Misspecification in Survival and Cure Models

Emil Aas Stoltenberg
Joint work with Per A. Mykland

Department of Mathematics
University of Oslo

Building Bridges at Bislett
May 2017

One story in two settings

$t_1^*, \ldots, t_n^*$ independent event times with absolutely continuous distribution function $F(t)$, density $f(t)$, and hazard rate $\alpha(t) = f(t)/(1 - F(t))$.

(A) Survival analysis

$$S(t) = \Pr\{t_i^* \geq t\} = 1 - F(t).$$

Everybody dies: $\lim_{t\to\infty} S(t) = 0$.

(B) Cure models.

$$S_{\text{pop}}(t) = 1 - \pi + \pi S(t),$$

A fraction $1 - \pi$ never die:

$$\lim_{t\to\infty} S_{\text{pop}}(t) = 1 - \pi > 0.$$

Both π and S can be functions of covariates.

This talk: Start with (A), then outline how we attempt to transfer the techniques used in (A) to the more advanced (B).

(A) A counting process approach

Finite time interval $[0, T]$. We observe

$$t_i = \min\{t_i^*, c_i\}, \quad \delta_i = I\{t_i^* \leq c_i\},$$

with $c_i \sim G$ independent of the lifetimes. Define a counting process and a left-continuous at risk process

$$N(t) = \sum_{i=1}^{n} I\{t_i \leq t, \delta_i = 1\}, \qquad Y(t) = \sum_{i=1}^{n} I\{t_i \geq t\}.$$

Consider two contiguous measures P_0 and $P_{A,n}$ with associated intensity processes

$$\begin{aligned} P_0 &: \ \lambda_0(t) = Y(t)\alpha(t) \\ P_{A,n} &: \ \lambda_{A,n}(t) = Y(t)(\alpha(t) + r(t)/\sqrt{n}), \end{aligned}$$

with $r(t)$ a continuous function.

(A) The likelihood ratio

By Jacod's theorem (Andersen et al., 1993), we can write

$$L_t^n = \frac{dP_{A,n}}{dP_0}\bigg|_{\mathcal{F}_t} = L_0 \left\{ \prod_{s \le t} \frac{\lambda_{A,n}(s)^{\Delta N(s)}}{\lambda_0(s)^{\Delta N(s)}} \right\} \frac{e^{-\int_0^t \lambda_{A,n}(u)\,du}}{e^{-\int_0^t \lambda_0(u)\,du}}.$$

Define the P_0-martingales $M(t) = N(t) - \int_0^t Y(s)\alpha(s)\,ds$ and

$$M_t^n = \frac{1}{\sqrt{n}} \int_0^t \frac{r(s)}{\alpha(s)}\,dM(s).$$

Then

$$\log L_t^n = M_t^n - \frac{1}{2}\langle M^n, M^n \rangle_t + o_p(1).$$

Under suitable conditions, M_t^n converges in distribution to $\int_0^t r(s)/\alpha(s)\,dW_s$, where W is a mean zero Gaussian process with independent increments and $\mathrm{Var}(dW_s) = y(s)\alpha(s, \theta_0)\,ds$. The quadratic variation

$$\langle M^n, M^n \rangle_t \to_p \langle W, W \rangle_t = \int_0^t y(s)\alpha(s)\,ds,$$

where $y(s)$ is the limit in probability of $Y(s)/n$.

(A) Parametric and nonparametric wide models

Let P_0 be fully parametric, with associated intensity $\lambda_0(t,\theta) = Y(t)\alpha(t,\theta)$. The bigger sequence of models $P_{A,n}$ have associated intensity processes:

(i) Nonparametric extension,

$$\lambda_{A,n}(t,\theta,r) = Y(t)(\alpha(t,\theta) + r(t)/\sqrt{n}),$$

where r is an unknown function; or

(ii) Parametric extension,

$$\lambda_{A,n}(t,\theta,\gamma) = Y(t)(\alpha(t,\theta,\gamma) + r(t)/\sqrt{n}),\ \gamma = \gamma_0 + h/\sqrt{n},$$

with $r(t) = h^\top \partial\alpha(t,\theta,\gamma_0)/\partial\gamma$ and $\alpha(t,\theta,\gamma_0) = \alpha(t,\theta)$, i.e. the local misspecification framework of Claeskens and Hjort (2003). Let $\hat{\theta}_{\mathrm{ml}}$ denote the MLE based on small model P_0; then

$$\sqrt{n}(\hat{\theta}_{\mathrm{ml}} - \theta_0) \to_d h\Sigma^{-1}\tau + \Sigma^{-1/2}\mathrm{N}_p(0, I_p),$$

under $P_{A,n}$. Not obvious how to estimate the bias when the bigger model is nonparametric.

