 2018–2019

Examen Final – Janvier 2019

Durée 2h00 – Sans documents

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.

Formulaire

Loi	Notation	Densité
Beta	$\mathcal{B}eta(a,b)$	$f_{a,b}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\mathbb{1}_{x\in]0,1[}$
Gamma	$\gamma(a,b)$	$f_{a,b}(x) = \frac{1}{\Gamma(a)} x^{a-1} e^{-bx} b^a \mathbb{1}_{x>0}$
Exponentielle	$\mathcal{E}(\lambda)$	$f_\lambda(x) = \lambda e^{-\lambda x}\mathbb{1}_{x>0}$
Normale	$\mathcal{N}(\mu,\sigma^2)$	$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$
Poisson	$\mathcal{P}oi(\lambda)$	$f_\lambda(x) = \frac{e^{-\lambda}}{x!}\lambda^x \mathbb{1}_{x\in\mathbb{N}}$

French – English Lexicon

- échantillon – sample
- famille exponentielle – exponential family
- statistique complète – complete statistic
- statistique exhaustive – sufficient statistic
- statistique libre – ancilary statistic
- vraisemblance – likelihood

Exercice 1 (QCM – 4 pts). **Donner la bonne réponse en la justifiant. Seules les réponses justifiées seront validées. Il n'y a pas de points négatifs.**

1. *(1 pt)* Soit $X_1, \dots, X_n$ un échantillon *i.i.d.* de loi $\mathcal{P}oi(\lambda)$. La forme générique de la loi conjuguée *a posteriori* est

(a) $\lambda \sim \gamma(a+n, b+\sum_{i=1}^n x_i)$

(b) $\lambda \sim \mathcal{P}oi(a+\sum_{i=1}^n x_i)$

(c) $\lambda \sim \mathcal{N}((a+\sum_{i=1}^n x_i)/(b+n), bn/b+n)$

(d) $\lambda \sim \gamma(a+\sum_{i=1}^n x_i, b+n)$

2. *(1 pt)* Soit $X_1, \dots, X_n$ un échantillon *i.i.d.* de loi $\mathcal{E}(\lambda)$. On observe $(Z_1, \dots, Z_n) = (\mathbb{1}(X_1 \le \delta), \dots, \mathbb{1}(X_n \le \delta))$, avec δ connu. Si $z_1 = 1$ et $(z_2, \dots, z_n) = (0, \dots, 0)$, l'estimateur du maximum de vraisemblance de λ est donné par $\hat\lambda = \dots$

(a) $-\log\left(\frac{n-1}{n}\right)/\delta$ **(b)** $\frac{(n-1)\delta}{n}$ **(c)** $\frac{\log(\delta)}{n-1}$ **(d)** $\frac{n+1}{\delta}$ **(e)** 0

3. *(1 pt)* Soit $X_1, \dots, X_n$ un échantillon *i.i.d.* tel que $\mathbb{P}(X_i = 0) = p$ et $\mathbb{P}(X_i > x) = (1-p)e^{-\lambda x}$, $(x > 0)$. Une statistique exhaustive pour ce modèle est

(a) $\bar{X}_n$

(b) $\sum_{i=1}^n X_i / \sum_{i=1}^n \mathbb{1}_{X_i > 0}$

(c) $(\sum_{i=1}^n \mathbb{1}_0(X_i), \sum_{i=1}^n \mathbb{1}_{X_i>0} X_i)$

(d) $(\sum_{i=1}^n \mathbb{1}_0(X_i), \sum_{i=1}^n \mathbb{1}_{X_i>0})$

4. *(1 pt)* Soit X de loi normale $\mathcal{N}(\mu, 1/v)$ avec μ inconnu et v connu. On considère sur μ une loi *a priori* $\mathcal{N}(m, 1/p)$. Lequels des codes ci-dessous permet d'évaluer la densité *a posteriori* $\pi(\mu \mid x)$ au point $\mu = 0.3$?

(a)
```
pnorm(0.3, (m*p^2 + x*v^2)/(x + v), sqrt((p*v)^2/(p^2+v^2)))
```

(b)
```
dnorm(0.3, (m*p + x*v)/(p+v), sqrt(1/(p+v)))
```

(c)
```
pnorm(0.3, (m*p + x*v)/(p+v), sqrt(1/(p+v)))
```

(d)
```
dnorm(0.3, (m*v + x*p)/(p+v), sqrt(1/(p+v)))
```

Exercice 2 (5 pts). Soit $X_1, \dots, X_n$, $n > 1$, un échantillon *i.i.d.* de loi uniforme sur $[0, \theta]$. Dans cet exercice, on cherche à construire un estimateur de $\lambda = 1/\theta$. Les statistiques d'ordre notées $X_{(1)}, \dots, X_{(n)}$ sont obtenues en faisant un tri croissant de l'échantillon, en particulier $X_{(1)} = \min\{X_1, \dots, X_n\}$ et $X_{(n)} = \max\{X_1, \dots, X_n\}$. Si l'on note F la fonction de répartition de X et f sa densité, on rappelle que la densité de la k-ème statistique d'ordre, $k \in \{1, \dots, n\}$, est

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} F(x)^{k-1} \{1 - F(x)\}^{n-k} f(x).$$

1. **(a)** *(1 pt)* Montrer qu'à l'exception de $X_{(1)}$, toutes les statistiques d'ordre admettent un moment inverse d'ordre 1 défini par $\mathbb{E}\left[1/X_{(k)}\right]$ et les calculer.

(b) *(0.5 pt)* En déduire que pour tout $k \geq 2$, $\hat{\lambda}_{n,k} = (k-1)/(nX_{(k)})$ est un estimateur sans biais de λ.

