

Examen de « Survie »

Master Mathématiques et Applications

Parcours Statistique

Année 2023 - 2024

Durée : 2h. Les calculatrices et téléphones portables sont interdits

Exercice 1 – Donner la définition de l'estimateur de Kaplan-Meier. Vous prendrez soin de définir clairement et précisément toutes les notations requises.

Exercice 2 – Pour $n = 9$ individus on dispose des données suivantes.

0.6	1.7 ⁺	1.6 ⁺	1.6	1.2	1.1	1.2	0.9 ⁺	1.4
-----	------------------	------------------	-----	-----	-----	-----	------------------	-----

Ces valeurs correspondent aux $n = 9$ observations $\{t_1^*, \dots, t_n^*\}$ de la variable aléatoire $T^* = \min(T, C)$. L'exposant + signifie que la valeur a été censurée à droite (i.e., l'événement n'est pas observé).

- 1) Donner la valeur m ainsi que l'ensemble $\{t_{(1)}^* < \dots < t_{(m)}^*\}$ des observations distinctes et ordonnées.

Pour tout $i \in \{1, \dots, m\}$, on note O_i le nombre d'événements observés à l'instant $t_{(i)}^*$ et N_i le nombre d'individus à risque à l'instant $t_{(i)}^*$.

- 2) Donner les valeurs O_1, \dots, O_m et N_1, \dots, N_m .
- 3) Calculer la valeur observée de l'estimateur de Kaplan-Meier au point $t_{(3)}^*$. (Vous donnerez le résultat sous forme d'une fraction irréductible.)

Exercice 3 – Soient T et C deux variables aléatoires indépendantes et positives. On pose $T^* := \min(T, C)$ et $\Delta = \mathbb{I}_{\{T \leq C\}}$. On suppose que T suit une loi exponentielle de paramètre $\lambda > 0$ (i.e., d'espérance $1/\lambda$) et que C suit une loi exponentielle de paramètre $\theta\lambda$ avec $\theta > 0$. On note $f_T(\cdot)$ (resp. $f_C(\cdot)$) la densité de la loi de T (resp. C) et $S_T(\cdot)$ (resp. $S_C(\cdot)$) la fonction de survie de la loi de T (resp. C).

- 1) Montrer que

$$\mathbb{P}(T \leq C) = \int_0^\infty S_C(x) f_T(x) dx = \int_0^\infty (1 - S_T(x)) f_C(x) dx.$$

(Aide : utiliser les probabilités conditionnelles $\mathbb{P}(T \leq C \mid T = x)$ et $\mathbb{P}(T \leq C \mid C = x)$).

- 2) En déduire que $\mathbb{P}(T \leq C) = 1/(1 + \theta)$.
- 3) Quelle est la loi de T^* ?

Soient $(\mathbf{T}^*, \mathbf{\Delta}) = \{(T_1^*, \Delta_1), \dots, (T_n^*, \Delta_n)\}$ des couples indépendants de même loi que le couple (T^*, Δ) . On note $(\mathbf{t}^*, \mathbf{d}) = \{(t_1^*, d_1), \dots, (t_n^*, d_n)\}$ les observations associées. On pose

$$\bar{T}_n^* := \frac{1}{n} \sum_{i=1}^n T_i^* \quad \text{et} \quad \bar{\Delta}_n := \frac{1}{n} \sum_{i=1}^n \Delta_i.$$

Les observations des variables aléatoires \bar{T}_n^* et $\bar{\Delta}_n$ sont notées \bar{t}_n^* et \bar{d}_n .

- 4) Quel est l'effet du paramètre θ sur $n\mathbb{E}(\bar{\Delta}_n)$ qui est le nombre attendu d'événements dans un échantillon de taille n ?

- 5) En utilisant la question 2), proposez un estimateur (i.e., une fonction de $(\mathbf{T}^*, \mathbf{\Delta})$) du paramètre θ .
- 6) En utilisant l'expression de l'espérance de T^* , proposez un estimateur de λ dont l'expression dépendra de \bar{T}_n^* et de l'estimateur de θ obtenu à la question précédente.

