Threshold Regression Models for Time-to-Events Data

Mei-Ling Ting Lee, University of Maryland, College Park

Presented at the Stata Conference at Washington DC, 2009

MLTLEE@UMD.EDU

Outline

- Cox model has been well used to analyzing time-to-event data. It has, however, limitations.
- An example to demonstrate the usefulness of the first-hitting time based threshold regression (TR) model.
- Brief Introduction of the TR model
- Tao Xiao will present Stata codes for the alternative model.

A non-proportional hazard example: Time to infection of kidney dialysis patients with different catheterization procedures

(Nahman *et al* 1992, Klein & Moesberger 2003)

- Surgical group:

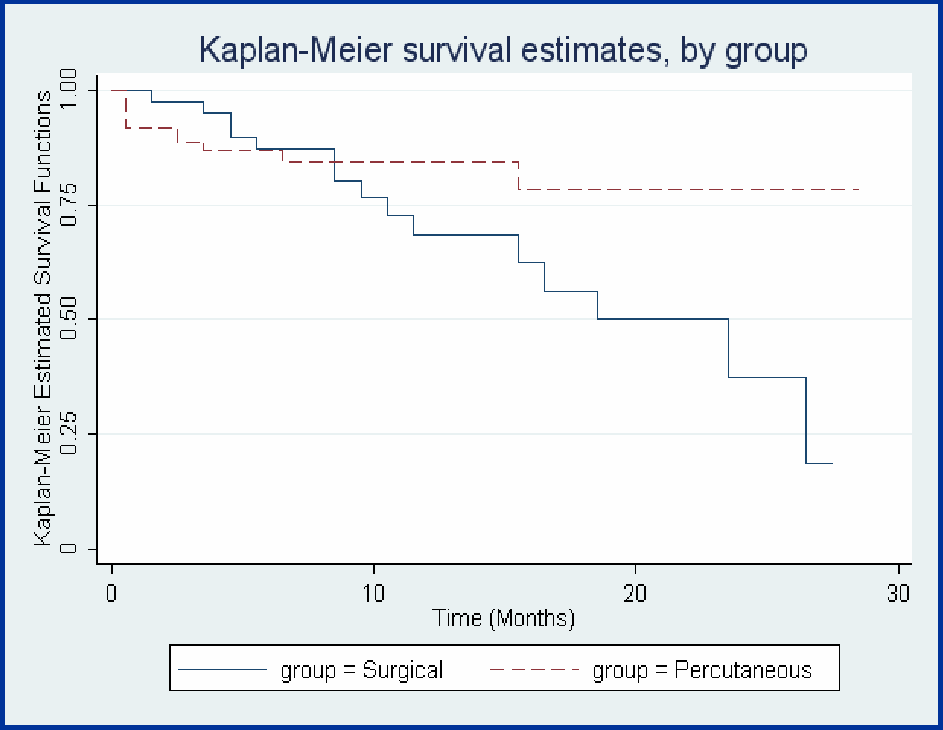
  43 patients utilized a surgically placed catheter

