A Brief Tour of Modern Survival Analysis

L.J. Wei, AM
Harvard University

Happy Birthday and “retirement” to Joe!

Practice more golfing!

Padgett, Pena, Aalen, Ebrahimi, Keiding, Klein,
McKeague, Oakes, Samaniego, Strawderman, J.L. Wang,
M.C. Wang

- What is special about survival analysis?
- Censoring(incomplete observations)
- Right censoring
- Left censoring
- Double censoring
- Truncation
- Interval censoring
- Informative censoring

- One sample problem (estimating survival, hazard function; life table analysis)
- Kaplan-Meier estimate
- Breslow and Crowley, Gill (large sample justification)
- Tsiatis (identifibility)
- Van Ryzin et al. (Bayesian)
- Confidence band based on KM estimate (Hall & Wellner; Parzen et al.)
- Quality of life adjusted KM
- KM with correlated data (Ying et al.)

- Cumulative hazard function estimate: Nelson-Aalen
- Hazard or density function estimate: Andersen, Borgan, Gill and Keiding; Tanner and Wong; Yandell; Padgett
- Confidence band (strong approximation)

- Two-sample problems
- Gehan, Gilbert, Efron test (Permutation test, U-statistic)
- Mantel test; logrank test (combining 2×2 tables at death times)
- Justification of logrank test (Crowley thesis)
- Putting different weights on those 2×2 tables Taron-Ware; Peto-Peto-Prentice; Fleming and Harrington; Gill; Kosorok et al.

- K-sample problem (Breslow for Gehan type; Gill)

- Two-sample estimation problem
- Proportional hazards (Reid)
- Scale-changed model: $T_1 =^d \theta T_2$
- Accelerated failure time model
- Louis; Wei and Gail: by inverting the rank test
- Padgett et al (minimum distance estimate)

- Group sequential analysis; interim looks
- Pocock; O'Brein and Fleming
- Slud and Wei (flexible boundaries);Lan and DeMets
- Repeated confidence intervals (quantitative monitoring)
- Conditional power (B-value)
- quantitative predictions

- Regression problem
- Without censoring, linear regression?
- Zelen: modeling the hazard
- Cox's proportional hazards model

$$\lambda(t) = \lambda_0(t) \exp(\beta' z(t))$$

- Score function; Case-controls?
- Justification for partial likelihood inference: Tsiatis
- Martingale approach: Aalen, Gill, Andersen-Gill, Sen

•Data: $\{X, \Delta, Z\}, X = \min(T, C)$

Counting process $N(t) = I(X \leq t)\Delta$

$M(t) = N(t) - A(t)$,$A(t) = \int_0^t I(X \geq s) d\Lambda(s)$

•Standardized KM, Nelson, logrank, and partial likelihood score can be written as

$$\sum_i \int_0^t H_i(s) dM_i(s)$$

- Why is the Cox model so popular?
- It is semi-parametric
- Allow time-dependent covariate (internal and external)
- Justification for the large sample theory (also efficiency)
- Commercial software available

- Prediction with the Cox model: given z what is the survival probability function? Lin, Fleming et al.
- Estimating the set of predictors z whose patients tend to be long (or short) term survivors? (Tian, Wang et al.)
- Tree classification

- Model checking?
- What is residual (observed minus expected)?
Martingale residuals $\hat{M} = N(t) - \hat{A}(t)$
- It is difficult to interpret the raw residual plot (Therneau et al.).
- Cumulative sums of residuals indexed by a component of z or $\hat{\beta}'z$ (Lin, Ying et al.)
- Variable selection (non-linear nature, O'Quigley)?
- Use prediction for model selection?

•Extensions of the Cox model (non-proportional hazards)

•Time dependent covariates model:

$\lambda(t) = \lambda_0(t)\exp(\beta' z + \gamma v(t))$

•Frailty model: $\lambda(t) = \lambda_0(t)\exp(\beta' z + W)$

•Time-dependent coefficient: $\lambda(t) = \lambda_0(t)\exp(\beta(t)' z(t))$
(Dynamic model, Zucker, Scheike, Aalen, Borgan, Hjort, Y. Sun)

•Measurement error (Prentice, X. Lin)

- Other non-proportional hazards models
- Additive hazards model (Aalen)
- Mixture of additive and multiplicative model (Martinussen)

•Linear regression model (AFT)

$$\log(T) = \beta' z + \epsilon$$

•Rank estimation (Prentice; Tsiatis, Ritov, Ying, Lai, Lin)

Consider "residuals" $\{X_i - \beta' Z_i, \Delta_i, Z_i\}$ and invert weighted logrank statistics

•The resulting estimating function for β is not continuous.

