© Théo Jalabert

Modèles de durée / Examen / Janvier 2023

Durée 2h – aucun document n'est autorisé

La qualité de la rédaction, des justifications apportées et de la présentation de la copie seront prises en compte dans la notation. Chaque réponse doit être correctement justifiée : une réponse juste sans justification sera considérée comme fausse.

Quelques propriétés du modèle de Weibull

On 'intéresse au modèle de Weibull ; dont la fonction de hasard est de la forme :

$$\mu(t) = \lambda\alpha t^{\alpha-1},\ \alpha,\lambda > 0$$

Question n°1 (1 point) : Calculez la fonction de survie.

Question n°2 (3 point) : Calculez les moments non centrés d'ordre *k*, $E\left(X^k\right)$ à l'aide de la fonction $\Gamma(x) = \int_0^{+\infty} u^{x-1}e^{-u}du$. En déduire l'espérance, la variance et le coefficient de variation. Que peut-on dire de ce dernier ?

Question n°2 (1 point) : Démontrez que si la variable *T* est distribuée selon une loi exponentielle de paramètre $\lambda > 0$ alors $T^{1/\alpha}$ suit $W(\alpha,\lambda)$.

Question n°3 (4 points) : Prouver que, si $(X_1,\ldots,X_n)$ est un échantillon d'une loi de fonction de répartition G sur $]0,+\infty[$ qui vérifie :

$$\lim_{x\to 0^+} \frac{G(x)}{\lambda x^\alpha} = 1,$$

alors $n^{1/\alpha}X_{(1)}$ converge en loi lorsque *n* tend vers l'infini vers une distribution $W(\alpha,\lambda)$. On rappelle que $X_{(1)} = \min\left(X_i, 1 \le i \le n\right)$.

On s'intéresse maintenant à l'estimation des paramètres (α,λ).

Question n°4 (2 points) : Trouver une relation entre (α,λ) et $\ln\left(-\ln S(t)\right)$ et en déduire un estimateur graphique de (α,λ).

Question n°5 (3 points) : Rappelez la forme générale de la log-vraisemblance d'un modèle avec une censure aléatoire droite non informative et une troncature gauche. Vous redonnerez la définition de « censure aléatoire droite non informative » et « troncature gauche ».

Question n°6 (4 points) : Écrire les conditions du premier ordre du programme de maximisation de la log-vraisemblance. Que peut-on en déduire dans le cas particulier $\alpha = 1$? Dans le cas général ?

Question n°7 (2 points) : Décrire un algorithme permettant de résoudre le système d'équations ci-dessus.

© Théo Jalabert

Quelques propriétés du modèle de Weibull

$$\mu(t) = \lambda \alpha t^{\alpha-1} \qquad \alpha, \lambda \in (\mathbb{R}_+^*)^2$$

① $S(t) = \exp\left(-\int_0^t \mu(s)\,ds\right) = \exp\left(-\int_0^t \lambda\alpha s^{\alpha-1} ds\right)$

$$= \exp(-\lambda t^\alpha)$$

② $f : t \mapsto \mu(t) S(t)$ est la fonct° densité d'un Weibull

$\Rightarrow$ Ici $f(t) = \lambda \alpha t^{\alpha-1} e^{-\lambda t^\alpha}$

On a $\Gamma : x \mapsto \int_0^\infty u^{x-1} e^{-u} du$

$u = \lambda x^\alpha \Rightarrow du = \lambda \alpha x^{\alpha-1}$

$\Rightarrow x = \left(\frac{1}{\lambda} u\right)^{1/\alpha}$

Soit $X \sim \text{Weibull}(\lambda, \alpha)$

$\rightarrow \mathbb{E}[X^k] = \int_0^\infty x^k f(x)\,dx = \int_0^\infty x^k \lambda\alpha x^{\alpha-1} e^{-\lambda x^\alpha} dx$

$$= \frac{1}{\lambda^{k/\alpha}} \int_0^\infty u^{k/\alpha} e^{-u} du$$

$$= \lambda^{-k/\alpha} \int_0^\infty u^{\left(\frac{k}{\alpha}+1\right)-1} e^{-u} du = \lambda^{-k/\alpha}\, \Gamma\left(\frac{k}{\alpha}+1\right)$$

$\Rightarrow \mathbb{E}[X] = \lambda^{-1/\alpha}\, \Gamma(1 + 1/\alpha)$

$\mathbb{E}[X^2] = \lambda^{-2/\alpha}\, \Gamma(1 + 2/\alpha)$

$\Rightarrow \mathbb{V}[X] = \lambda^{-2/\alpha} \Gamma(1 + 2/\alpha) - \lambda^{-2/\alpha} \Gamma(1 + 1/\alpha)^2$

$$= \lambda^{-2/\alpha} \left(\Gamma(1 + 2/\alpha) - \Gamma(1 + 1/\alpha)^2 \right)$$

Le coeff de variat° vaut $\dfrac{\sqrt{\mathbb{V}[X]}}{\mathbb{E}[X]}$ qui est $\perp\!\!\!\perp$ de λ.