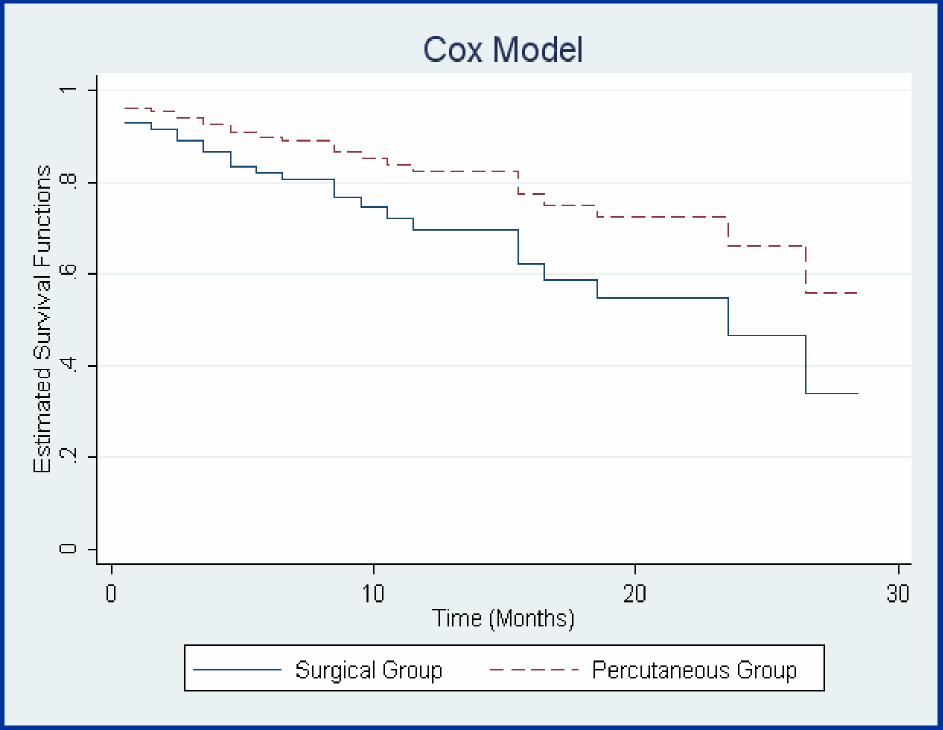
- Percutaneous group:

  76 patients utilized a percutaneous placement of their catheter

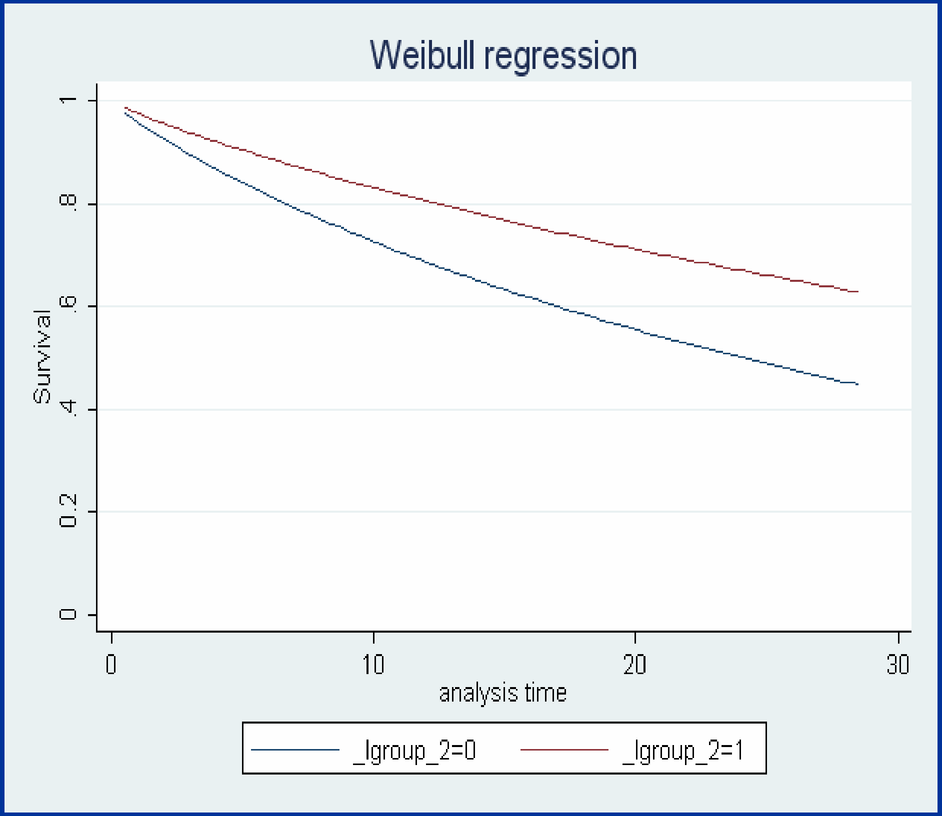
The survival time is defined by the time to cutaneous exit-site infection.

