Une vision de l'estimateur de Kaplan–Meier comme estimateur de maximum de vraisemblance empirique

Mathieu Ribatet

Soient $(T_{*,1}, C_1), \dots, (T_{*,n}, C_n)$ n réalisations i.i.d. où T_* a pour densité f, fonction de survie S et C de densité g et de fonction de survie H. Ainsi une observation t_i a pour contribution à la vraisemblance

$$f(t_i;\theta)H(t_i) \quad \text{si } \delta_i = 1, \qquad g(t_i)S(t_i;\theta) \quad \text{si } \delta = 0.$$

Sous l'hypothèse (souvent raisonnable) que la censure ne dépend pas de θ, la vraisemblance (à une constante multiplicative près) est donnée par

$$L(\theta) = \prod_{i=1}^n f(t_i;\theta)^{\delta_i} S(t_i;\theta)^{1-\delta_i} = \prod_{i=1}^n \left\{ \frac{f(t_i;\theta)}{S(t_i;\theta)} \right\}^{\delta_i} S(t_i;\theta) = \prod_{i=1}^n \lambda(t_i;\theta)^{\delta_i} S(t_i;\theta).$$

L'approche par vraisemblance empirique consiste à supposer que le taux de panne est de la forme

$$\lambda(t) = \sum_{j=1}^J \lambda_j \delta_{t_{(j)}}(t), \qquad t > 0,$$

où $t_{(1)} < t_{(2)} < \cdots < t_{(J)}$ les J instants de décès distincts issu de notre n–échantillon.

Puisque le taux de panne est défini par $\lambda(t) = f(t)/\Pr(T_* \geq t)$, on a pour notre cas bien précis que $\lambda_j = \Pr(T_* = t_{(j)} \mid T_* \geq t_{(j)})$ ou encore que $1 - \lambda_j = \Pr(T_* > t_{(j)} \mid T_* \geq t_{(j)})$.

De plus pour tout $t \in (t_{(j)}, t_{(j+1)}]$ on a

$$\begin{aligned}
S(t) &= \Pr(T_* > t) = \Pr(T_* > t_{(j)}) = \Pr(T_* > t_{(j)} \mid T_* \geq t_{(j)}) \Pr(T_* \geq t_{(j)}) \\
&= (1-\lambda_j) \Pr(T_* \geq t_{(j)}) \\
&= (1-\lambda_j) \Pr(T_* > t_{(j-1)}) \\
&= (1-\lambda_j) \Pr(T_* > t_{(j-1)} \mid T_* \geq t_{(j-1)}) \Pr(T_* \geq t_{(j-1)}) \\
&= (1-\lambda_j)(1-\lambda_{j-1}) \Pr(T_* \geq t_{(j-1)}) \\
&= \cdots \\
&= \prod_{j \colon t_{(j)} < t} (1-\lambda_j),
\end{aligned}$$

avec la convention que $\lambda_0 = 0$ de sorte que pour tout $t < t_{(1)}$, on a bien $S(t) = 1$.

En utilisant ce résultat dans l'expression de $L(\theta)$, où ici $\theta = (\lambda_1, \dots, \lambda_J)^\top$, on trouve alors

$$L(\theta) = \prod_{i=1}^n \lambda(T_i)^{\delta_i} S(T_i) = \prod_{j=1}^J \lambda_j^{d_j} \times \prod_{i=1}^n \prod_{j \colon T_{(j)} < T_i} (1-\lambda_j) = \prod_{j=1}^J \lambda_j^{d_j} \times \prod_{j=1}^J (1-\lambda_j)^{r_j - d_j},$$

puisque dans le deuxième produit chaque λ_j apparaît $|\{i \colon T_i > T_{(j)}\}|$ soit exactement $r_j - d_j$.

La log–vraisemblance est donc

$$\ell(\theta) = \sum_{j=1}^J \left\{ d_j \ln \lambda_j + (r_j - d_j) \ln(1-\lambda_j) \right\}.$$

Pour tout $j \in \{1, \ldots, J\}$, on a

$$\begin{aligned} \frac{\partial \ell(\theta)}{\partial \lambda_j} = 0 &\iff \frac{d_j}{\lambda_j} + \frac{r_j - d_j}{1 - \lambda_j} = 0 \\ &\iff (1 - \lambda_j) d_j - \lambda_j (r_j - d_j) = 0 \\ &\iff \lambda_j = \frac{d_j}{r_j}. \end{aligned}$$

Ainsi l'estimateur du maximum de vraisemblance (empirique) pour $\theta = (\lambda_1, \ldots, \lambda_J)^\top$ est donné par $\hat{\theta} = (d_1/r_1, \ldots, d_J/r_J)^\top$ et sa matrice variance–covariance est estimée via la matrice observée d'information de Fisher

$$\widehat{\mathrm{Var}(\hat{\theta}_j)} = -\left(\frac{\partial^2 \ell(\hat{\theta})}{\partial \lambda_i \partial \lambda_j} \right)^{-1} = \begin{bmatrix} \frac{r_1}{\hat{\lambda}_1 (1 - \hat{\lambda}_1)} & & 0 \\ & \ddots & \\ 0 & & \frac{r_J}{\hat{\lambda}_J (1 - \hat{\lambda}_J)} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{\hat{\lambda}_1 (1 - \hat{\lambda}_1)}{r_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\hat{\lambda}_J (1 - \hat{\lambda}_J)}{r_J} \end{bmatrix}.$$

On notera au passage que le $\hat{\lambda}_1, \ldots, \hat{\lambda}_J$ sont asymptotiquement mutuellement indépendants.