3-Lecture Minicourse on Statistics of Survival Data

Eric Slud

I. (11/6) **Death Hazards & Competing Risks**

Concepts:

(i) Statistical Estimation as mathematical problem,

(ii) Identifiability, nonparametric vs. nonparametric.

II. (11/13) **Population Cohorts & Martingales**

Concepts:

(iii) Counting process models,

(iv) "Innovations" and Statistics.

III. (11/20) **Models and Likelihoods with ∞-Dimensional Parameters**

Concepts:

(v) Nuisance parameters,

(vi) Asymptotic Relative Efficiency.

Lecture Slides (incl. annotated references) at :

www.math.umd.edu/∼evs/SurvSlid3.pdf

Parametric vs. Nonparametric Trade-off

Return to survival-data setting of the first lecture to focus the question of how much 'efficiency' is lost by **nonparametric** statistical estimation of survival probability.

Data: $T_i = \min(X_i, C_i),\ \Delta_i = I_{[X_i \le C_i]},\ Z_i,\ 1 \le i \le n$

event time, death-indicator, treatment-grp indicator

First Objective: estimation of $P(X_1 > t)$ including 95% Confidence Interval, under assumption either of indep. X_i, C_i or a more detailed parametric model.

Compare estimates based on popular parametric model

- **Exponential** which says $f_X(x) = \lambda e^{-\lambda x},\ x > 0$, or more general
- **Weibull** saying $f_X(x) = \lambda\,\gamma\,x^{\gamma-1}\,e^{-\lambda x^\gamma},\ x > 0$

vs. Kaplan-Meier estimate (no other assumptions).

Methodology: statistical theory provides asymptotic prob. dist'n for estimator and 95% confidence interval in each setting, which can be compared through σ :

$$\sqrt{n}\ (\tilde{p} - P(X_1 > t))\ \xrightarrow{\mathcal{D}}\ \mathcal{N}(0,\, \sigma^2)$$

$$P(X_1 > t)\ \in\ \left(\tilde{p}\ -\ 1.96\,\frac{\sigma}{\sqrt{n}},\, \tilde{p}\ +\ 1.96\,\frac{\sigma}{\sqrt{n}}\right)$$

References

Background:

Andersen, Borgan, Gill, and Keiding (1993) **Statistical Models based on Counting Processes**

Cox, D.R. (1972) *Jour. Roy. Statist. Soc.* **B**

R. Miller (1980) **Survival Analysis**

Sources for Current Lecture:

R. Miller (1983) "What price Kaplan-Meier ?" *Biometrics*, param. vs. nonparam. efficiency

Slud & Kong (1997) *Biometrika* treatment effectiveness testing using 'adaptively' fitted misspecified Cox models

Slud & Korn (1997) *Biometrika* testing in 2-grp case w ∞-dim nuisance parameters in setting with 'post-randomization' variables

Slud & Vonta (2002) Consistency of the NPML estimator in the right-censored transformation model. *Preprint, available from web-page.*

Parametric Max. Likelihood Th'y

Under general conditions satisfied here, with $\vartheta = (\lambda, \gamma)$ in Weibull case (which incl. Exponential when $\gamma = 1$),

$$\hat{\vartheta} \;=\; \arg\max_{\vartheta} \sum_{i=1}^{n} \log f_{T,\Delta}(T_i, \Delta_i)$$

is 'optimal' asympt. normal with covariance matrix the inverse of

$$\int \sum_{j=0}^{1} \Big(- \nabla_{\vartheta}^{\otimes 2} \log f_{T,\Delta}(t,j) \Big) \; f_{T,\Delta}(t,j)\, dt$$

In this setting

$$f_{T,\Delta}(t,j,\vartheta) \;=\; \begin{cases} f_X(t,\vartheta)\, S_C(t) & \text{if} \quad j = 1 \\ S_X(t,\vartheta)\, f_C(t) & \text{if} \quad j = 0 \end{cases}$$

Censoring dist.'n unknown ('nuisance parameter') but not depending on ϑ so ignored in likelihood. Function to maximize in $\vartheta = (\lambda, \gamma)$ becomes

$$\sum_{i=1}^{n} \; \{\log(\lambda\gamma) \,+\, \Delta_i\, (\gamma - 1)\, \log(T_i) \,-\, \lambda\, T_i^{\gamma}\}$$