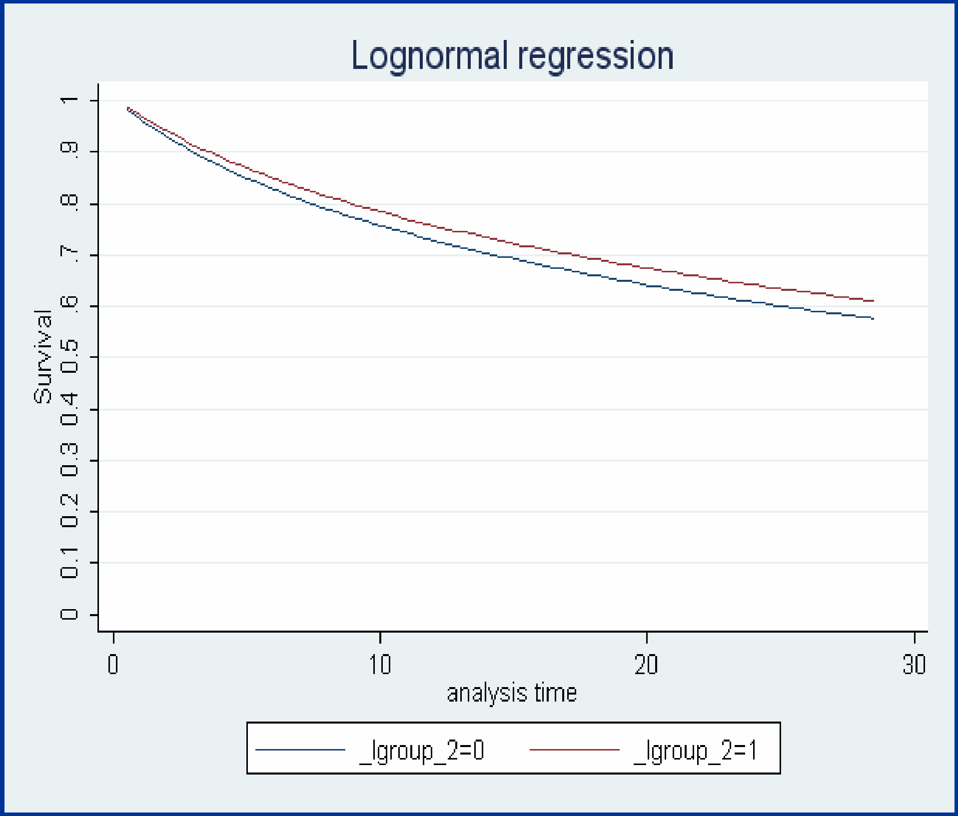
Kaplan-Meier Estimate versus PH Cox Model