② Soit $T \sim e^{-\lambda}$ $\quad \lambda > 0$

Mq $T^{1/\alpha} \sim W(\alpha, \lambda)$

$S_{T^{1/\alpha}}(x) = \mathbb{P}(T^{1/\alpha} > x) = \mathbb{P}(T > x^\alpha) = \underbrace{e^{-\lambda x^\alpha}}_{\text{fonct° de survie d'une } W(\alpha,\lambda)}$

③ Soit $(X_1, \dots, X_n)$ un échantillon de loi de f.d.r G sur $\mathbb{R}_+^*$ tq

$$\lim_{x \to 0^+} \frac{G(x)}{\lambda x^\alpha} = 1$$

$\Rightarrow n^{1/\alpha} X_{(1)} \xrightarrow[n\to\infty]{\mathcal{L}} W(\alpha, \lambda)$? où $X_{(1)} = \min\limits_{i \in [\![1,n]\!]} (X_i)$

© Théo Jalabert

$\rightarrow$ On a $\mathbb{P}(n^{1/\alpha} X_{(1)} > x) = \mathbb{P}(X_{(1)} > x n^{-1/\alpha})$

$$= \mathbb{P}(\min_{i \in [\![1,n]\!]} X_i > x n^{-1/\alpha})$$

$$\underset{X_i \text{ iid}}{=} \mathbb{P}(X_1 > x n^{-1/\alpha})^n$$

$$= [1 - G(x n^{-1/\alpha})]^n$$

$$\Rightarrow \ln(\mathbb{P}(n^{1/\alpha} X_{(1)} > x)) = n \ln(1 - G(x n^{-1/\alpha}))$$

On se place dans le cas où $n \in \mathcal{V}(+\infty)$

$$= n \ln(1 - \lambda (x n^{-1/\alpha})^\alpha + o(\tfrac{1}{n}))$$

$$= n[-\lambda x^\alpha n^{-1} + o(\tfrac{1}{n})]$$

$$= -\lambda x^\alpha + o(1)$$

D'où $\lim_{n \to \infty} \mathbb{P}(n^{1/\alpha} X_{(1)} > x) = e^{-\lambda x^\alpha}$

(4) On a mq $S : t \longmapsto e^{-\lambda t^\alpha}$

$$\Rightarrow \ln(-\ln(S(t))) = \ln(\lambda t^\alpha)$$

$$= \ln(\lambda) + \alpha \ln(t)$$

$\Rightarrow$ Si $\hat{S}(t)$ designe un estimateur empirique de la fonct° de survie, les point $(\ln(t), \ln(-\ln(\hat{S}(t))))$ doivent être approximativem^t alignés sur une droite d'ordonnée à l'origine $\ln(\lambda)$ et de pente α.

(5) $\ln(\mathcal{L}(\theta)) = \sum_{i=1}^{n} d_i \ln(\mu_\theta(t_i)) + \ln(S_\theta(t_i)) - \ln(S_\theta(e_i))$

<u>Censure aléatoire droite non informative :</u>

<u>Censure droite</u> : Si X est la variable d'intérêt, l'observation de la censure C indique que $X \geq C$

<u>Censure aléatoire</u> : Soit un échantillon de durées de vie $(X_1, \ldots, X_n)$ et un second échantillon indep composé de variables positives $(C_1, \ldots, C_n)$, on dit qu'il y a Censure de type III / aléatoire si au lieu d'observer directement $(X_1, \ldots, X_n)$ on observe $(T_1, D_1), \ldots, (T_n, D_n)$ avec :

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

<u>Censure non informative</u> : Cela signifie que la loi de censure est indépendante du paramètre θ.
$\Leftrightarrow$ Les lois de X et C n'ont pas de paramètres commun.

<u>Troncature gauche</u>

On dit qu'il y a troncature gauche lorsque la variable d'intérêt n'est pas observable lorsqu'elle est inférieure à un seuil $c > 0$

© Théo Jalabert

(6) On trouve en notant (e_i, t_i, d_i) l'instant d'entrée dans l'observation (troncature gauche), le dernier instant observé et l'indicatrice de non censure :

$$\frac{\partial}{\partial \lambda} \ln(L) = \frac{d}{\lambda} - \sum_{i=1}^{n} (t_i^{\alpha} - e_i^{\alpha})$$

$$\frac{\partial}{\partial \alpha} \ln(L) = \frac{d}{\alpha} + \sum_{i=1}^{n} d_i \ln(t_i) - \lambda\alpha \sum_{i=1}^{n} (t_i^{\alpha-1} - e_i^{\alpha-1}) \qquad \text{où } d = \sum_{i=1}^{n} d_i$$

Si $\alpha = 1$, on retrouve l'estimateur du paramètre d'une loi exponentielle

$$\hat{\lambda} = \frac{d}{\sum_{i=1}^{n} (t_i - e_i)}$$

Dans le cas général, les équat° de vraisemblance ne donnent pas de solut° explicite, il faut se tourner vers une approximat° numérique.

(7) La méthode la + simple est d'utiliser l'algorithme de Newton-Raphson, basé sur la récurrence :

$$\theta_{k+1} = \theta_k - \left[\frac{\partial^2}{\partial\theta\partial\theta'} \ln(\mathcal{L}(\theta_k))\right]^{-1} \frac{\partial \ln(\mathcal{L}(\theta_k))}{\partial\theta} \qquad \text{où } \theta = (\alpha, \lambda)$$

Dans le cas de la distribut° de Weibull, on a

$$\frac{\partial^2}{\partial\theta\partial\theta'} \ln(\mathcal{L}(\theta_k)) = \begin{pmatrix} -\frac{d}{\alpha^2} - \lambda\sum_{i=1}^{n}(t_i^{\alpha-1} - e_i^{\alpha-1}) - \lambda\alpha\sum_{i=1}^{n}(t_i^{\alpha-1}\ln(t_i) - e_i^{\alpha-1}\ln(e_i)) & \alpha\sum_{i=1}^{n}(t_i^{\alpha-1} - e_i^{\alpha-1}) \\ \alpha\sum_{i=1}^{n}(t_i^{\alpha-1} - e_i^{\alpha-1}) & -\frac{d}{\lambda^2} \end{pmatrix}$$

Pour les valeurs init de l'algo, on peut par exemple prendre l'estimateur graphique de la Q4