Université Claude Bernard - Lyon 1
ISFA - M2SAF
Frédéric Planchet et Maxime Ben-Brik

Année 2021-2022

Travaux dirigés : modèles de durée
Séance n°2

Exercise 1 Censure et troncature.

Pour chacun des exemples suivants, décrire en détail le type de censure et / ou le type de troncature considérés.

1. On considère une expérience effectuée sur une population de rats pour tester la dangerosité d'une substance. La variable d'intérêt est leur durée de survie après ingestion de la substance. Cette expérience est effectuée sur 60 jours. Considérer les cas suivants :
   — Rat 1 : un rat survivant au terme des 60 jours ;
   — Rat 2 : un rat mort le $32^{\text{ème}}$ jour, mais d'une cause sans lien avec la substance ;
   — Rat 3 : un rat mort le $10^{\text{ème}}$ jour du fait de la substance.
2. On s'intéresse à la mortalité d'un portefeuille d'assurance pour une garantie en cas de décès. La population est observée entre le 01/01/2010 et le 01/01/2016. Considérer les cas suivants :
   — Assuré 1 : un assuré ayant souscrit son contrat le 01/01/2009 et décédé le 01/01/2012 ;
   — Assuré 2 : un assuré ayant souscrit son contrat le 01/01/2011 et décédé le 30/06/2016 ;
   — Assuré 3 : un assuré ayant souscrit son contrat le 01/01/2008 et décédé le 01/01/2009 ;
   — Assuré 4 : un assuré ayant souscrit son contrat le 01/01/2009 et l'ayant racheté le 01/01/2013 pour raison de santé ;
   — Assuré 5 : un assuré ayant souscrit son contrat le 01/01/2009, s'étant présenté le 01/01/2013 pour renouveler son contrat et pour lequel son épouse demande ensuite le versement de prestations le 01/01/2015.
3. On s'intéresse pour un contrat d'assurance arrêt de travail à la durée de survie en invalidité. Le contrat comprend une période de franchise de 3 mois [*]. La population est observée entre le 01/01/2010 et le 01/01/2016. Considérer les cas suivants :
   — Assuré 1 : un assuré entré en invalidité le 01/01/2011 et décédé le 01/01/2012 ;
   — Assuré 2 : un assuré entré en invalidité le 01/01/2011 et décédé dans les trois mois ;
   — Assuré 3 : un assuré entré en invalidité le 01/12/2015.

Exercise 2 Vraisemblance et données manquantes.

Dans la plupart des analyse de survie, les observations comprennent des données manquantes. Soit un échantillon de n individus de durée de vie respective $T_1, \ldots, T_n$ i.i.d. Ces observations suivent une loi de densité f définie à partir du paramètre $\boldsymbol{\theta}$.

---

*. La date décès n'est pas communiquée si le décès survient pendant cette période.

1. Dans chacun des cas suivant, écrire l'expression générale de la vraisemblance, puis introduire progressivement (a) l'hypothèse d'indépendance du processus lié aux données manquantes et (b) le fait que ce processus ne dépende pas de $\boldsymbol{\theta}$ (non-informatif).
   — censure C à droite de type I ;
   — troncature individuelle à gauche L_i ;
   — censure individuelle à gauche L_i et à droite R_i.
2. On suppose à présent que les observations sont uniquement soumises à censure **indépendante** à droite C_i. On fait l'hypothèse que les T_i et les C_i ont pour fonction de hasard respective $h_T(t) = \alpha t^{\alpha-1}$ et $h_C(t) = \beta t^{\beta-1}$ (Weibull). Écrire la log-vraisemblance du modèle et en déduire l'équation vérifie par l'estimateur du maximum de vraisemblance de $\boldsymbol{\theta} = (\alpha, \beta)$.

Exercise 3 Modèle de Gompertz-Makeham.

Soit T la variable aléatoire positive représentant la durée de vie d'un individu que l'on suppose définie à partir du modèle de Gompertz-Makeham dont la fonction de hasard vaut $h(t) = a + bc^t$, avec a, b et c trois paramètres strictement positifs.

