© Théo Jalabert

Censure de type I : censure fixe

Soit un échantillon de durée de survie $(X_1, \dots, X_n)$ et $C > 0$ fixé ; la vraisemblance du modèle associé aux observat° $(T_1, D_1), \dots, (T_n, D_n)$ avec :

$$T_i = X_i \wedge C \quad \text{et} \quad D_i = \begin{cases} 1 & \text{si } X_i \le C \\ 0 & \text{si } X_i > C \end{cases}$$

possède une composante continue et une discrète elle s'écrit :

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} \underbrace{f_\theta(T_i)^{D_i}}_{\text{observat° de la sortie avant censure : c'est la densité qui intervient}} S_\theta(C)^{1-D_i}$$

fct° de survie à la date de la censure si l'obs° intervient après la censure.

Preuve :

Il suffit de calculer $P(T_i \in [t_i, t_i + dt_i], D_i = d_i)$ Comme $D_i \in \{0,1\}$, on calcule sur $[0, C]$

$$P(T_i \in [t_i, t_i+dt_i], D_i = 1) = P(X_i \wedge C \in [t_i, t_i + dt_i], X_i \le C)$$
$$= P(X_i \in [t_i, t_i + dt_i]) = f_\theta(t_i)\, dt_i$$

params liés à X. On peut toujours supposer dt_i suffisamment petit pour que $t_i + dt_i \le C$.

$$\text{et } P(T_i \in [t_i, t_i+dt_i], D_i = 0) = P(X_i \wedge C \in [t_i, t_i+dt_i], X_i > C)$$
$$= P(X_i > C)$$
$$= S_\theta(C)$$

On peut donc écrire $P(T_i \in [t_i, t_i + dt_i], D_i = d_i) = f_\theta(t_i)^{d_i} S_\theta(C)^{1-d_i}$

On peut retrouver cette express° par $P(T_i > t_i, D_i = 1) = P(X_i > t_i, X_i \le C)$
$$= \int_{t_i}^{C} f_\theta(u)\, du$$

et dans le cas $D_i = 0$ comme alors $T_i = C$ => pas de densité mais $P = S_\theta(C)$

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

Autrement : $P(T_i > t_i, D_i = 1) = P(X_i \wedge C > t_i, X_i \le C)$
$$= P(X_i > t_i, X_i \le C)$$
$$= P(t_i < X_i \le C)$$
$$= \int_{t_i}^{C} f_X(u)\, du = F_X(C) - F_X(t_i)$$
$$= 1 - S_X(C) - (1 - S_X(t_i))$$
$$= S_X(t_i) - S_X(C)$$

$$\Rightarrow -\frac{d}{dt_i} P(T_i > t_i, D_i = 1) = f_X(t_i)$$

On a $\mathcal{L}(\theta) = \prod_{i=1}^{n} f_\theta(T_i)^{D_i} S_\theta(T_i)^{1-D_i}$

$$\Rightarrow \mathcal{L}(\theta) = \prod_{i=1}^{n} S_\theta(T_i)\, h_\theta(T_i)^{D_i} \quad \text{car } f = hS$$

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

© Théo Jalabert

Censure de type III : censure aléatoire

$$T_i = X_i \wedge C_i \qquad D_i = \mathbb{1}_{X_i \leq C_i}$$

$$\begin{aligned}
\mathbb{P}(T_i > t_i, D_i = 1) &= \mathbb{P}(X_i \wedge C_i > t_i, X_i \leq C_i) \\
&= \mathbb{P}(X_i > t_i, X_i \leq C_i) \\
&= \mathbb{P}(t_i < X_i \leq C_i) \\
&= \mathbb{E}[\mathbb{P}(t_i < X_i \leq c) \mid C_i = c] \\
&= \int_{t_i}^{\infty} \left(\int_{t_i}^{c} f_X(u)\, du \right) f_C(c)\, dc \overset{\text{Fubini}}{=} \int_{t_i}^{\infty} \int_{u}^{\infty} f_C(c) f_X(u)\, dc\, du \\
&= \int_{t_i}^{\infty} f_X(u) S_C(u)\, du = S_X(t_i) S_C(u)
\end{aligned}$$

$t_i \leq c < \infty$
$t_i \leq u \leq c$

$$\Rightarrow -\frac{d}{dt_i} \mathbb{P}(T_i > t_i, D_i = 1) = f_X(t_i) S_C(t_i)$$

De même, $-\frac{d}{dt_i} \mathbb{P}(T_i > t_i, D_i = 0) = f_C(t_i) S_X(t_i)$

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

$$\begin{aligned}
\Rightarrow \mathcal{L}(\theta) &\propto \prod_{i=1}^{m} f_\theta(T_i)^{D_i} S_\theta(C_i)^{1-D_i} \\
&\propto \prod_{i=1}^{m} S_\theta(T_i) h_\theta(T_i)^{D_i}
\end{aligned}$$

Lorsque $D_i = 0$ on a $T_i = C_i$

La prise en compte de covariables (pour censure de type III)

Lorsque le modèle comporte p variables explicatives (covariables) $Z = (Z_1, \dots, Z_n)$ on fait l'hyp que la loi conditionnelle de X sachant Z dépend d'un paramètre θ.