Pour tout $h > 0$, on introduit la fonction de vraisemblance observée

$$L_h(\lambda, \theta \mid (\mathbf{t}^*, \mathbf{d})) := \frac{1}{h^n} \prod_{i=1}^n \mathbb{P}(\{t_i^* - h < T \leq t_i^*\} \cap \{\Delta = d_i\}).$$

- 7) Montrer, en justifiant correctement chaque étape, que

$$L_h(\lambda, \theta \mid (\mathbf{t}^*, \mathbf{d})) := \frac{1}{h^n} \prod_{i=1}^n [\mathbb{P}(\{t_i^* - h < T \leq t_i^*\} \cap \{T \leq C\})]^{d_i} \\ \times [\mathbb{P}(\{t_i^* - h < C \leq t_i^*\} \cap \{T > C\})]^{1-d_i}$$

- 8) Donner l'expression de la limite

$$\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}(\{t_i^* - h < T \leq t_i^*\} \cap \{T \leq C\}).$$

Vous justifierez correctement votre réponse.

- 9) En déduire que la vraisemblance observée du modèle est

$$L_0(\lambda, \theta \mid (\mathbf{t}^*, \mathbf{d})) := \lim_{h \rightarrow 0} L_h(\lambda, \theta \mid (\mathbf{t}^*, \mathbf{d})) \\ = \prod_{i=1}^n [f_T(t_i^*) S_C(t_i^*)]^{d_i} [f_C(t_i^*) S_T(t_i^*)]^{1-d_i}.$$

- 10) Donner l'expression (en fonction de λ , θ , \bar{T}_n^* et $\bar{\Delta}_n$) de la log-vraisemblance

$$\mathcal{L}_0(\lambda, \theta \mid (\mathbf{T}^*, \mathbf{\Delta})) := \log(L_0(\lambda, \theta \mid (\mathbf{T}^*, \mathbf{\Delta}))).$$

- 11) Donner l'expression des estimateurs $\hat{\lambda}_n$ et $\hat{\theta}_n$ et vérifier qu'ils coïncident avec ceux trouvés aux questions 5) et 6).

En supposant que les hypothèses de régularité requises sont satisfaites, on rappelle que

$$[\mathcal{I}_n(\theta, \lambda)]^{1/2} \left[\begin{pmatrix} \hat{\theta}_n \\ \hat{\lambda}_n \end{pmatrix} - \begin{pmatrix} \theta \\ \lambda \end{pmatrix} \right] \xrightarrow{d} \mathcal{N}(0, I_2)$$

où I_2 est la matrice identité de dimension 2 et $\mathcal{I}_n(\theta, \lambda)$ est la matrice d'information de Fisher donnée par

$$\mathcal{I}_n(\theta, \lambda) = -\mathbb{E} \begin{pmatrix} \frac{\partial^2}{\partial \theta^2} \mathcal{L}_0(\lambda, \theta \mid (\mathbf{T}^*, \mathbf{\Delta})) & \frac{\partial^2}{\partial \theta \partial \lambda} \mathcal{L}_0(\lambda, \theta \mid (\mathbf{T}^*, \mathbf{\Delta})) \\ \frac{\partial^2}{\partial \theta \partial \lambda} \mathcal{L}_0(\lambda, \theta \mid (\mathbf{T}^*, \mathbf{\Delta})) & \frac{\partial^2}{\partial \lambda^2} \mathcal{L}_0(\lambda, \theta \mid (\mathbf{T}^*, \mathbf{\Delta})) \end{pmatrix}$$

- 12) Donner l'expression de la matrice d'information de Fisher en fonction de n , θ , λ , $\mathbb{E}(T^*)$ et $\mathbb{P}(T > C)$.

- 13) Montrer que

$$\frac{\bar{\Delta}_n^2}{1 - \bar{\Delta}_n} \xrightarrow{\mathbb{P}} \frac{\mathbb{P}(T > C)}{\theta^2}.$$

On introduit la matrice

$$\widehat{\mathcal{I}}_n = n \begin{pmatrix} \overline{\Delta}_n^2 / (1 - \overline{\Delta}_n) & \overline{T}_n^* \\ \overline{T}_n^* & (\overline{T}_n^* / \overline{\Delta}_n)^2 \end{pmatrix}$$

On admettra que $\widehat{\mathcal{I}}_n^{-1} \mathcal{I}_n(\theta, \lambda) \xrightarrow{\mathbb{P}} I_2$.

14) Montrer que

$$\sqrt{n} \frac{\overline{T}_n^*}{\overline{\Delta}_n} (\widehat{\lambda}_n - \lambda) \xrightarrow{d} \mathcal{N}(0, 1).$$

15) Proposer un test de l'hypothèse nulle $H_0 : \mathbb{E}(T) = 1$ contre l'hypothèse alternative $H_1 : \mathbb{E}(T) \neq 1$ avec un risque de première espèce égal à α .