1. Rappeler pour ce modèle l'expression de sa fonction de survie, sa densité et sa fonction de survie conditionnelle à l'évènement $T > t_0$.
2. Ce modèle vous semble t-il adapté à la modélisation de la vie humaine ?
3. En présence de censure indépendante à droite non-informative, écrire la log-vraisemblance du modèle et calculer sa dérivée première pour une échantillon $(y_1, d_1, \ldots, y_n, d_n)$.

Exercice 1: Décrire type de censure et/ou type de troncature considérés

1) * Rat 1: Censure à droite de type 1
* Rat 2: Censure à droite aléatoire
* Rat 3: Observation complète

2) * Assuré 1: Troncature gauche (observation se fait à partir de l'âge au 01/01/2010)
* Assuré 2: Troncature gauche (car observation à partir de l'âge au 01/01/2011) et on a une censure à droite de type 1 (décès au 30/06/2016 or fin d'observation au 01/01/2016)
* Assuré 3: Observation non incluse.
* Assuré 4: Troncature gauche (observation qui se fait à partir de l'âge au 01/01/2010) puis on a aussi une censure à droite aléatoire potentiellement non indépendante
* Assuré 5: Troncature gauche (observat° qui se fait à partir de l'âge au 01/01/2010) puis potentielle censure à droite car le décès est vraisemblablement survenu entre le 01/01/2013 et le 01/01/2015

3) * Assuré 1: Observation complète
* Assuré 2: Censure à gauche aléatoire
* Assuré 3: Troncature à droite (car seul les décès > 3mois de la fin de l'étude peuvent être observés)

Exercice 2:

1) Censure de type I:

On observe $(Y_i, D_i)\ \forall i$ où $Y_i = T_i \wedge C$ et $D_i = \mathbb{1}_{\{Y_i = T_i\}}$

La contribution à la vraisemblance individuelle s'écrit

$$
\begin{aligned}
\mathcal{L}_i(\theta) &= \mathbb{P}(Y_i = y_i, D_i = d_i ; \theta) \\
&= \mathbb{P}(T_i = y_i, D_i = 1 ; \theta)^{d_i}\, \mathbb{P}(C = y_i, D_i = 0 ; \theta)^{1-d_i} \\
&= \mathbb{P}(T_i = y_i, T_i \leq C ; \theta)^{d_i}\, \mathbb{P}(C = y_i, C \leq T_i ; \theta)^{1-d_i}
\end{aligned}
$$

Si la censure est indépendante, il vient:

$$\mathcal{L}_i(\theta) = \left(f_T(y_i;\theta) S_C(y_i;\theta)\right)^{d_i} \left(f_C(y_i;\theta) S_T(y_i;\theta)\right)^{1-d_i}$$

Exercise 2 Vraisemblance et données manquantes.

Dans la plupart des analyse de survie, les observations comprennent des données manquantes. Soit un échantillon de n individus de durée de vie respective $T_1, \ldots, T_n$ i.i.d. Ces observations suivent une loi de densité f définie à partir du paramètre $\boldsymbol{\theta}$.

∗. La date décès n'est pas communiquée si le décès survient pendant cette période.

1. Dans chacun des cas suivant, écrire l'expression générale de la vraisemblance, puis introduire progressivement (a) l'hypothèse d'indépendance du processus lié aux données manquantes et (b) le fait que ce processus ne dépende pas de $\boldsymbol{\theta}$ (non-informatif).
   — censure C à droite de type I ;
   — troncature individuelle à gauche L_i ;
   — censure individuelle à gauche L_i et à droite R_i.
2. On suppose à présent que les observations sont uniquement soumises à censure **indépendante** à droite C_i. On fait l'hypothèse que les T_i et les C_i ont pour fonction de hasard respective $h_T(t) = \alpha t^{\alpha-1}$ et $h_C(t) = \beta t^{\beta-1}$ (Weibull). Écrire la log-vraisemblance du modèle et en déduire l'équation vérifie par l'estimateur du maximum de vraisemblance de $\boldsymbol{\theta} = (\alpha, \beta)$.

Exercice 3 :

1) $S(t) = e^{-\int_0^t (a+bc^u)\,du} = \exp\left(-at - b\frac{1}{\ln(c)}(c^t-1)\right)$

$f(t) = (a+bc^t)\exp\left(-at - b\frac{1}{\ln(c)}(c^t-1)\right)$

Fonction de survie conditionnelle : $\frac{S(t+t_0)}{S(t_0)} = \exp\left(-at - \frac{bc^{t_0}}{\ln(c)}(c^t-1)\right)$

Choix raisonnable de a_0, b_0 : $q_x = q_x^{Ref}$

$\rightarrow a_0 = 1$
$b_0 = 0$

table ajustée sur la table de Ref.