2. *(0.5 pt)* Donner une statistique exhaustive pour l'échantillon $X_1, \dots, X_n$.

3. *(1.5 pt)* Montrer que $\hat{\lambda}_{n,n}$ est l'estimateur de variance minimale uniformément parmi tous les estimateurs sans biais.

4. *(1.5 pt)* Justifier que pour tout $k \geq 2$, l'amélioration de l'estimateur $\hat{\lambda}_{k,n}$ par le théorème de Rao-Blackwell conduit à l'estimateur $\hat{\lambda}_{n,n}$.

Exercice 3 (6.5 pts). Dans une classe, il y a $n = X_1$ étudiants. On note X_t le nombre d'étudiants présents la semaine t, sachant que chaque semaine un étudiant décide, indépendament des autres et du passé, de revenir la semaine suivante avec probabilité p. Lorsqu'un étudiant a décidé de ne pas venir, il ne reviendra plus jamais. $X_t \mid X_{t-1} = x_{t-1}$ suit une loi binomiale $\mathcal{B}(x_{t-1}, p)$, dont la densité est notée par la suite $f(x_t \mid x_{t-1}, p)$. Notons que les X_t ne sont pas indépendants.

Pour chaque étudiant $i \in \{1, \dots, n\}$, le nombre T_i de cours auquel il aura assisté avant de ne plus venir suit une loi géométrique de paramètre $1-p$, c'est à dire que $P(T_i = k) = (1-p)p^{k-1}$, pour $k > 0$. On a alors $E(T_1) = 1/(1-p)$ et $Var(T_1) = p/(1-p)^2$.

1. *(0.5 pt)* Montrer que la loi géométrique est dans la famille exponentielle. En déduire S une statistique exhaustive de dimension 1 quelque soit la taille de l'échantillon.

2. *(1 pt)* On admet que la vraisemblance s'écrit $L(p \mid x_1, \ldots, x_{s_0}) = \prod_{t=2}^{s_0} f(x_t \mid x_{t-1}, p)$, avec $s_0 = \inf\{t \mid X_t = 0\}$. Exprimer cette vraisemblance en fonction des T_i et montrer que l'estimateur du maximum de vraisemblance est donné par

$$\hat{p}_n = 1 - \frac{n}{\sum_{i=1}^n T_i}.$$

3. (a) *(0.5 pt)* Montrer que cet estimateur est biaisé.

(b) *(1 pt)* Écrire une procédure R qui estime le biais de cet estimateur.

4. *(1.5 pt)* Donner un intervalle de confiance asymptotique pour p au niveau de confiance $1-\alpha$.

5. On mène une inférence bayésienne sur le paramètre p.

(a) *(1 pt)* Montrer que pour n'importe quel choix de loi *a priori* sur p, la loi *a posteriori* de p sachant $x_1, \ldots, x_{s_0}$ est égale à la loi *a posteriori* de p sachant S.

(b) *(0.5 pt)* On met une loi *a priori* $\mathcal{B}eta(\alpha, \beta)$ sur p. Quelle est la loi *a posteriori* de p ?

(c) *(0.5 pt)* Montrer que la moyenne de la loi *a posteriori* $\tilde{p}_n$ et $\hat{p}_n$ sont équivalents lorsque $n \to \infty$.

Exercice 4 (4.5 pts). On souhaite calibrer un appareil de voltage connu pour être d'autant moins précis que la mesure à prendre est grande. On suppose que n mesures de voltage $X_1, \ldots, X_n$ forment un échantillon *i.i.d.* de loi $\mathcal{N}(m, \theta m^2)$, avec $m \in \mathbb{R}$ connu et $\theta > 0$ inconnu.

On pourra utiliser sans démonstration les résultats suivants :

- Si Y suit une loi normale $\mathcal{N}(0, \sigma^2)$, alors Y^2 suit une loi gamma $\gamma(\frac{1}{2}, \frac{1}{2\sigma^2})$.
- Si Y et Z sont des variables aléatoires indépendantes de loi respectives $\gamma(\alpha, \lambda)$ et $\gamma(\beta, \lambda)$, alors $Y + Z$ est de loi $\gamma(\alpha + \beta, \lambda)$.
- Pour tout $a > 0$, $\Gamma(a+1) = a\Gamma(a)$.

1. *(1 pt)* Donner la vraisemblance $L(\theta \mid x_1, \ldots, x_n, m)$. En déduire $\hat{\theta}_n$ l'estimateur du maximum de vraisemblance.

2. *(1 pt)* Donner la loi de $\hat{\theta}_n$. En déduire que $\hat{\theta}_n$ est sans biais.

3. (a) *(1 pt)* Calculer l'information de Fisher $I_X(\theta)$. En déduire la borne de Cramer-Rao.

(b) *(0.5 pt)* Conclure que $\hat{\theta}_n$ est l'estimateur de variance minimale uniformément parmi les estimateurs sans biais.

4. On considère une loi a priori sur θ dont la densité est définie par :

$$\pi(\theta) = \frac{b^a}{\Gamma(a)} \left(\frac{1}{\theta}\right)^{a+1} e^{-\frac{b}{\theta}} \mathbb{1}_{\theta > 0}, \quad \text{avec } a > 0 \text{ et } b > 0 \text{ fixés.}$$

Cette loi est appelée loi inverse-gamma de paramètre (a, b).

(a) *(0.5 pt)* Donner la loi *a posteriori* de θ sachant $x_1, \ldots, x_n$.

(b) *(0.5 pt)* Calculer l'estimateur de $\theta_{\text{MAP}} = \arg\max_\theta \pi(\theta \mid x_1, \ldots, x_n)$.