(A) Nelson-Aalen under the big model

Let $A_n(t)$ denote the cumulative hazard under the big model. Under $P_{A,n}$,

$$\sqrt{n}(\hat{A}(t) - A_n(t)) \to_d \mathrm{N}(0, \sigma_t^2),$$

$$\sqrt{n}(A(t, \hat{\theta}_{\mathrm{ml}}) - A_n(t)) \to_d \kappa(t)^\top \Sigma^{-1}\tau - R(t) + \mathrm{N}(0, v_t^2),$$

with $\hat{A}(t)$ the Nelson-Aalen estimator, $R(t) = \int_0^t r(s)\,ds$ and $\sigma_t^2 \geq v_t^2$. The quantity

$$Z_n(t) = \sqrt{n}(A(t, \hat{\theta}_{\mathrm{ml}}) - \hat{A}(t)),$$

converges in distribution to $\kappa(t)^\top \Sigma^{-1}\tau - R(t) + \mathrm{N}(0, \sigma_t^2 - v_t^2)$ under $P_{A,n}$, but is not consistent for the bias of $A(t, \hat{\theta}_{\mathrm{ml}})$. Since $Z_n(t)$ is $\mathrm{N}(0, \sigma_t^2 - v_t^2)$ under P_0 and $\sigma_t^2 - v_t^2$ can be estimated, $Z_n(t)$ provides a test of the smaller parametric model.

Current work:

(i) Estimate bias of $\hat{\theta}_{\mathrm{ml}}$ under nonparametric alternative.

(ii) Stochastic intensity processes.

(iii) Estimators of functionals of A.

(B) The cure model

Usual assumption in survival analysis: Everyone experiences the event,

$$S(t) \to 0 \quad \text{when} \quad t \to \infty.$$

Not always realistic:

- Some patients are fully cured after treatment.
- Some people will never be convinced.
- Some kids are born immune to a given disease.
- A binary variable whose value is not known for all subjects at the time of analysis (Farewell, 1977).

Two populations:

- Cured individuals.
- Susceptible individuals.

As before, we observe $(t_1, \delta_1, x_1), \ldots, (t_n, \delta_n, x_n)$. Note that

$$\delta_i = 1 \Rightarrow \text{Susceptible}, \quad \text{but} \quad \delta_i = 0 \Rightarrow \text{don't know}.$$

(B) The cure model

Think of a binary variable

$$U_i = \begin{cases} 1, & \text{Susceptible}, \\ 0, & \text{Cured}, \end{cases}$$

with $\Pr(U_i = 1) = \pi(x_i^\top \beta)$, and $\pi(z) = e^z/(1+e^z)$. Event times t_i^* have frailty type hazard $U_i\alpha(t)$, and the population survival function is

$$S_{\text{pop}}(t) = 1 - \pi(x_i) + \pi(x_i)S(t).$$

Note that S_{pop} is not proper: $S_{\text{pop}}(t) \underset{t\to\infty}{\to} 1 - \pi(x_i) > 0$. The likelihood for β and $A(t) = \int_0^t \alpha(s)\,ds$ is

$$L_n(\beta, A) = \prod_{i=1}^n \left\{\pi(x_i)\alpha(t_i)e^{-A(t_i)}\right\}^{\delta_i}\left\{1 - \pi(x_i) + \pi(x_i)e^{-A(t_i)}\right\}^{1-\delta_i}.$$

Large sample behaviour of estimators with $A_i(t) = \int_0^t \alpha_0(s)e^{z_i'\gamma}\,ds$ (the Cox-model) have been studied by Fang et al. (2005) and Lu (2008).

(B) A counting process approach to the cure model

Advantages:

(i) Amenable to the study of model misspecification/inference under contiguous alternatives;

(ii) Possible to study the model with a stochastic intensity process.

Define

$$N_i(t) = I\{t_i \leq t, \delta_i = 1\}, \qquad Y_i(t) = I\{t_i \geq t\}.$$

As before, we consider contiguous measures P_0 and $P_{A,n}$ with associated intensity processes

$$\lambda_0(t) = Y(t)\alpha(t, \theta_0), \quad \lambda_{A,n}(t) = Y(t)(\alpha(t, \theta_0) + r(t)/\sqrt{n}),$$

Let $\mathcal{F}_t = \sigma(N_i(s), Y_i(s), s \leq t, 1 \leq i \leq n)$ be the filtration generated by the observable data.