Covariances found from integral which assumes specific censoring dist. Then for fixed t regard $S_X(t) = \exp(-\hat{\lambda}\, t^{\hat{\gamma}})$ as known smooth function $g(\vartheta)$ estimated by $\tilde{p} = g(\hat{\vartheta})$. By linearization ('Delta method')

$$\text{asympt. Var}(\sqrt{n}(\tilde{p} - p)) \;=\; \nabla_{\vartheta}'\, g(\vartheta)\; \text{avar}(\hat{\vartheta})\; \nabla_{\vartheta}\, g(\vartheta)$$

Nonparametric Kaplan-Meier Th'y

Recall that $S_X(t) = \exp(-\int_0^t h_X(x)\,dx)$. Define the **cumulative hazard function**

$$H_X(t) \;=\; \int_0^t h_X(x)\,dx \;=\; -\,ln(S_X(t))$$

KM estimator of $S_X(t)$ from survival data is equiv. to

$$\hat{H}_X(t) \;=\; \int_0^t \frac{dN(t)}{Y(t)}$$

where

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

Recall from last time: can view survival on $(t, t+\delta)$ for each surviving individual as an indep. coin-toss: failure occurs with prob. $\approx \delta \cdot h_X(t)$ each, so overall prob. of an observed failure is $\delta \cdot Y(t)\,h_X(t)$. Hence

$$N(t) - \int_0^t Y(x)\,h_X(x)\,dx \qquad \text{is a martingale, as is}$$

$$\sqrt{n}\,(\hat{H}_X(t) - H_X(t)) \;=\; \sqrt{n}\int_0^t \frac{dN(x) - h_X(x)\,Y(x)\,dx}{Y(x)}$$

From this, can prove asympt. normality and find variance formula, leading to

$$\sqrt{n}\,\left(\hat{S}_X^{KM}(t) - S_X(t)\right) \;\xrightarrow{\mathcal{D}}\; \mathcal{N}\left(0,\; S_X^2(t)\int_0^t \frac{f_X(x)\,dx}{S_X^2(x) S_C(x)}\right)$$

Efficiency Comparison

Since Conf. Int. widths are proportional to $\sigma_{est}/\sqrt{n}$, equal widths can be achieved if another estimator with *avar* σ^2_{alt} is applied on sample of size n_{alt}, where

$$\frac{n_{alt}}{n} \;=\; \mathbf{ARE}(\hat{\vartheta}_{alt},\, \hat{\vartheta}_{est}) \;=\; \sigma^2_{alt}/\sigma^2_{est}$$

Miller's (1983) ARE comparisons (Weibull vs KM) — in case of Expon censoring — are:

ARE's of KM versus parametric MLE of survival prob. at 3 quantiles $0.5,\, 0.25, 0, 0.1$ *for Weibull* S_X *and Exponential* S_C*. Shape parameter* $= \gamma$*.*

			Quantiles		
γ	Known	Cens. %	Med.	Upper Quartile	Upper Decile
1	Y	50	.64	.51	.21
		25	.56	.64	.46
		0	.48	.64	.59
2	Y	50	.57	.58	.40
		25	.52	.63	.52
		0	.48	.64	.59
2	N	50	.60	.62	.56
		25	.63	.63	.62
		0	.66	.64	.65

Multiplicative Intensity Model

Cox (1972), Aalen (1978) introduced the class of models

$$E(N(t+dt) - N(t) \,|\, Z,\, (Y(s), V(s) :\, s < t))$$
$$=\; Y(t-)\, e^{\beta' Z + \gamma' V(t-)}\, \lambda(t)\, dt$$

Idea: parameters (β, γ) to be fitted describe effect on prognosis of individual subjects, while the (infinite-dimensional) **nuisance hazard function** $\lambda(t)$ describes the general background population. Exponent usable as *prognostic index*.

