Marginal Models for Cox Proportional Hazards Regression

Alex McLain
Department of Statistics, University of South Carolina
Research support provided by NIH Grant 2 R01 GM56182
Advisor: Dr. Edsel A. Peña

October 11, 2007

Outline

- Introduction
- Cox Proportional Hazard Model
- Marginal Models
- Current Results
- Conclusion

Recurrent Event Setting

- Recurrent event data can be found in many applications.
  1. Reliability (Bugs in software, repairs of a automobile)
  2. Biomedical (Recurrences of a tumor, relapses of a drug user)
  3. Social Sciences (Party switching of a seat in congress, Repeat criminal offenses)

Data Structure

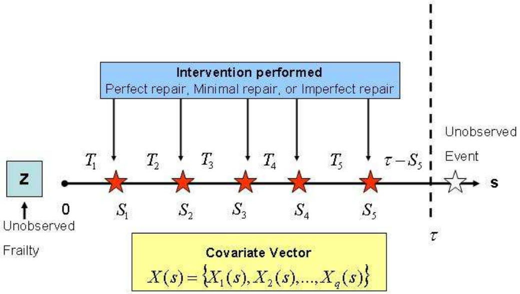

Figure: Graphical view of a unit with recurrent events.

- T_i are the interevent times, S_i are the calendar times, and τ_i is the censoring time

Methods

Analysis of Recurrent Event Data:

- Parametric (using Stocker and Peña (2007), under exponential, extreme value, lognormal, etc.)
- Nonparametric (Kaplan-Meier, Nelson-Aalen, Peña et al. (2001) etc..)
- Semi-parametric (Cox proportional hazards model)

Introduction

- In the Cox proportional hazards model, we assume that the hazard rate has the following form:

$$\lambda(t|x) = \lambda_0(t)\exp\{\beta x\}$$

  where β is a q dimensional vector of parameters, and $\lambda_0(t)$ can be specified parametrically or assumed to be nonparametric.
- In this model the observable data for the ith unit is (Z_i, δ_i, x_i). Where $Z_i = T_i \wedge \tau_i$, $\delta_i = I(T_i \leq \tau_i)$ and x_i is a q dimensional covariate vector.
- The goal of this model is to make inference about λ_0 and/or β.

Methods of Inference

- From Cox (1972), Andersen and Gill (1982) and Andersen et al. (1993) we have;

$$N_i(s) = \sum_{k=1}^{\infty} I(S_{ik} \leq s \text{ and } S_{ik} \leq \tau_i) \quad \& \quad Y_i(s) = I(s \leq \tau_i)$$

As the counting process and at risk indicator. Then from Andersen and Gill (1982) we have that;

$$M_i(s) = N_i(s) - \int_0^s Y_i(v)\lambda(v|x_i)dv \qquad (1)$$

Is a zero mean square integrable martingale, under certain conditions.

Parametric Inference in the Cox Model

- If $\lambda_0(t)$ is parametric (i.e. $\lambda_0(t) = \lambda_0(t;\theta)$ for some $\theta \in \Theta$). Then inference can be made through the full likelihood, which in terms of Jacod (1974/75) is:

$$L_F(\theta, \beta) = \prod_{i=1}^{n} \left\{ \prod_{s=0}^{\infty} [\lambda_0(s;\theta)e^{\beta x_i}]^{dN_i(t)} \exp\{-\int_0^{\infty} Y_i(u)\lambda_0(u;\theta)e^{\beta x_i} du\} \right\}$$

- L_F is then maximized with respect to θ and β.

Semi-Parametric Inference in the Cox Model

- If $\lambda_0(\cdot)$ is nonparametric then we are left with a semi-parametric model. In order to make inference on β we use the partial likelihood of Cox (1972):

$$L_P(\alpha) = \prod_{i=1}^{n} \prod_{t=0}^{\infty} \left[\frac{e^{\alpha x_i}}{\sum_{j=1}^{n} Y_i(t) e^{\alpha x_j}} \right]^{dN_i t} \tag{2}$$

and score

$$U_p(\alpha) = \sum_{i=1}^{n} \int_0^{\infty} \left[x_i - \frac{S^{(1)}(t;\alpha)}{S^{(0)}(t;\alpha)} \right] dN_i(t) \tag{3}$$

where $S^{(m)}(t;\alpha) = \sum_{i=1}^{n} x_j^m Y_j(t) e^{\alpha x_j}$.

- L_P is then maximized with respect to α.

Marginal Modeling

- For a set of data **Y** there is a true full model:

$$P(\cdot|\theta) \quad \textit{for} \quad \theta \in \Theta$$

- For the kth event in **Y** there is a true marginal model generated from the full model:

$$P_j(\cdot|\theta) \quad \textit{for} \quad \theta \in \Theta$$

- Recently it has become common to assume a marginal model on the kth event of the data:

$$Q_j(\cdot|\theta_j) \quad \textit{for} \quad \theta_j \in \Theta_j$$

- Marginal model offer an ease of interpretation, and are sometimes more computationally efficient.

Questions of marginal models

What is the overall effect of using marginal models?

- Is there loss of efficiency?
- Should marginal estimates be combined?
- What is the bias of these models when incorrectly specified?

