naturel des paramètres : *natural parameter space*
- *i.i.d.* : *independent and identically distributed*
- statistique libre : *ancillary statistic*
- statistique exhaustive : *sufficient statistic*
- statistique complète : *complete statistic*
- vraisemblance : *likelihood*

Statistiques d'ordre Étant données $X_1, \ldots, X_n$ *i.i.d.* de densité f et de fonction de répartition F, les densités de $\min(X_1, \ldots, X_n)$ et $\max(X_1, \ldots, X_n)$ sont respectivement

$$f_1(x) = n[1-F(x)]^{n-1} f(x) \quad \text{et} \quad f_n(x) = nF(x)^{n-1} f(x).$$

PARTIE À RABATTRE

Exercice 1

...... /**8.5**

Dans cet exercice il vous est demandé de donner la ou les bonnes réponses. Seules les réponses justifiées seront validées, au prorata du nombre de bonnes réponses. Il n'y a pas de points négatifs, mais toute réponse fausse conduit à une note nulle.

1. Soit la densité définie sur $\mathbb{R}$ par $f(x \mid \alpha, \theta) = \alpha\theta^{-\alpha}x^{\alpha-1}$ pour $0 \leq x \leq \theta$. Alors $f(\cdot \mid \alpha, \theta)$:

(a) forme une famille exponentielle de dimension 2,
(b) forme une famille exponentielle de dimension 1,
(c) ne forme pas une famille exponentielle,
(d) forme une famille exponentielle uniquement lorsque α est fixé.

...... /**0.5**

2. Soit un jeu de données `x` correspondant à des observation d'un modèle statistique de fonction de répartition F. On dispose d'un échantillon bootstrap `mstar` de la médiane de $(X_1^\star, \ldots, X_n^\star)$ pour $X_1^\star, \ldots, X_n^\star$ *i.i.d.* suivant la fonction de répartition empirique $\widehat{F}_n$. Quel code permet d'obtenir un intervalle de confiance bootstrap empirique (*empirical bootstrap confidence interval*) de la médiane de la loi F au niveau 95% ?

(a)
```
quantile(mstar, c(.05, .95))
```
(b)
```
median(x) - quantile(mstar - median(x), c(.95, .05))
```
(c)
```
quantile(mstar, c(.025, .975))
```
(d)
```
median(x) - quantile(mstar - median(x), c(.975, .025))
```

...... /**1**

3. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables *i.i.d.* de loi uniforme sur $[0,\theta]$, avec $\theta > 0$. Parmi les variables suivantes, donner celle(s) qui sui(ven)t asymptotiquement la loi $\mathcal{N}(0,1)$.

(a) $\sqrt{3n}\left(1-\theta/(2\overline{X}_n)\right)$,

(b) $2\sqrt{3n}\left(1-\theta/\overline{X}_n\right)$,

(c) $\sqrt{n}\left(1-(2\overline{X}_n)/\theta\right)$,

(d) $\frac{\sqrt{n}}{3}\left(1-(2\overline{X}_n)/\theta\right)$.

(e) Aucune

...... /**1**

4. Soit X qui suit la loi de Laplace $\mathcal{L}(\mu, b)$ où le paramètre de position $\mu \in \mathbb{R}$ est connu. L'information de Fisher apportée par $T(X) = |X - \mu|$ sur b est :

(a) μb^2, (b) $(2 - b^2)/b^4$, (c) $1/b^2$, (d) $(b - 2\mu)/b^3$.

...... $\big/$ **1**

5. Soient X_1, X_2, X_3, X_4 *i.i.d.* suivant la loi de Poisson $\mathcal{P}(\lambda)$, avec $\lambda \in \mathbb{R}_+^*$. On pose

$$T(X_1,\ldots,X_4) = \log\left(1+\sum_{i=1}^{4} X_i\right) \quad \text{et} \quad R(X_1,\ldots,X_4) = \log\left[\exp\left(-\frac{X_1+X_2}{2}\right)+\frac{1}{2}\exp\left(X_4-X_3\right)\right].$$

(a) $T(X_1,\ldots,X_4)$ est exhaustive,

(b) $T(X_1,\ldots,X_4)$ n'est pas exhaustive,

(c) $R(X_1,\ldots,X_4)$ est exhaustive,

(d) $R(X_1,\ldots,X_4)$ n'est pas exhaustive.

...... $/$**1**

6. Soient X_1, X_2, X_3, X_4 *i.i.d.* suivant la loi $\mathcal{N}(\mu, \sigma^2)$, avec $\mu \in \mathbb{R}$ et $\sigma^2 \in \mathbb{R}_+^*$ inconnus. On note $\mathrm{med}(X_1, \dots, X_4)$ la médiane de $X_1, \dots, X_4$. Parmi les statistiques suivantes, lesquelles sont libres ?

(a) X_1/X_2,

(b) $X_1 - X_2$,

(c) $(X_1 - X_2)/(X_3 - X_4)$,

(d) $\sum_{i=1}^4 (X_i - X_1)^2$,

(e) $\dfrac{4 \times \mathrm{med}(X_1, \dots, X_4) - \sum_{i=1}^4 X_i}{\max(X_1, \dots, X_4) - \min(X_1, \dots, X_4)}$.

