Dauphine | PSL
UNIVERSITÉ PARIS

NOM : ....

PRÉNOM : ....

(lisiblement)

MIDO - L3 Math. Appliquées
2020–2021

Statistical modelling

Examen partiel du 29 octobre 2020 (durée 2h00)

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. La fin de la copie peut être utilisée pour un complément éventuel.

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 des copies et intercalaires doit se faire au moment de la remise de chaque copie par les enseignants et surveillants. Il ne sera pas accordé de délai pour cette raison en fin d'épreuve.

Exercise 1 (9.5 points, les questions peuvent être résolues indépendamment)

On réalise une étude de qualité sur une chaîne de production de composants électroniques utilisés dans le secteur de l'aéronautique. Pour cela, on compte le nombre de composants défectueux dans des lots pris au hasard. Les résultats de l'expérience sont sauvegardés dans un fichier `dataset.csv` dont les premières lignes sont données ci-dessous (la colonne `Defective` correspond aux nombres de composants défectueux du lot indiqué dans la colonne `Reference`) :

```
Reference,Defective
QcP-s-1120,2
QcP-s-9120,0
QcP-s-2220,3
```

1. (0.5 pt) Écrire le code R qui permet de charger, à partir de ce fichier, les données concernant le nombre de pièces défectueuses dans un vecteur `x`.

2. (0.5 pt) Parmi les codes ci-dessous, donner, en justifiant, celui qui permet de tracer la distribution de l'échantillon contenu dans `x`.

(a) `hist(table(x) / length(x), freq = FALSE)`

(b) `barplot(x)`

(c) `hist(x, freq = FALSE)`

(d) `barplot(table(x) / length(x))`

PARTIE
À
RABATTRE

On considère le modèle statistique à support $\{1, 2, 3, 4\}$

$$\mathcal{P} = \left\{ \mathbb{P}_\theta = \theta^2 \delta_1 + (1-\theta)^2 \delta_2 + \theta(1-\theta)(\delta_3 + \delta_4) \;\middle|\; \theta \in]0;1[\right\}.$$

3. (0.5 pt) Écrire le code `R` d'une fonction `rgen(`n`, `θ`)` permettant de simuler n réalisations indépendantes suivant $\mathbb{P}_\theta$.

On note $X_1, \ldots, X_n$ des variables aléatoires indépendantes et identiquements distribuée de loi $\mathbb{P}_\theta$. On considère les estimateurs

$$\hat{\delta}_n = \sqrt{\frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{X_k = 1\}}} \quad \text{et} \quad \hat{\theta}_n = g(X_1, \ldots, X_n), \quad \text{où la fonction } g \text{ est donnée par le code :}$$

```
g <- function(x) { return(0.5 * (1 + mean(x == 1) - mean(x == 2))) }
```

4. (0.5 pt) Déduire du code ci-dessus l'expression de $\hat{\theta}_n$.

5. (1.5 pt) Montrer que les estimateurs $\hat{\delta}_n$ et $\hat{\theta}_n$ convergent vers θ en probabilité.

6. (1 pt) Rappeler la loi de $\sum_{k=1}^{n} \mathbb{1}_{\{X_k=1\}}$ et donner le code `R` correspondant à une estimation bootstrap paramétrique du biais de $\hat{\delta}_n$. On notera k le nombre d'échantillons bootstrap utilisés.
Bonus (0.5 pt) : code sans boucle `for`.

7. (3 pts) Trouver une suite c_n indépendante de θ telle que

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

En déduire un intervalle de confiance asymptotique pour θ au niveau $1-\alpha$.

8. (2 pts) Trouver une fonction h inversible sur $]0;1[$ telle que

$$\sqrt{n}\left[h\left(\hat{\theta}_n\right) - h\left(\theta\right)\right] \xrightarrow[n\to+\infty]{d} \mathcal{N}(0,1).$$

En déduire un intervalle de confiance asymptotique pour θ au niveau $1-\alpha$.

Exercice 2 (5 points) *Dans cet exercice il vous est demandé de donner la ou les bonnes réponses, seules les réponses justifiées seront validées. Il n'y a pas de points négatifs.*

1. Si `x` est un échantillon continu de taille `n = 13` et si on étudie la variabilité de la médiane de l'échantillon par bootstrap, le nombre de valeurs possibles d'une réalisation bootstrap de cette médiane vaut

(a) `8.9161e+12` (b) `479001600` (c) `1352078` (d) `1` (e) `13`

2. Si X dénote une variable aléatoire de la loi de Poisson de paramètre 2, quelle commande R retourne $\mathbb{P}[X = 4]$?

(a) `ppois(2, 4)`
(b) `dpois(4,2)`
(c) `dpois(2,4)`
(d) `rpois(4,2)`
(e) `ppois(4,2)`
(f) `qpois(4,2)`

3. Pour montrer que $\lim_{n\to\infty} \mathbb{P}[|X_n - X| > \varepsilon] = 0$, à ε fixé, une condition minimale est que

(a) $X_n \to X$ en distribution
(b) $X_n \to X$ en probabilité
(c) $X_n \to X$ presque sûrement
(d) $\mathbb{E}[|X_n - X|]$ existe pour n assez grand

5. Pour construire une variable aléatoire X suivant une loi exponentielle $\mathcal{E}xp(\lambda)$, avec $\lambda > 0$, à partir d'une variable uniforme $\mathcal{U}(0, 1)$, U, on peut utiliser la transformation

(a) $X = \lambda \log(U)$
(b) $X = -\log[(1 - U)]^\lambda$
(c) $X = -\log(1 - U)/\lambda$
(d) $X = \log[\lambda(1 - U)]$
(e) $X = -\log(U/\lambda)$
(f) $X = \exp\{-\log(U)^\lambda\}$

4. La loi de Laplace a pour densité $f(x|\mu,\sigma) = \frac{1}{2\sigma}\exp(-|x-\mu|/\sigma)$, $\sigma > 0$. Quelle affirmation est exacte ? La loi de Laplace

(a) à μ connu forme une famille exponentielle de paramètre naturel $\tau(\sigma) = -\frac{1}{\sigma}$.

