Parametric Survival-Data Likelihood

Suppose that underlying (latent) waiting times X_i, C_i to death and censoring are **independent** for each subject $i = 1, \ldots, n$; densities are $f_X = f_X(\cdot, \vartheta)$, f_C and survival functions are $S_X = S_X(\cdot, \vartheta)$, S_C. For observed data $(T_i, \Delta_i) = (\min(X_i, C_i),\ I_{[X_i \le C_i]}$:

$$Lik(\vartheta) \;=\; \prod_{i=1}^{n} \{(f_X(T_i, \vartheta)\, S_C(T_i))^{\Delta_i}\, (f_C(T_i)\, S_X(T_i, \vartheta))^{1-\Delta_i}\}$$

Censoring density f_C is unknown but does not contain parameters to be estimated, so dropping factors f_C, S_C and writing $f_X = h_X S_X = h_X\, e^{-H_X}$ leaves

$$\text{logLik}(\vartheta) \;=\; \sum_{i=1}^{n} (\Delta_i\, \log h_X(T_i,\, \vartheta) \,-\, H_X(T_i, \vartheta))$$

$$= \int \{\, \log h_X(t, \vartheta)\, dN(t) \;-\; \sum_{i=1}^{n} I_{[T_i \ge t]}\, h_X(t, \vartheta)\, dt\}$$

where

$$N(t) \;=\; \sum_{i=1}^{n} \Delta_i\, I_{[T_i \le t]}$$

defines the *observed death counting process*, and the *at-risk process* is

$$Y(t) \;=\; \sum_{i=1}^{n} I_{[T_i \ge t]}$$

MLE's from Parametric Survival Likelihood

$$\text{logLik}(\vartheta) \;=\; \int \left(\log h_X(t,\vartheta)\, dN(t) \;-\; Y(t)\, h_X(t,\vartheta)\, dt \right)$$

leading to likelihood score equation

$$\mathbf{0} \;=\; \int \nabla_\vartheta \, \log h_X(t,\vartheta)\, (dN(t) \;-\; Y(t)\, h_X(t,\vartheta)\, dt)$$

with solution $\hat{\vartheta}$ satisfying in large samples:

$$-\nabla^{\otimes 2}\, \text{logLik}(\vartheta_0)\, (\hat{\vartheta} - \vartheta_0) \;\approx\; \nabla \text{logLik}(\vartheta_0)$$

(using notation $\mathbf{v}^{\otimes 2} = \mathbf{v}\mathbf{v}'$), which leads to:

$$\hat{\vartheta} \;-\; \vartheta_0 \;\approx\; \Big(-\int \nabla^{\otimes 2} \, \log h(t,\vartheta_0)\, dM(t) +$$

$$\int (\nabla \log h(t,\vartheta_0))^{\otimes 2}\, Y(t)\, h(t,\vartheta_0)\, dt \Big)^{-1} \int \nabla \log h(t,\vartheta_0)\, dM(t)$$

where $dM(t) \;=\; dN(t) - Y(t)\, h_X(t,\vartheta_0)\, dt$ is the integrator for martingale *stochastic integrals* and will be seen to have the property that when the model with hazard $h_X(t,\vartheta_0)$ actually governs the data, for each square-integrable function g (wrt $f(t,\vartheta_0)\, dt = h(t,\vartheta_0) S(t,\vartheta_0) dt$) and all large n,

$$E \Big(\int g(t)\, dM(t) \Big)^2 \;=\; \mathcal{O}(n \int g^2(t)\, f(t,\vartheta_0)\, dt)$$

Under the $f(t,\vartheta_0)$ *model, 1st term in* $\nabla^{\otimes 2}$*logLik above is* $\mathcal{O}(\sqrt{n})$*, can be ignored because 2nd is* $\mathcal{O}(n)$.

Specialization to Weibull Log-Lik

Likelihood in Weibull(λ, γ) case $[h(t, \vartheta) = \lambda\gamma\, t^{\gamma-1}]$ gives

$$\int \begin{pmatrix} 1/\lambda \\ 1/\gamma + \log(t) \end{pmatrix} (dN(t) - Y(t)\,\lambda\gamma\, t^{\gamma-1}\, dt) \;=\; 0$$

First equation uniquely determines $\hat{\lambda}$ in terms of $\hat{\gamma}$ by

$$\sum_{i=1}^{n} \Delta_i \;=\; N(\infty) \;=\; \int \sum_{i=1}^{n} I[T_i \geq t]\, d(\lambda\, t^{\gamma}) \;=\; \lambda \sum_{i=1}^{n} T_i^{\gamma}$$

Second eq'n becomes

$$\int \log(t)\, dN(t)/N(\infty) \;=\; -\gamma^{-1} + \sum_{i=1}^{n} T_i^{\gamma}\, \log(T_i) \,/\, \sum_{i=1}^{n} T_i^{\gamma}$$

and right-hand side is strictly $\nearrow$ in γ.

RESULT. For large n, if Weibull(λ_0, γ_0) model holds, then

$$\sqrt{n}\,(\hat{\vartheta} - \vartheta_0) \;\approx\; \mathcal{N}\left(0, \left\{\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1/\lambda_0 \\ 1/\gamma_0 + \log(T_i) \end{pmatrix}^{\otimes 2} \Delta_i \right\}^{-1}\right)$$

If not, large-sample limit ϑ_0 *of* $\hat{\vartheta}$ *still exists and* **robust** *misspecified-model variance can also be estimated simply.* (White 1982)

Exponential Case

Consider data satisfying Weibull$(\lambda_a, 1)$ = Expon(λ_a) among all surviving uncensored to time a and right-censored at time $b = a + \delta$ $(a \geq 0,\ \delta \leq \infty)$, i.e. :

$$(T_i^*, \Delta_i^*) \equiv (\min(b, T_i), \Delta_i\, I_{[T_i \leq b]}) \ : \quad 1 \leq i \leq n, \quad T_i \geq a$$

This is *left-truncated right-censored* dataset for which maximum likelihood estimator of hazard λ_a specializes (from Weibull formulas with $\gamma = 1$ fixed) to:

$$\sum_{i=1}^n I_{[a \leq T_i \leq b]}\, \Delta_i \,/\, \sum_{i=1}^n I_{[a \leq T_i]} \min(T_i - a, b - a)$$

or number of observed deaths in interval divided by total time on test within exposure-interval $[a, b]$.

Result: $\hat{\lambda}_a \ = \ (N(b) - N(a)) \,/\, \int_a^b Y(t)\, dt$

since $\int_a^b Y(t)\, dt \ = \ \sum_{i=1}^n I_{[T_i \geq a]} \min(T_i - a, b - a)$

This says for very small δ that the instantaneous hazard rate $h(a)$ at a is generally estimated by

$$\hat{h}(a) \ = \ (N(a + \delta) - N(a)) \,/\, (\delta\, Y(a))$$