(B) The likelihood ratio

Q_0 and $Q_{A,n}$ are probability measures defined on $\mathcal{F}_t$. The associated intensity processes are

$$\lambda_{0,i}(t) = E_{Q_0}[U_i \mid \mathcal{F}_{t-}]Y_i(t)\alpha(t,\theta_0),$$
$$\lambda_{A,i}(t) = E_{Q_A}[U_i \mid \mathcal{F}_{t-}]Y_i(t)(\alpha(t,\theta_0) + r(t)/\sqrt{n}).$$

The likelihood ratio is

$$L_t^n = \frac{dQ_{A,n}}{dQ_0}\bigg|_{\mathcal{F}_t} = L_0 \prod_{i=1}^{n} \left\{ \prod_{s \le t} \frac{\lambda_{A,i}(s)^{\Delta N_i(s)}}{\lambda_{0,i}(s)^{\Delta N_i(s)}} \right\} \frac{e^{-\int_0^t \lambda_{A,i}(u)\,du}}{e^{-\int_0^t \lambda_{0,i}(u)\,du}}$$

We show that

$$\log L_t^n = Z_{t,n} - \frac{1}{2}\langle Z_n, Z_n\rangle_t + o_p(1), \qquad (1)$$

where $Z_{t,n}$ is a martingale and $\langle Z_n, Z_n\rangle_t$ its quadratic variation. As before, $\log L_t^n$ can be used to derive estimators and study their performance under local misspecification.

(B) Limit of likelihood ratio

Define the Q_0 martingales

$$\mu_{n,t}^{(1)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t \frac{1}{\alpha(s,\theta_0)} \, dM_i(s),$$

$$\mu_{n,t}^{(2)} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t E_{Q_0}[1 - U_i \mid \mathcal{F}_{s-}] dM_i(s),$$

where

$$dM_i(s) = dN_i(s) - Y_i(s) E_{Q_0}[U_i \mid \mathcal{F}_{s-}] \alpha(s,\theta_0) \, ds.$$

The martingales $(\mu_{n,t}^{(1)}, \mu_{n,t}^{(2)})$ converge to a Gaussian process $(\mu_t^{(1)}, \mu_t^{(2)})$. The limit of the log likelihood ratio is

$$\begin{aligned}\log L_t = & \int_0^t r(s) d\mu_s^{(1)} + \int_0^t R(s-) d\mu_s^{(2)} - \frac{1}{2} \int_0^t r(s)^2 d\langle \mu^{(1)}, \mu^{(1)} \rangle_s \\ & - \frac{1}{2} \int_0^t R(s-)^2 d\langle \mu^{(2)}, \mu^{(2)} \rangle_s + \int_0^t r(s) R(s-) d\langle \mu^{(1)}, \mu^{(2)} \rangle_s .\end{aligned}$$

(B) Use of $\log L_t^n$ and more

Let $U_1, \ldots, U_n$ be iid Bernoulli random variables. The oracle estimator $\tilde{A}(t)$ of Lu (2008) is given by

$$\tilde{A}(t) = \int_0^t \frac{dN(s)}{Y(s) E_{Q_0}[U \mid \mathcal{F}_{s-}]}.$$

Under $Q_{A,n}$,

$$\sqrt{n}(\tilde{A}(t) - A_n^*(t,\theta)) \to_d \text{bias} + \mathrm{N}(0, \sigma_t^2),$$

where bias $= -\int_0^t E_{Q_0}[1 - U \mid \mathcal{F}_s] R(s-)\alpha(s,\theta_0)\, ds$.

Next step:

$$\frac{dQ_{A',\beta'}}{dQ_{A,\beta}} = \frac{dQ_{A',\beta}}{dQ_{A,\beta}} \frac{dQ_{A',\beta'}}{dQ_{A',\beta}},$$

where

$$\log \frac{dQ_{A',\beta'}}{dQ_{A',\beta}} = X_{t,n} - \frac{1}{2}\langle X_n, X_n \rangle_t + o_p(1),$$

and

$$X_{t,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t u^\top x_i E[1 - U_i \mid \mathcal{F}_{s-}]\, dM_i(s).$$

Bibliography

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer.

Claeskens, G. and Hjort, N. L. (2003). The Focused Information Criterion. *Journal of the American Statistical Association*, 98:900–916.

Fang, H.-B., Li, G., and Sun, J. (2005). Maximum Likelihood Estimation in a Semiparametric Logistic/Proportional-Hazards Mixture Model. *Scandinavian Journal of Statistics*, 32:59–75.

Farewell, V. T. (1977). A model for a binary variable with time-censored observations. *Biometrika*, pages 43–46.

Lu, W. (2008). Maximum likelihood estimation in the proportional hazards cure model. *Annals of the Institute of Statistical Mathematics*, 60:545–574.