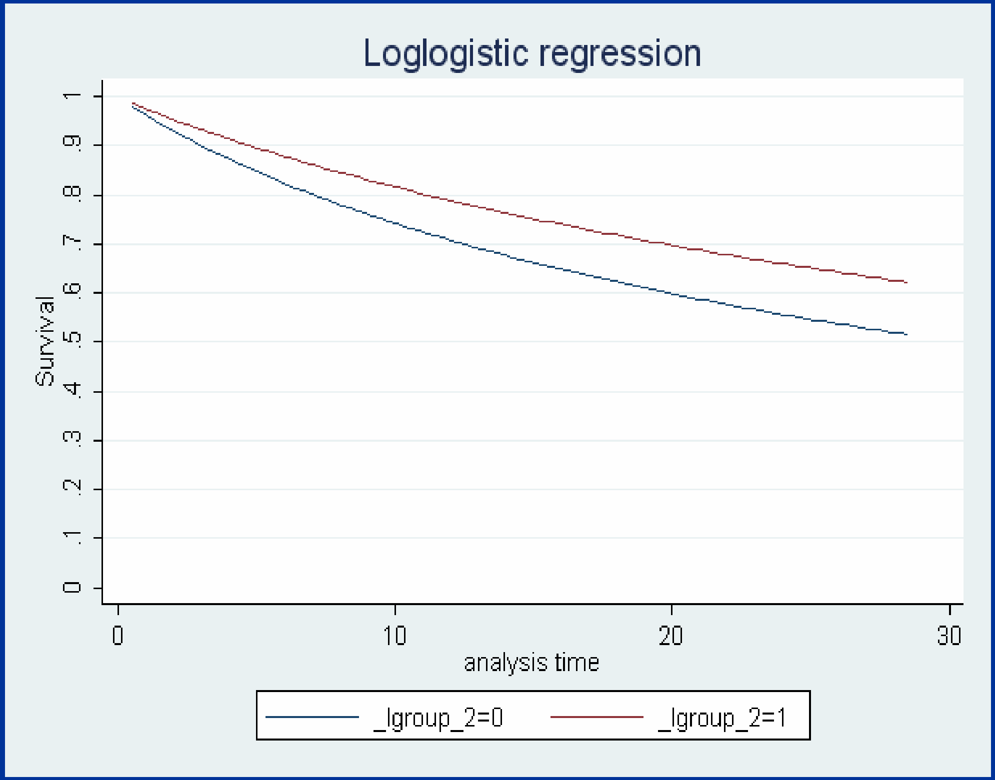


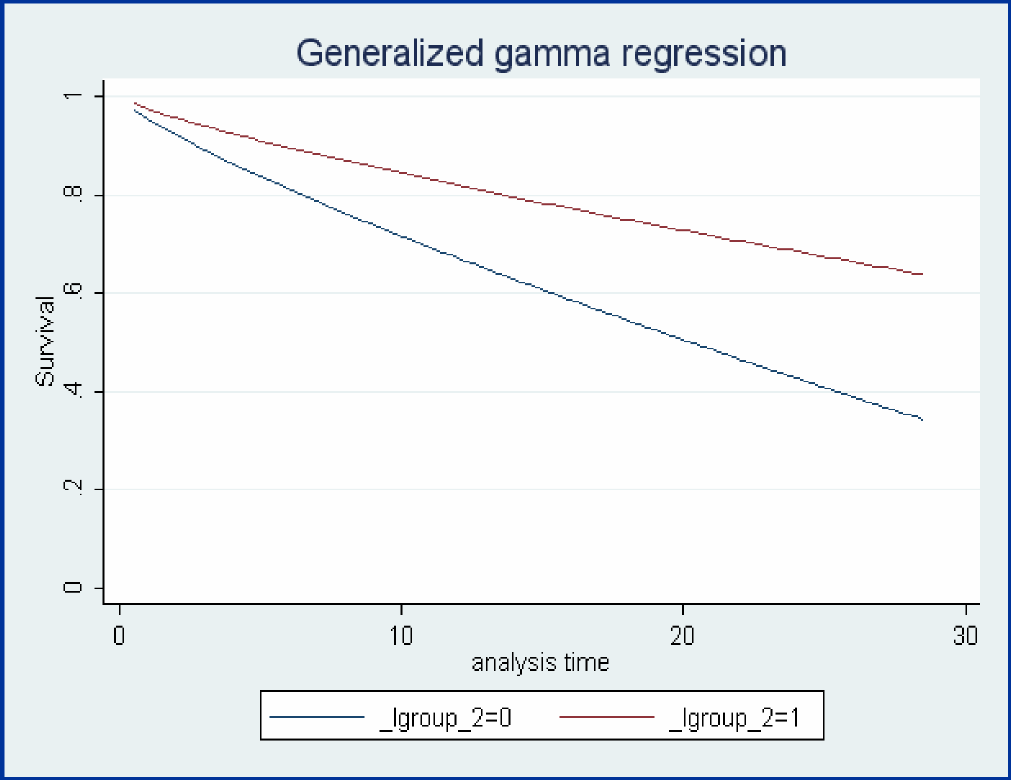
Weibull versus Lognormal



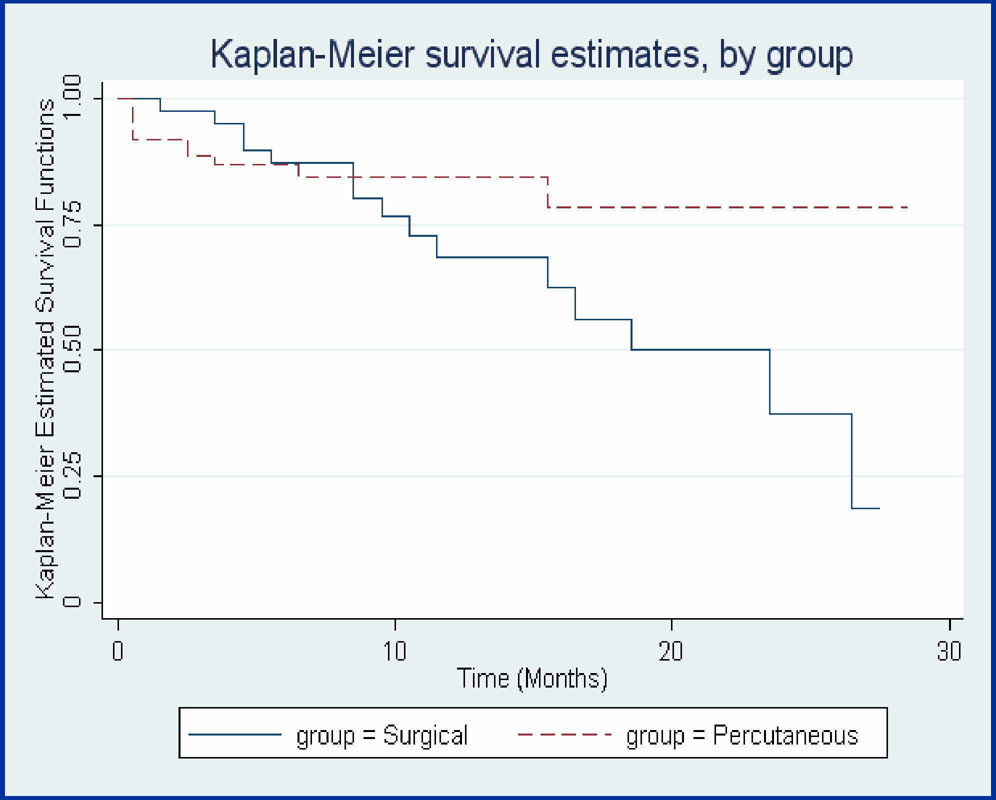