(b) forme une famille exponentielle de paramètre naturel $\tau(\mu,\sigma) = (-\mu,\sigma^2)$.

(c) forme une famille exponentielle de paramètre naturel $\tau(\mu,\sigma) = (-\mu/\sigma,\sigma^2)$.

(d) ne forme pas une famille exponentielle

Exercice 3 (11 points) La loi inverse gaussienne $IG(\mu, \lambda)$ est définie par la densité

$$f(x;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right] \mathbb{1}_{\{x\in\mathbb{R}_+^*\}}, \quad \text{avec} \quad \mu > 0 \quad \text{et} \quad \lambda > 0.$$

Dans le cas particulier où $\lambda = \mu^2$, on parle de variable inverse gaussienne à un seul paramètre $IG(\mu, \mu^2)$.

Formulaire : dans cet exercice, on pourra utiliser, sans la démontrer, l'identité suivante

$$\int_0^{+\infty} x^{-3/2} \exp\left[-\frac{1}{2}\left(ax + \frac{b}{x}\right)\right] \mathrm{d}x = 2\left(\frac{a}{b}\right)^{\frac{1}{4}} K(\sqrt{ab}), \quad \textit{avec, pour } x \in R_+^*, \quad K(x) = \sqrt{\frac{\pi}{2x}}\exp(-x).$$

1. (2 pts) Montrer que la loi $IG(\mu, \lambda)$ constitue une famille exponentielle de dimension deux. Donner un paramètre naturel, une statistique naturelle, et démontrer que la représentation est minimale et régulière.

2. (0.5 pt) Dans le cas d'une loi inverse gaussienne à un seul paramètre, donner la dimension de la famille exponentielle correspondante et une statistique naturelle.

3. (2 pt) Pour $t < \lambda/(2\mu^2)$, montrer que la fonction génératrice des moments $M_X(t)$ de la loi inverse gaussienne $IG(\mu, \lambda)$ vaut

$$\exp\left[{}^{\lambda}/_{\mu}\left\{1 - (1 - {}^{2t\mu^2}/_{\lambda})^{1/2}\right\}\right] .$$

4. (1 pt) En déduire que pour X de loi inverse gaussienne $IG(\mu, \lambda)$, on a $\mathbb{E}[X] = \mu$ et $\mathbb{E}(1/X) = {}^{1}/_{mu} + {}^{1}/_{\lambda}$.

5. (1 pt) Montrer que si X suit la loi $IG(\mu, \lambda)$, alors pour $\alpha > 0$, αX est de loi $IG(\alpha\mu, \alpha\lambda)$. En déduire qu'une variable inverse gaussienne de paramètres μ et λ peut être transformée en variable inverse gaussienne Z à un seul paramètre μ_0 à préciser en fonction de μ et λ.

6. (1 pt) Montrer que si $X_1, \ldots, X_n$ sont des variables aléatoires *i.i.d.* suivant $IG(\mu, \lambda)$, alors

$$S_n = \textstyle\sum_{i=1}^n X_i \sim IG(n\mu, n^2\lambda).$$

En déduire le comportement asymptotique de la moyenne empirique $\bar{X}_n$ quand n tend vers l'infini.

7.a (1.5 pt) Si $X \sim IG(\mu, \lambda)$, soit $Y = \min\{X, \mu^2/X\}$. Montrer que $Y \leq \mu$ avec probabilité 1 et donner la densité de la loi de Y. Quand μ est fixé, est ce que cette loi constitue une famille exponentielle ? Si oui, donner sa dimension et une statistique naturelle.

7.b (0.5 pt) Si $X \sim IG(\mu, \lambda)$, $Y = \min\{X, \mu^2/X\}$ et

$$Z = \frac{\lambda(X-\mu)^2}{\mu^2 X}$$

montrer que Z a la même loi que

$$W = \frac{\lambda(Y-\mu)^2}{\mu^2 Y}$$

(On ne demande pas de trouver cette loi.)

7.c (1.5 pt) Admettant que, si $X \sim IG(\mu, \lambda)$, la variable aléatoire

$$Z = \frac{\lambda(X-\mu)^2}{\mu^2 X}$$

suit une loi $\chi^2(1)$, montrer la validité de la fonction de géneration de $IG(\mu, \lambda)$ suivante

```
z <- rnorm(1)**2
x <- mu + (mu**2 * z - mu * sqrt(4 * mu * lambda * z + (mu * z)**2)) / (2 * lambda)
if (runif(1) > mu / (mu + x))  x <- mu**2 / x
```

Bonus (2 pts) : Démontrer que la loi de Z est bien une loi $\chi^2(1)$.