Modèles de durée / Examen / Janvier 2018

Durée 2h – aucun document n'est autorisé

Vraisemblance d'un modèle paramétrique tronqué et censuré

La qualité de la rédaction, des justifications apportées et de la présentation de la copie seront prises en compte dans la notation.

On considère une variable de durée X dont la fonction de hasard sous-jacente est notée h_X et la fonction de survie S_X.

Les durées de survie $(X_1,\dots,X_n)$ sont censurées à droite par des durées $(C_1,\dots,C_n)$ indépendantes de l'échantillon d'intérêt et, au lieu d'observer directement $(X_1,\dots,X_n)$ on observe $(T_1,D_1),\dots,(T_n,D_n)$ avec :

$$T_i = X_i \wedge C_i \text{ et } D_i = \begin{cases} 1 \; si \; X_i \leq C_i \\ 0 \; si \; X_i > C_i \end{cases}$$

Question n°1 (4 points) : Déterminez l'expression de la vraisemblance en fonction des fonctions de survie et de hasard de *X* et de C.

Question n°2 (4 points) : Quelle hypothèse supplémentaire faut-il faire pour éliminer (au sens de « considérer comme une constante pour l'estimation de la loi de X ») la loi de la censure dans l'expression ci-dessus ? Donnez l'expression qu'on en déduit de la log-vraisemblance en fonction de $S_X = S_\theta$ et $h_X = h_\theta$. Donnez également un exemple dans lequel la loi de la censure ne peut pas être éliminée de la vraisemblance, alors que la censure est indépendante de la durée d'intérêt. Connaissez-vous une expérience dans laquelle la censure n'est pas indépendante de la durée modélisée ?

On suppose dans la suite que la condition permettant d'éliminer la loi de la censure de l'expression de la log-vraisemblance est vérifiée.

Question n°3 (2 points) : Qu'appelle-t-on troncature gauche ? Donnez un exemple de troncature gauche dans un contexte d'assurance. Comment faut-il adapter l'expression de la log-vraisemblance établie à la question précédente dans ce cas ?

Question n°4 (4 points) : On fait l'hypothèse que la fonction de hasard est constante sur un intervalle $[x, x+1[$; à l'aide de ce qui précède, calculez la log-vraisemblance pour l'estimation du paramètre θ, valeur de la fonction de hasard sur $[x, x+1[$ et en déduire l'estimateur du maximum de vraisemblance de la fonction de hasard sur cet intervalle. Quel lien avec la loi de Poisson peut-on faire ? Quel intérêt présente ce lien ?

Question n°5 (2 points) : Rappelez les définitions de la fonction de survie conditionnelle S_x et de la probabilité conditionnelle de sortie q_x ? Quel est le lien entre S_x et q_x ?

Question n°6 (2 points) : En utilisant les résultats des questions 4 et 5, indiquez quel est l'estimateur du maximum de vraisemblance de q_x lorsque la fonction de hasard est constante sur $[x, x+1[$; quelle hypothèse supplémentaire faut-il faire pour que l'on puisse utiliser l'estimateur de Hoem, $\hat{q}_x = \frac{D_x}{E_x}$?

Question n°7 (2 points) : Si l'on souhaite utiliser l'approximation ci-dessus comme estimateur de q_x même lorsque l'hypothèse supplémentaire de la question 6 n'est pas satisfaite, quelle correction faut-il apporter au calcul de l'exposition au risque ? Donnez un exemple de l'intérêt de cette correction.

Annale 2018

Question 1 :

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} P(T_i \in [t_i, t_i+dt_i]) \frac{1}{dt_i}$$

$$= \prod_{i=1}^{n} \frac{1}{dt_i} P(T_i \in [t_i, t_i+dt_i], D_i=1)^{D_i} \, P(T_i \in [t_i, t_i+dt_i], D_i=0)^{1-D_i}$$

* $P(T_i \in [t_i, t_i+dt_i], D_i=1) = P(X_i \wedge C_i \in [t_i, t_i+dt_i],\ X_i \leq C_i)$
$= P(X_i \in [t_i, t_i+dt_i],\ t_i \leq C_i)$
$= P(X_i \in [t_i, t_i+dt_i])\, P(C_i \geq t_i)$
$= f_X(\theta, T_i)\, S_C(\theta, T_i)\, dt_i$