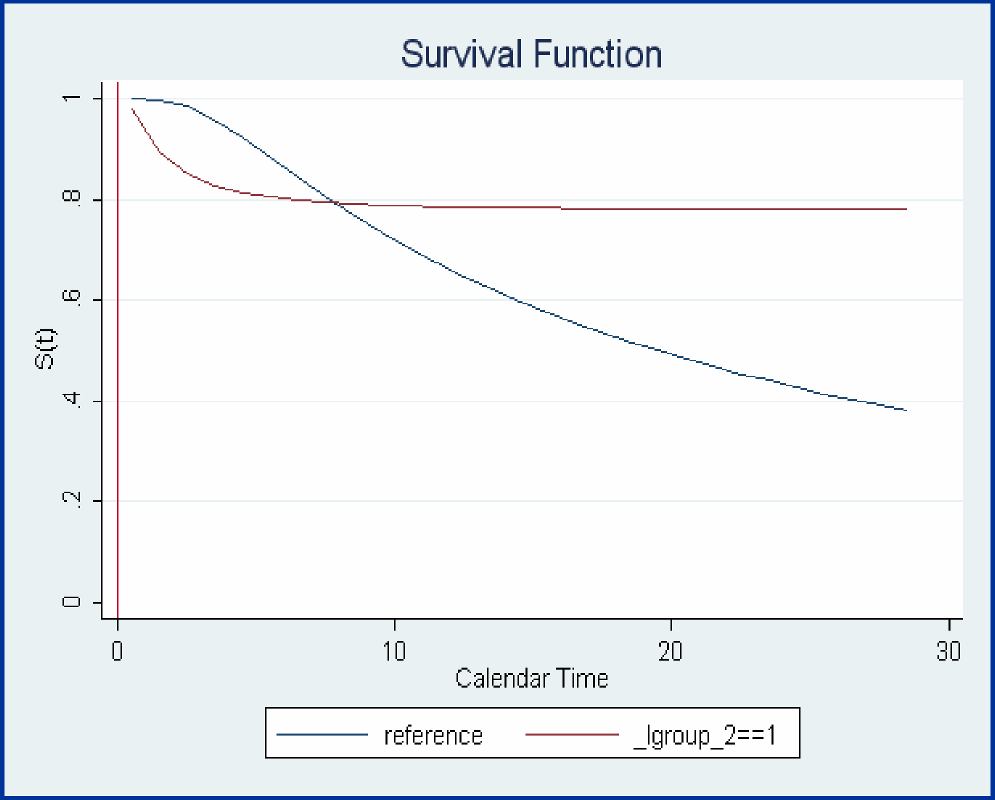
Loglogistic versus Gamma





Kaplan-Meier Estimate versus First-hitting-time based Threshold Regression Model





First-hitting Time Based Threshold Regression:

Modeling Event Times by a Stochastic Process Reaching a Boundary

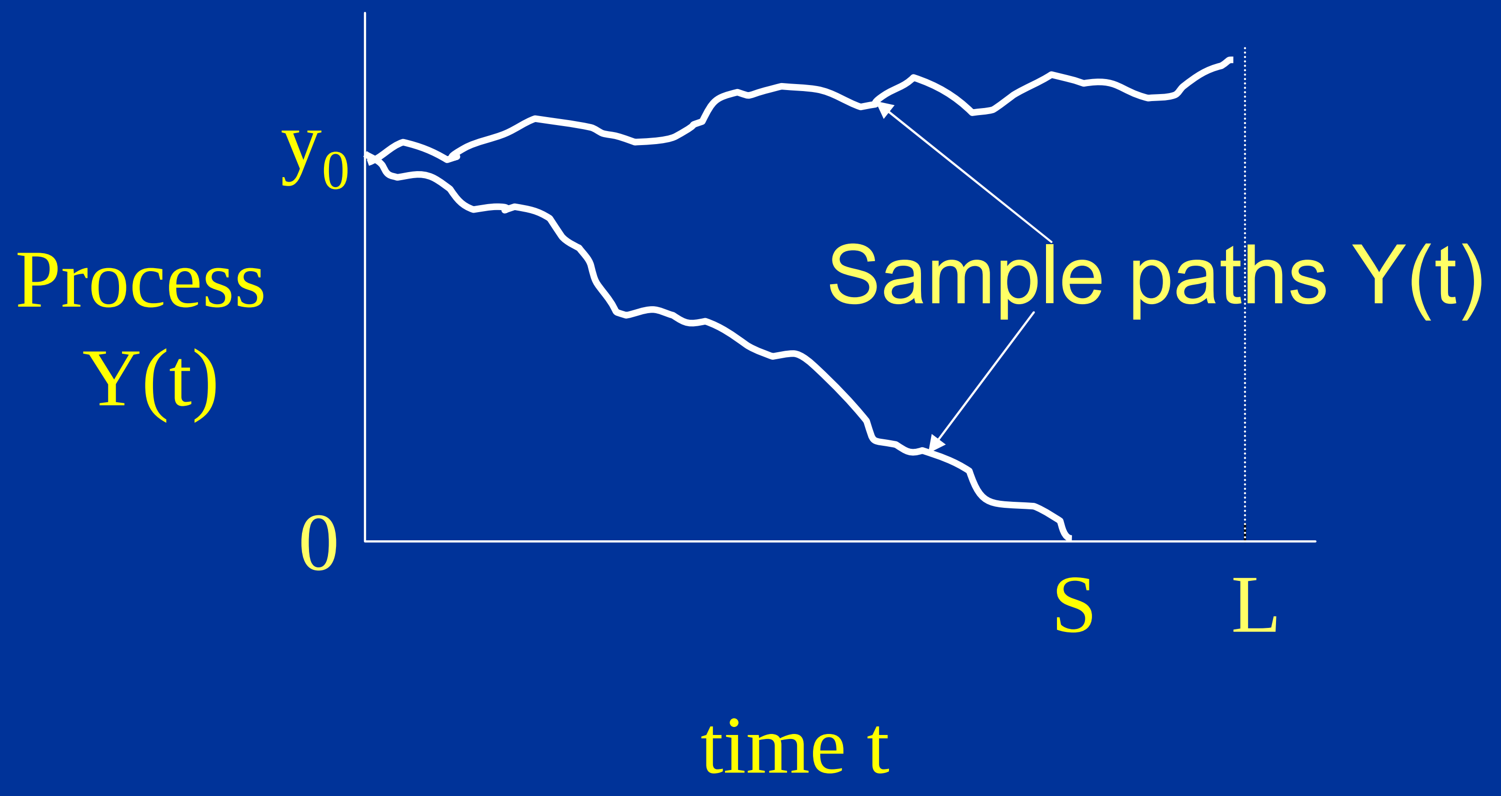
(Lee & Whitmore 2006, *Statistical Sciences*)

