

Solutions aux exercices additionnels

1. a) Soit $x_1, \dots, x_{100}$ les 100 durées de vie observées. La vraisemblance est

$$L(\lambda) = \prod_{i=1}^{100} \lambda e^{-\lambda x_i} = \lambda^{100} e^{-\lambda \sum_{i=1}^{100} x_i}, \quad \lambda > 0,$$

d'où $\ell(\lambda) = \log L(\lambda) = 100 \log \lambda - \lambda \sum_{i=1}^{100} x_i$, de dérivée $\ell'(\lambda) = 100/\lambda - \sum_{i=1}^{100} x_i$. L'estimateur de maximum de vraisemblance de λ est donc $100/\sum_{i=1}^{100} x_i$, soit l'inverse de la moyenne empirique. On trouve ici $\hat{\lambda} = 100/88912 \simeq 0.0012$.

Il est facile à vérifier que $\hat{\lambda}$ donne bien un maximum de ℓ .

- b) L'information observée est

$$j(\lambda) = -\ell''(\lambda) = \frac{100}{\lambda^2},$$

donc $j(\lambda)^{-1/2} = \lambda/10$. Un IC approximatif au niveau $\alpha = 90\%$ est donc

$$\left[\hat{\lambda} - \frac{\hat{\lambda}}{10} z_{1-\alpha/2}, \hat{\lambda} + \frac{\hat{\lambda}}{10} z_{1-\alpha/2} \right] = [0.0011, 0.0013].$$

c) $\lambda = 0.001$ se trouvant pas dans l'intervalle de confiance à 90%, il semble que ce n'est pas plausible à 90%.

2. a) Soit X_i l'état du composant i (0 si défectueux, 1 sinon). Le nombre de composants défectueux parmi n , $\sum_{i=1}^n X_i$, suit une loi binomiale de paramètres n et p . L'estimateur de maximum de vraisemblance de p est donc $\hat{p} = \bar{X}$ par un calcul vu dans les cours.

- b) L'erreur quadratique moyenne est la somme de la variance et du biais carré, soit

$$E\{(\hat{p} - p)^2\} = \text{var}(\hat{p}) + \{E(\hat{p}) - p\}^2,$$

et puisque les X_j sont des variables de Bernoulli indépendantes avec $E(X_j) = p$ et $\text{var}(X_j) = p(1-p)$, on a

$$E(\hat{p}) = nE(X_j)/n = p, \quad \text{var}(\hat{p}) = \frac{1}{n^2} \sum \text{var}(X_j) = np(1-p)/n^2 = p(1-p)/n,$$

alors

$$E\{(\hat{p} - p)^2\} = p(1-p)/n.$$

Donc $E\{(\hat{p} - p)^2\} \leq 0.002$ si et seulement si $p(1-p) \leq 0.002 \times n$. A fortiori $E\{(\hat{p} - p)^2\} \leq 0.002$ si $n \geq \frac{0.08(0.92)}{2 \times 10^{-3}}$, puisque $p \leq 0.08$. Donc un échantillon de taille $n = 37$ suffit.

3. a) On a

$$L(\theta) = f(t_1, \dots, t_n; \theta) = \prod_{j=1}^n f(t_j; \theta) = \theta^{-2n} \exp\left(-\sum_{i=1}^n t_i/\theta\right) \prod_{j=1}^n t_j, \quad \theta > 0,$$

et donc la log vraisemblance est

$$\ell(\theta) = -2n \log \theta - \theta^{-1} \sum_{j=1}^n t_j + \sum_{j=1}^n \log t_j,$$

dont les dérivées sont

$$\ell'(\theta) = -2n/\theta + \theta^{-2} \sum_{j=1}^n t_j, \quad \ell''(\theta) = 2n/\theta^2 - 2\theta^{-3} \sum_{j=1}^n t_j, \quad \theta > 0,$$

et ainsi $\hat{\theta} = \sum_{i=1}^n T_i / (2n)$. Il est facile à vérifier que $\hat{\theta}$ donne bien un maximum.

b) On a

$$E(\hat{\theta}) = \frac{1}{2n} \sum_{i=1}^n E(T_i) = \frac{1}{2} E(T),$$

et

$$E(T) = \int_0^\infty y f(y; \theta) dy = \int_0^\infty \frac{y^2}{\theta^2} e^{-y/\theta} dy = \frac{2}{\theta} \int_0^\infty y e^{-y/\theta} dy = 2\theta,$$

ce qui nous donne $E(\hat{\theta}) = \theta$. Donc $\hat{\theta}$ est sans biais, et ainsi l'erreur quadratique moyenne coïncide avec la variance de l'estimateur:

$$\text{var}(\hat{\theta}) = \frac{1}{4} \text{var}\left(\sum_{i=1}^n T_i / n\right) = \frac{1}{4n} \text{var}(T).$$

On vérifie aisément que $\text{var}(T) = 2\theta^2$, et ainsi l'erreur quadratique moyenne est $\theta^2 / (2n)$.

4. a) Les pannes étant rares et leur répartition étant aléatoire, on peut supposer que, pour θ fixé, X suit une loi de Poisson de paramètre θ :

$$f(x | \theta) = \theta^x e^{-\theta} / x!, \quad x = 0, 1, 2, \dots$$

b) Puisque la loi gamma(a, b) a pour moyenne a/b et variance a/b^2 , il faut ici que $a/b = 4$ et $a/b^2 = 2^2$ et donc que $a = 4$, $b = 1$.

c) On dispose de $n = 2$ observations de X , avec les valeurs $x_1 = 2$ (nombre de pannes la semaine dernière) et $x_2 = 0$ (nombre de pannes cette semaine). La densité a posteriori de θ est donc

$$\pi(\theta | x_1, x_2) = \frac{f(x_1 | \theta) f(x_2 | \theta) \pi(\theta)}{f(x_1, x_2)} = \frac{\theta^{x_1}}{x_1!} e^{-\theta} \frac{\theta^{x_2}}{x_2!} e^{-\theta} \times \frac{\theta^{a-1} b^a}{\Gamma(a)} e^{-b\theta} \propto \theta^{a+x_1+x_2-1} e^{-(b+2)\theta}, \quad \theta > 0,$$

où nous avons ignoré les constants ne dépendant pas de θ . Nous voyons que la densité a posteriori est aussi une densité de gamma, avec (a, b) remplacé par $(a + x_1 + x_2, b + 2) = (6, 3)$. Donc

$$\pi(\theta | x_1, x_2) = 3^6 \theta^5 e^{-3\theta} / \Gamma(6), \quad \theta > 0,$$

dont l'espérance a posteriori est $6/3 = 2$ pannes en moyenne par semaine et la variance a posteriori est $6/3^2 = 2/3$. Donc l'information x_1, x_2 a réduit l'espérance de 4 à 3 et aussi la variance de 1 à $2/3$.