- Estimation and inferences about β is not easy.
- With Gehan weight for the estimating equation, there is a unique root $\hat{\beta}$
- General logrank estimating function with monotone weighting function (Jin, Lin, Ying et al.)
- Non-monotone weighting: weighted bootstrapping (Tian, Jun Liu et al.)

- Prediction of survival function given z with AFT model (Y. Park)
- Model checking for the AFT model
  Cumulative sums of ordinary residuals indexed by a component of Z (Leon)
- Box-Cox transformation: $g_\tau(T) = \beta' Z + \epsilon$ (T. Cai et al.)
- Time-dependent covariates (Robins, Tsiatis, Slud)
- Time-dependent regression coefficient (quantile regression)

•Linear transformation models

•Cox model: $log\{-logS_Z(t)\} = h(t) + \beta' Z$

•Proportional odds model: $-logit\{S_Z(t)\} = h(t) + \beta' Z$

•General models: $g(S_Z(t)) = h(t) + \beta' Z$

$$h(T) = -\beta' Z + \epsilon$$

the survival function of ϵ is g^{-1}

(Dabrowska & Doksum, Cheng et al., Scharfstein, Horowitz, Chen et al.)

• Time-dependent covariates (it is different from Cox's!)

•Time-dependent regression coefficient (Y. Park)

- Quantile regression:
- Median regression: median$(T) = \beta' Z$ (Ying, et al; Tsiatis et al. Robins)
- Estimate for quantiles may not be monotone

Multivariate Failure Time Data

- Multivariate KM estimate (Prentice, Cai, Van der Laan, Robins)

•Distinct types of failures

Times to various serious bacteria infections: HIV patient who may have bacteremia, pneumonia, syphilis, meningitis

•Can we get a single score (like non-censored case) for each patient based on possibly censored multiple event times?

- For each event time, use the Cox model, combining the estimates of a specific component of the regression parameters (Wei, Lin & Weissfeld)
- With more elaborate working correlation model (Prentice and Cai)
- General marginal modeling approach (other survival models, AFT, transformation)

•Clustered failure time data

•Genetics studies (linkage or association studies)

•Each family is a cluster, each family member contributes a possibly censored event time

$$\lambda_{ij}(t) = \lambda_0(t)\exp(\beta' Z_{ij})$$

(Lee et al.; Cai and Prentice; X. Lin)

•General approach (using other survival models)

- Recurrent event time data
- Modeling intensity function (Andersen-Gill)
- WLW
- Modeling the mean: $E\{N(t)\} = \mu(t)\exp(\beta' Z)$ (Lawless, Pepe, Cai, Lin, Ying)
- Problem with informative censoring or terminal event? (M. Wang)

- Random effects model (including frailty model)
- Cox model:$\lambda_{ij}(t) = \lambda_0(t) \exp(\beta' Z_{ij} + W_i)$ (Murphy & Sen; Parner)
- Linear transformation: $g(S_{ij}(t)) = h(t) + \beta' Z_{ij} + W_i$ (T. Cai)
- $T_{ij} = \beta' Z + W_i + \epsilon_{ij}$ (almost non-identifiable problem?)
- What is the advantage of random effects model over marginal model? (prediction)
- What is the disadvantage? (complex inference procedures)

- Competing risks
- Dependent censoring
- Cumulative incidence function

$F(t) = n^{-1} \sum_i I(X_i \leq t; E_i = 1)$ (Gray)

- Need all the incidence functions to make cost-benefit analysis
- Regression models (Fine, Gray)

More work on:

- Modeling event time with time-dependent bio-markers
(Dynamic treatment strategy: Robins, Murphy)
- Complex censoring
- Informative censoring
- Model selection (variables)

•Empirical Bayes?

Complicated posterior?

•Using frequenstist's approach to eliminate "nuisance" and approximate the likelihood (Efron)?

•Not sure any gain from large sample theory?

•Panel data?

•A stochastic process $\{X(t), Z(t), t \geq 0\}$
$X(t)$ is the response and $Z(t)$ is the covariate

•For each patient, we only open the window at time points $\{t_1, \cdots, t_K\}$ (Marked point process)

•Modeling $X(t)$:

$$E(X(t)|Z(t)) = \mu(t)\exp(\beta' Z(t))$$

How to make inferences about β and prediction for $X(t)$ given Z?

50 years after KM; 40 years after Mantel test; 30 years after the Cox model; 20 years after Andersen-Gill-Aalen-Sen; one year after Kalbfleisch & Prentice(2nd Edition)

Predicting 50^{+} more fruitful years to come!