- Example: Equipment Failure

Equipment fails when its cumulative wear first reaches a failure threshold.

- Example: Health research

People died at heart failure, lung failure, etc



Two sample paths of a stochastic process of interest:

(1) One path experiences 'failure' at first hitting time S

(2) One path is 'surviving' at end of follow up at time L

Parameters for the FHT Model

Model parameters for the latent process $Y(t)$:

- Process parameters: $\theta = (\mu, \sigma^2)$, where

□ μ is the mean drift and σ^2 is the variance

- Baseline level of process: $Y(0) = y_0$

- Because Y(t) is latent, we set $\sigma^2 = 1$.

Likelihood Inference for the FHT Model

The likelihood contribution of each sample subject is as follows.

- If the subject fails at S=s:

$$f(s \mid y_0, \mu) = \Pr[\text{ first-hitting-time in } (s, s+ds) \,]$$

- If the subject survives beyond time *L*:

$$1 - F(L \mid y_0, \mu) = \Pr[\text{ no first-hitting-time before } L \,]$$

$$\ln L(\theta, x_0) = \sum_{i=1}^{n} \left\{ d_i \ln f\left(t_i \middle| \theta, x_0\right) + (1 - d_i) \ln \overline{F}\left(t_i \middle| \theta, x_0\right) \right\}.$$

where

d_i is the failure indicator for subject i

t_i is a censored survival time $(t_i = s_i$ if subject i fails$)$

f and $\overline{F}$ denote the FHT p.d.f and complementary c.d.f.

Threhold Regression

Link Functions: parametric or semi-parametric

Possible Link functions for the baseline parameter $Y(0)$ and drift parameter μ include

- **Linear combinations of covariates $X_1,\dots, X_p$**
- **polynomial combinations of $X_1, \dots, X_p$**
- **Regression splines**
- **Penalized regression splines**
- **Random effects**

Threshold regression (TR)

Regression estimates for parameters of:

1. Process Y(t): Wiener process, gamma process, etc

2. Boundary: straight lines or curves

3. Time scale: calendar or running time, analytical time

References

- **Aalen O.O. and Gjessing H.K. (2001). Understanding the shape of the hazard rate: a process point of view. *Statistical Science*, 16: 1-22.**

- **Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*, Second Edition, Wiley.**

- **Lee, M.-L. T. and G. A. Whitmore (2006). Threshold regression for survival analysis: modeling event times by a stochastic process reaching a boundary. *Statistical Sciences.***

- **Aalen O.O., Borgon O, and Gjessing H.K (2008). *Survival and Event History Analysis: A process Point of View*. Springer.**