Research Topics Related to Today's Lecture:

- *Theoretical:* large-sample theory of *efficient* estimators for *semiparametric* models like these with ∞-dim nuisance parameters. Efron (1977), Johansen (1983) and others proved *efficiency* of Cox's (1972) estimator of $\hat{\beta}$ based on maximizing *logLik* with $\Lambda(t) = \int_0^t \lambda(s)\, ds$ replaced by

$$\hat{\Lambda}(t) \;=\; \int_0^t \left\{ \sum_i Y_i(s) e^{\beta' Z_i + \gamma' V_i(s-)} \right\}^{-1} dN(s)$$

which amounts to maximizing *Partial Likelihood*:

$$\prod_{i:\, \Delta_i = 1} \left\{ e^{\beta' Z_i + \gamma' V_i(T_i -)} \Big/ \sum_{j:\, T_j \ge T_i} e^{\beta' Z_j + \gamma' V_i(T_j -)} \right\}$$

- *Misspecified Cox Models* : can fit Cox model for adjusting treatment comparisons (like PBC example, last lecture), even when the model is not valid. (Lin & Wei 1989; Slud & Kong 1997)

- *Variant Model: Pop'n Subgps w Related Parameters*: Slud & Korn (1997) studied the model:

  $Z_{1,i}$ = treatment group indicator

  $Z_{2,i}$ = 'post-randomization' indicator (e.g. indicator of initial tumor shrinkage within 3 mos. after treatment)

  $$h_{X|Z}(t|\mathbf{z} = (j,k)) \;=\; e^{j\,\beta_k}\,\lambda_k(t)$$

  with $\beta_i,\ \beta_2,\ \lambda_1,\ \lambda_2$ unknown.

- Kopylev (1997 PhD thesis) studied estimation of $S_X(t)$ when multiplicative Intensity model holds with time-dependent covariates $V_i(t)$. *Moderate-sample trials in which V_i summarizes patient-management regime are a growth area for 'data mining'.*

Misspecified and Adaptive Analysis

Setting: Two-group trial, with covariates; assume treatment indicator Z_i independent of covariates given $Y_i(t) = 1$ (eg random treatment allocation). Can calculate asympt. variance *under the null hypothesis* $\vartheta = 0$ *of the coeff* $\hat{\vartheta}$ *of* Z_i *in Cox-model incl. variables* W_i*: which is still approx. normal with mean* 0.

Estimate using 'working model'

$$h_{X|Z,W}(t|z,w) \; = \; e^{\vartheta z + \beta W} \, \lambda(t)$$

Assume only that *some* model $e^{\vartheta z} \, \lambda(t, V)$ holds.

Can analyze *and estimate* from data with Z_i masked:

ARE ratio $\quad avar_{\text{work}}(\hat{\vartheta}) / avar_{\text{true}}(\hat{\vartheta})$

Can do this for several models, covariate-sets W, and choose the best one based on Z*-masked data;* **then** *test* $\vartheta = 0$ *using* $\hat{\vartheta}$ *in the best of these models !*

FDA as regulatory authority still needs persuading about validity of this approach, developed in Kong & Slud (1997).

Variant Models, ∞-dim Nuisance

$Z_{1,i}$ = treatment group indicator (half in each grp)

$Z_{2,i}$ = 'post-randomization' indicator (e.g. indicator of initial tumor shrinkage within 3 mos. after treatment)

$$h_{X|Z}(t|\mathbf{z} = (j,k)) \;=\; e^{j\,\beta_k}\,\lambda_k(t)$$

with β_i, β_2, λ_1, λ_2 unknown (Slud & Korn 1997).

ARE comparison. Nonparam. 2-group problem if $Z_{i,2}$ ignored, Kaplan-Meier estimator for each gp $S_{X|Z_1}(t|z = j)$. Alternatively, estimate unknowns in Cox model for each $Z_{2,i} = k$ group, to obtain difference of survival curves. Can calculate *avar* of $\Delta(t) \equiv S_{X|Z_1=1}(t) - S_{X|Z_1=0}(t)$ both ways, form **ARE** ratio !

Results: either using data beyond t_0, or not.

ARE's of KM versus model-based estimator of $\Delta(t_0)$, *uncensored case.*

$S_{X\|z=1}(t_0)$	$S_{X\|Z=0}(t_0)$	ARE	ARE_{trunc}
.9	.9	.10	1.00
.9	.5	.50	0.98
.9	.1	.76	0.94
.5	.1	.60	0.90
.5	.5	.48	0.96
.1	.1	.59	0.66