...... /**1**

7. Soit X de densité $f(\cdot \mid \theta)$, $\theta \in \mathbb{R}$. Soient $T(X)$ une statistique exhaustive pour θ et $S(X)$ une statistique libre non constante telle que $S(X) = g(T(X))$, avec g une fonction mesurable. Alors nécessairement,

(a) $T(X)$ est complète,

(b) $T(X)$ est minimale exhaustive,

(c) $T(X)$ et $S(X)$ sont indépendantes,

(d) $S(X)$ est minimale exhaustive,

(e) $T(X)$ est libre,

(f) Aucune de ces propositions n'est vraie.

...... /**1**

8. Soit $X_1, \ldots, X_n$ variables aléatoires *i.i.d.* suivant la loi $\mathcal{E}(e^\theta)$, $\theta \in \mathbb{R}$. On pose pour $k \in [\![1, n]\!]$ et $\delta \in \mathbb{R}_+^*$ connu, $Y_k = \mathbb{1}_{\{X_k \leq \delta\}}$. On observe $(y_1, \ldots, y_{n-1}) = (1, \ldots, 1)$ et $y_n = 0$. L'estimateur du maximum de vraisemblance de θ est

(a) $\log(\delta)/(n-1)$,

(b) $-\log[(n-1)/n]/\delta$,

(c) $\log[\log(n)/\delta]$,

(d) $\log\{-\log[(n-1)/n]/\delta\}$.

...... /**1**

9. Soient $X_1, \ldots, X_n$ *i.i.d.* suivant la loi de densité $f(x \mid \theta) = \exp[-(x-\theta)]\mathbb{1}\{x \geq \theta\}$. Alors :

(a) $\min(X_1, \ldots, X_n)$ est l'estimateur du maximum de vraisemblance de θ et il est sans biais,

(b) $\min(X_1, \ldots, X_n) - \frac{1}{n}$ est un estimateur sans biais de θ,

(c) $\max(X_1, \ldots, X_n)$ est l'estimateur du maximum de vraisemblance de θ et il est sans biais,

(d) $\frac{n}{n-1} \min(X_1, \ldots, X_n)$ est un estimateur sans biais de θ.

...... $/$ **1**

Exercice 2

....../**11.5**

La loi de Maxwell de paramètre $b \in \mathbb{R}_+^*$ admet pour densité sur $\mathbb{R}_+^*$

$$f(x \mid b) = \sqrt{\frac{2}{\pi}} \frac{x^2}{b^3} \exp\left(-\frac{x^2}{2b^2}\right) \mathbb{1}_{\{x>0\}}.$$

Dans cet exercice, on suppose que **le paramètre** b **est inconnu**. On note pour $n \in \mathbb{N}$, $X_1, \ldots, X_n$ un ensemble de variables aléatoires *i.i.d.* suivant $f(\cdot \mid b)$

1. Montrer que $\{f(\cdot \mid b); b \in \mathbb{R}_+^*\}$ forme une famille exponentielle. La famille ainsi définie est-elle sous forme canonique ? Si non, donner la forme canonique. La famille est-elle minimale ? Régulière ?

....../**1.5**

2. Montrer qu'il existe un unique estimateur du maximum de vraisemblance de b donné par

$$\widehat{b}_n = \sqrt{\frac{1}{3n} \sum_{i=1}^{n} X_i^2}.$$

...... /**1**

3. Montrer que $\widehat{b}_n \xrightarrow[n\to\infty]{\mathbb{P}} b$.

...... /**1**

4. Calculer l'information de Fisher apportée par X_1 sur b et celle apportée par X_1 sur le paramètre de la forme canonique.

...... $/\mathbf{1}$

5. En déduire qu'il existe une suite $(c_n)_{n\in\mathbb{N}^*}$, dont l'expression est une fonction de $X_1, \dots, X_n$ uniquement, telle que

$$c_n\left(\widehat{b}_n - b\right) \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1).$$

...... /**1**

6. Montrer qu'il existe une fonction h bijective sur $\mathbb{R}_+^*$ telle que pour $\alpha \in]0,1[$,

$$\mathbb{P}\left[\sqrt{n}\left|h(\widehat{b}_n) - h(b)\right| \leq q_{1-\alpha/2}\right] \xrightarrow[n\to\infty]{} 1-\alpha,$$

où $q_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi normale $\mathcal{N}(0,1)$.

...... /**1**

7. Montrer que $\widehat{b}_n$ est une statistique complète et exhaustive pour b.

...... $/\mathbf{1}$

8. Montrer que pour tout $c \in \mathbb{R}_+^*$, cX_1 admet pour densité $f(\cdot \mid bc)$. En déduire que $(X_1/b)^2$ suit la loi gamma de paramètre $(3/2, 1/2)$ (*c.f.* Formulaire).

Indication. $2\Gamma(3/2) = \sqrt{\pi}$.

...... /**1**

9. On admet que pour tout $c \in \mathbb{R}_+^*$, $c\sum_{i=1}^n X_i^2$ suit la loi Gamma de paramètre $(3n/2, 1/(2cb^2))$. Calculer alors le biais de l'estimateur $\widehat{b}_n$ par rapport à b. En déduire un estimateur sans biais de b, noté $\widehat{\beta}_n$, s'exprimant comme une fonction mesurable de $\widehat{b}_n$.

...... /**1**

11. Montrer que $\widehat{\beta}_n$ est l'unique estimateur sans biais de b de variance minimale (UMVUE).

...... $/\mathbf{1}$

12. Montrer que pour toute fonction g linéaire par rapport à chacune de ses coordonnées, $g(X_1, \ldots, X_n)/\widehat{b}_n$ est une statistique indépendante de $\widehat{b}_n$.

...... $/\mathbf{1}$