

Research Interests

Pang Du

Department of Statistics
Virginia Tech

2007 SOS Faculty Seminar
September 18, 2007
Blacksburg, VA

Outline

- 1 Survival Analysis
- 2 Smoothing Splines
- 3 Other Interests

What Is Survival Analysis?

Survival analysis concerns the time to some event, generically called **failure time**.

Examples of failures:

- failure of a machine component
- a worker's re-employment after a jobless period
- cessation of smoking
- death for patient with heart disease

Most often used in biomedical studies and industrial life testing

Goals of Survival Analysis

- **Estimate** the distribution of failure time
- **Compare** the distributions of failure times among competing treatments with the goal of finding the best treatment
- **Understand** the relationship of failure time to important covariates

Why Survival Analysis?

- **Censoring**: failures of some subjects not observable due to the end of study
- Incomplete information on censored subject: don't know the exact failure time but know it must be later than the end of study
- Standard statistical techniques cannot be used

Notation

T : failure time; C : censoring time

Observables:

- $X = \min(T, C)$: observed time on study
- $\delta = I(T \leq C)$: failure indicator
- U : covariates

Interests:

- Estimate distribution of T
- Model the relationship of T to U

How? - The Targets

- Survival function:

$$S(t, u) = P(T > t | U = u)$$

- Hazard function:

$$\lambda(t, u) = \lim_{h \rightarrow 0} \frac{P(t \leq T < t + h | T \geq t, U = u)}{h}$$

Other Complications

- Missing patterns
 - left censoring, interval censoring
 - left truncation
- Multivariate survival analysis
 - competing risks
 - recurrent events
 - clustered data

Prospective of Survival Analysis

- Field still very strong
- According to Oakes (2001, *Biometrika*):
 - typical recent issue of *Biometrika* contains four or five papers
 - over 400 papers have been published in this general area in *Biometrika* since its inception in 1901
- More comprehensive, scholarly reviews: Hougaard (1999, *Biometrics*), Fleming and Lin (2000, *Biometrics*), Oakes (2001, *Biometrika*)

Available Methods

- Without covariate: Kaplan-Meier estimate (MLE, **nonparametric** method)
- With covariate: Cox proportional hazard models, accelerated failure time models (both **semi-parametric** methods)

My research: **nonparametric** methods by **smoothing splines** in the presence of covariate

General Form

Estimate a function $\eta(x)$ by minimizing

$$L(\eta) + \frac{\lambda}{2} J(\eta)$$

with respect to η in a (Hilbert) space of functions:

- $L(\eta)$: often negative log likelihood, representing goodness-of-fit,
- $J(\eta)$: **roughness penalty**. e.g., $J(\eta) = \int [\eta''(t)]^2 dt$ (without covariate).
- $\lambda > 0$: **smoothing parameter** balancing the tradeoff.

Hazard Estimation

Estimate $\eta_0(t, u) = \log h_0(t, u)$ by minimizing

$$-\frac{1}{n} \sum_{i=1}^n \left\{ \delta_i \eta(X_i, U_i) - \int_{D_i}^{X_i} e^{\eta(t, U_i)} dt \right\} + \frac{\lambda}{2} J(\eta)$$

a convex function of η , wrt η in a (Hilbert) space of functions of time and covariate.

Other Interests

- Monotone splines and receiver operating characteristic (ROC) curve estimation
- Spectral analysis