L'échantillon observé devient une séquence de triplets (T_i, D_i, Z_i) ; on reprend l'hyp de la censure non informative ; on suppose de + que X et C sont indep conditionnellement à Z et que C est non informative pour les paramètres de la loi conditionnelle de X sachant Z.
On suppose enfin que Z admet une densité qui dépend d'un paramètre ϕ, $f_Z(z,\phi)$

Dans ces condit°, l'express° de la vraisemb devient :

$$\mathcal{L}(\theta) = \prod_{i=1}^{m} h_{\theta|Z}(T_i)^{D_i} S_{\theta|Z}(T_i) f_Z(Z_i, \phi)$$

Lorsque la loi de T sachant Z et la loi de Z n'ont pas de param en commun, on retrouve simplement ce qu'il y a au-dessus dans laquelle la loi de X est remplacée par la loi conditionnelle de X sachant Z.
Se généralise facilement au cas de covariables dépendant du temps.

© Théo Jalabert

Un autre type de censure : "arrêt au $r^{ième}$ décès" (censure de type II)

On se place dans le cas où la date de fin d'observat° n'est pas définie à l'avance mais on s'arrête à la $r^{ième}$ sortie.
→ Date de fin d'exp est aléatoire et égale à $X_{(r)}$

On dit qu'il y a censure de type II si au lieu d'observer directe^t $(X_1, \dots, X_n)$ on observe $(T_1, D_1), \dots, (T_n, D_n)$ avec :

$$T_i = X_i \wedge X_{(r)} \quad \text{et} \quad D_i = \begin{cases} 1 & \text{si } X_i = T_i \\ 0 & \text{si } X_i \neq T_i \end{cases} = \mathbb{1}_{X_i \leq X_{(r)}}$$

avec $X_{(r)}$ la $r^{ième}$ stat d'ordre de l'échantillon $(X_1, \dots, X_n)$.

Pour la vraisemblance, il convient donc de choisir les instants de r sorties parmi les n observat°

$$\Rightarrow \mathcal{L}(\theta) = \frac{n!}{(n-r)!} \left[\prod_{i=1}^{r} f_\theta(X_{(i)}) \right] S_\theta(X_{(r)})^{n-r}$$

$$= \frac{n!}{(n-r)!} \prod_{i=1}^{n} f_\theta(T_i)^{D_i} S_\theta(T_i)^{1-D_i} = \frac{n!}{(n-r)!} \prod_{i=1}^{n} h_\theta(T_i)^{D_i} S_\theta(T_i)$$

Si la loi de référence est la loi exponentielle $\Rightarrow \mathcal{L}(\theta) = \frac{n!}{(n-r)!} \theta^r e^{-\theta T}$ avec $T = \sum_{i=1}^{r} T_{(i)} + (n-r) T_{(r)}$

Troncature

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

La distrib observée dans ce cas est donc la loi conditionnelle à l'év^t $\{c < T < C\}$
La fonct° de survie tronquée s'écrit donc :

$$S(t \mid c < T < C) = \begin{cases} 1 & \text{si } t < c \\ \dfrac{S(t) - S(C)}{S(c) - S(C)} & \text{si } c \leq t \leq C \\ 0 & \text{si } t > C \end{cases}$$

La fonct° de hasard a également le support $\{c < t < C\}$ et s'écrit $h(t \mid c < T < C) = h(t) \dfrac{S(t)}{S(t) - S(C)}$ ce qui montre que l'express° de h ne dépend pas de c.

La troncature droite augmente la fonct° de hasard, et s'il n'y a que de la troncature gauche (i.e $C = +\infty$) $\Rightarrow h$ pas modifiée.

Synthèse : troncature gauche et censure aléatoire droite.

Les individus ne sont en général pas observés depuis l'origine mais depuis l'âge (ou ancienneté) atteint au début de la période d'observat° notée E_i.
La censure C_i peut être inférieure à l'âge atteint à fin de période d'observat° si sortie anticipée (résiliat° par ex)

$$\rightarrow \mathcal{L}(\theta) = \prod_{i=1}^{n} h_{\theta,E}(t_i)^{d_i} S_{\theta,E}(t_i)$$

$$\rightarrow \ln(\mathcal{L}(\theta)) \propto \sum_{i=1}^{n} d_i \ln(h_\theta(t_i)) + \ln(S_\theta(t_i)) - \ln(S_\theta(e_i))$$ Comme $h_{\theta,E}(t_i) = h_\theta(t_i)$ et $S_{\theta,E}(t_i) = \dfrac{S_\theta(t_i)}{S_\theta(e_i)}$

Si tous les individus sont observés depuis l'origine $e_i = 0 \Rightarrow \ln(\mathcal{L}(\theta)) \propto \sum_{i=1}^{n} d_i \ln(h_\theta(t_i)) + \ln(S_\theta(t_i))$

→ cf Weibull pour exemple $f(t) = \lambda\alpha t^{\alpha-1} e^{-\lambda t^{\alpha}}$

$\Rightarrow S(t) = e^{-(\lambda t)^{\alpha}}$

© Théo Jalabert

Cas particulier du modèle exponentiel, lien avec modèle de Poisson.

On considère n individus pour lesquels on fait l'hyp que sur $[x, x+1[$ la fct° de hasard sous-jacente est cte.

$\Rightarrow$ à cte près, $\ln(\mathcal{L}(\theta)) = \sum_{i=1}^{n} [d_i \ln(\theta) - \theta(t_i - e_i)] = d_x \ln(\theta) - \theta E_x$

avec $d_x = \sum_{i=1}^{d} d_i$ et $E_x = \sum_{i=1}^{d} (t_i - e_i)$

On rq que tout se passe comme si D_x qui compte le nb de sorties sur $[x, x+1[\sim \mathcal{P}(\theta E_x)$

En effet, dans ce cas, $\ln(\mathbb{P}(D_x = d)) = \text{cte} + d_x \ln(\theta) - \theta E_x$.