* $P(T_i \in [t_i, t_i+dt_i], D_i=0) = P(X_i \wedge C_i \in [t_i, t_i+dt_i],\ X_i > C)$
$= P(C_i \in [t_i, t_i+dt_i],\ X_i > C)$
$= P(C_i \in [t_i, t_i+dt_i])\, P(X_i > C)$
$= f_C(\theta, T_i)\, S_X(\theta, C)\, dt_i$

$$\Rightarrow \mathcal{L}(\theta) = \prod_{i=1}^{n} [f_X(\theta, T_i)\, S_C(\theta, T_i)]^{D_i} \, [f_C(\theta, T_i)\, S_X(\theta, T_i)]^{1-D_i}$$

Quest° 2 :

On doit supposer que les lois de X et C n'ont pas de paramètre commun (censure non informative)

$$\Rightarrow \mathcal{L}(\theta) = cte \times \prod_{i=1}^{n} [f_\theta(T_i)]^{D_i} [S_\theta(C_i)]^{1-D_i}$$

T_i → car dans le cas $D_i=0$ $T_i = C$

$$= cte \times \prod_{i=1}^{n} [h_\theta(T_i) S_\theta(T_i)]^{D_i} [S_\theta(C_i)]^{1-D_i}$$

$$= cte \times \prod_{i=1}^{n} h_\theta(T_i)^{D_i} S_\theta(T_i)$$

$$\Rightarrow \ln(\mathcal{L}(\theta)) = \sum_{i=1}^{n} [D_i \ln(h_\theta(T_i)) + \ln(S_\theta(T_i))]$$

Exemple dans lequel la loi de censure ne peut pas être éliminée de la vraisemb. alors que la censure est $\perp$ de la durée d'intérêt :
→ Si on a $S_\theta(x) = S_C(x)^\beta$

Censure de type II (arrêt au r-ième décès) → loi de censure non indep de la durée modélisée.

Quest° 3 :

Troncature à gauche : la variable d'intérêt n'est pas observable lorsqu'elle est < à un seuil.

→ Ex : Contrat d'arrêt de travail avec une franchise.

En cas de troncature, il faut adapter les lois conditionnellem^t à l'ev^t que la variable soit comprise dans les bornes de l'expérience :

$$h_\theta(t \mid e<T) = h(t)$$

$$S_\theta(t \mid e<T) = \frac{S_\theta(t)}{S_\theta(e)}$$ où e désigne le seuil de troncature à gauche.

$$\Rightarrow \mathcal{L}(\theta) = cte \times \prod_{i=1}^{n} h_\theta(T_i)^{D_i} \frac{S_\theta(T_i)}{S_\theta(E_i)}$$

$$\Rightarrow \ln(\mathcal{L}(\theta)) \propto \sum_{i=1}^{n} [D_i \ln(h_\theta(T_i)) + \ln(S_\theta(T_i)) - \ln(S_\theta(E_i))]$$

Q4 :

Fonct° de hasard cte : $h_\theta(t) = \theta$

$\Rightarrow$ Modèle exponentiel : $S_\theta(t) = e^{-\theta t}$

$$\Rightarrow \ln(\mathcal{L}(\theta)) \propto \sum_{i=1}^{n} [D_i \ln(\theta) - \theta(T_i - E_i)]$$

$$\propto \ln(\theta) \sum_{i=1}^{n} D_i - \theta \sum_{i=1}^{n} (T_i - E_i)$$

Sur $[x, x+1[$,

$$D_x = \sum_{i=1}^{n} d_i(x)$$

$$E_x = \sum_{i=1}^{n} t_i(x) - e_i(x)$$ avec $t_i(x) = t_i \wedge (x+1)$

$e_i(x) = t_i \vee x$

$$\rightarrow \ln(\mathcal{L}(\theta)) \propto \ln(\theta) D_x - \theta E_x$$