Car dans modèle de Brass :

$y_x = \ln\left(\frac{q_x}{1-q_x}\right) = a\ln\left(\frac{q_x^{Ref}}{1-q_x^{Ref}}\right) + b$

$q_x(a,b) = \frac{e^{y_x(a,b)}}{1-e^{y_x(a,b)}}$

$y(x) = \ln(q_{x+1} - q_x)$

$1 - q_x = \frac{S(x+1)}{S(x)} = e^{-a - \frac{bc^x}{\ln(c)}(c-1)}$

$\gamma = \frac{b}{\ln(c)}(c-1)$

$\rightarrow \ln(1-q_x) = -a - \gamma c^x$

$\hat{=}\ q_x = -a - \gamma c^x$

$\hat{=}\ q_{x+1} = -a - \gamma c^{x+1}$

$\ln(q_{x+1} - q_x) = \ln(\gamma c^x (c-1))$

$= \ln(\gamma(c-1)) + x\ln(c)$

2. Dans la suite, on note $H(t) = \int_0^t h(u)\,du$ la fonction de hasard cumulée.
La loi de Gompertz-Makeham est usuellement utilisée pour modéliser la durée de vie humaine. Le second terme du taux de hasard (bc^t) correspond à la loi de Gompertz originale (1825) et est croissant pour $c > 1$ et $b > 0$. Il permet de traduire le vieillissement progressif de l'organisme. Le paramètre a ajouté par Makeham intègre les décès accidentels survenant aux âges plus jeunes. Si ce modèle est retenu pour la population générale, il ne permettra pas de prendre en compte les âges de la vie où le taux de hasard est potentiellement décroissant (mortalité infantile, bosses dues aux accidents chez les jeunes adultes...). Notons qu'il n'existe pas d'expression explicite pour la loi de Makeham-Gompertz pour l'espérance de vie et les moments d'ordre supérieurs.

3) La log-vraisemblance du modèle s'écrit :

$\ln(\mathcal{L}_\theta) = \sum_i [d_i \ln(h_\theta(t_i)) + \ln(S_\theta(t_i))]$

$$\ln(\mathcal{L}(a,b,c)) = \sum_{i}^{m} d_i \ln(a+bc^{y_i}) - ay_i - \frac{b}{\ln(c)}(c^{y_i}-1)$$

$$\Longrightarrow \frac{\partial \ln(\mathcal{L}(a,b,c))}{\partial a} = \sum_i^m \frac{d_i}{a+bc^{y_i}} - y_i$$

$$\frac{\partial \ln(\mathcal{L}(a,b,c))}{\partial b} = \sum_{i=1}^m \frac{d_i c^{y_i}}{a+bc^{y_i}} - \frac{1}{\ln(c)}(c^{y_i}-1)$$

$$\frac{\partial \ln(\mathcal{L}(a,b,c))}{\partial c} = \sum_{i=1}^{n} \frac{d_i y_i c^{y_i-1}}{a+bc^{y_i}} - \frac{b}{\ln(c)} y_i c^{y_i-1} + \frac{b}{\ln(c)} \times \frac{1}{c\ln(c)}(c^{y_i}-1)$$

Si on avait $X_{c_i} = X_i \mid X_i > c_i$

$S_\theta(t_i \mid e_i) = \frac{S_\theta(t_i)}{S_\theta(e_i)}$

$-\frac{d}{dt}\ln(S_\theta(t_i \mid e_i)) = h_\theta(t_i)$

$e(x) = \mathbb{E}[X - x \mid X > x] = \int_0^\infty S_x(u)\,du = \frac{1}{S(x)}\int_x^\infty S(u)\,du$

$S_x(u) = \frac{S(x+u)}{S(x)}$

$= \frac{1}{S(x)}\int_x^\infty e^{-au - \frac{b}{\ln(c)}(c^u-1)}\,du$

$\approx \frac{1}{S(x)}\sum_{y>x} S(y)$