Marginal Modeling with the Cox Model

There are two methods of marginal analysis with the Cox model:

***1.** Conditional (PWP) Model*

The PWP model what proposed in Prentice et al. (1981) and assumes that a subject is not at risk for the kth event until the k-1st event has occurred. It assumes that the intensity process for the ith subject and the kth event is:

$$\lambda_{ik}(t|\beta, x_i) = Y_{ik}^{p}(t)\lambda_0(t)e^{\beta x_i}$$

where $Y_{ik}^{p}(t) = I(S_{ik-1} < t \leq S_{ik-1} \; and \; t \leq \tau_i)$. The PWP model puts each inter-event into a different strata, and each subject will have K_i stratums.

Marginal Modeling with the Cox Model

2. *WLW Model*:
The WLW model, proposed in Wei et al. (1989), assumes that a subject is at risk for the kth event from the beginning of observation.
It assumes that the intensity process for the ith subject and the kth event is:

$$\lambda_{ik}(t|\beta, x_i) = Y_{ik}^{w}(t)\lambda_0(t)e^{\beta x_i}$$

where $Y_{ik}^{w}(t) = I(t \leq S_{ik} \; and \; t \leq \tau_i)$. The WLW model has k_{max} stratums for each subject.

Analysis

- Inference on the PWP & WLW are based on $L^{p}_{pk}(\alpha)$ and $L^{w}_{pk}(\alpha)$ the partial likelihood in (2) with the respective at risk process, for each strata. These models result in estimates $\hat{\beta}^{p}_{k}$ and $\hat{\beta}^{w}_{k}$ $\forall$ k.
- An overall estimate may be obtained by defining the log likelihood to be the sum of the stratum specific log likelihoods, for the given at risk;

$$\ell_p(\beta) = \sum_k \ell_{pk}(\beta)$$

- WLW('89) propose an "average effect" which combines the β_k's for an estimator with the smallest asymptotic variance among all linear estimators.

Properties

- It is clear from (1) and (3) that the score function in the partial likelihood can be rearranged to be:

$$U_p(\alpha) = \sum_{i=1}^{n} \int_0^\infty [x_i - E(t;\alpha)]\, dM_i(t) + \sum_{i=1}^{n} \int_0^\infty [x_i - E(t;\alpha)]\, Y_i(t)\lambda_0(t)e^{\beta x_i}\, dt \quad (4)$$

  where $E(t;\alpha) = \frac{S^{(1)}(t;\alpha)}{S^{(0)}(t;\alpha)}$.
- Hence, if the model is correctly specified $M_i(\cdot)$ is a zero mean martingale, and the second expression is zero at $\alpha = \beta$.
- In this case the score is a zero mean process at the true value of β.

Properties under an IID assumption

- Let $T_{ik} \sim F$ for some cdf F with positive support, and corresponding hazard $\lambda(t;\beta) = \lambda_0(t)e^{\beta x}$.
- Then $S_{ik} \sim F^{*k}$ where $*k$ denotes the kth convolution of F with hazard $\lambda_k(t;\beta)$
- Under these assumptions:

$$M_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}^w(v)\lambda_k(v;k,e^{\beta x_i})dv$$

Forms a zero mean square integrable martingale, under some regularity conditions.

Score Function

- In this case the score function becomes:

$$U_{pk}(\beta) = \sum_{i=1}^{n} \int_0^\infty [x_i - E(t;\beta)]\, dM_{ik}(t) + \sum_{i=1}^{n} \int_0^\infty [x_i - E(t;\beta)]\, Y_{ik}^{w}(v)\lambda_k(v;k,e^{\beta x_i})dv$$

and

$$\frac{1}{n}U_{pk}(\beta) \longrightarrow u_k(\beta) = E\left\{\int_0^\infty [x_i - e(t;\beta)]\, Y_{ik}^{w}(v)\lambda_k(v;k,e^{\beta x_i})dv\right\}$$

as $n \to \infty$, where $e(t;\alpha) = \frac{E[S^{(1)}(t;\alpha)]}{E[S^{(0)}(t;\alpha)]}$

Bias

- Solving $u_k(\beta)$ will give the asymptotic estimate of β when $n \rightarrow \infty$.
- When $Y_{ik} = Y_{ik}^p$ it can be shown that $u_k(\alpha)= 0$ at $\alpha = \beta$. Hence the PWP model is unbiased under the iid assumption for any F.
- When $Y_{ik} = Y_{ik}^w$ then $u_k(\alpha)=0$ does not necessarily occur when $\alpha = \beta$.
- When this is the case $u_k(\alpha)$ can be used to calculate the asymptotic bias

Bias under a Homogeneous Poisson Process

- The simplest for of a proportional hazard model is when $\lambda_0(t) = c$ or the Homogeneous Poisson Process.
- If our F is the cdf of an Exponential($e^{\beta x}$), then F^{*k} will be that of a Gamma(k, $e^{\beta x}$).
- Under these assumptions it can be shown that the solution for $u_k^w(\alpha)$, β_k^w will be larger in absolute value than β when k $>$ 1, for $x \sim Ber(p)$.

Simulated Bias of β_k^W

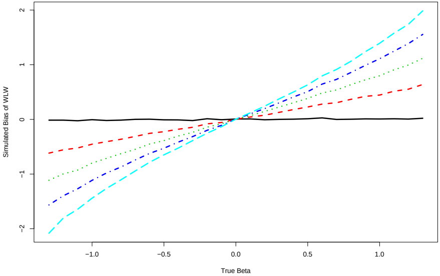

Figure: Simulated bias of the WLW model (for k=1, 2, 3, 4, and 5) under exponential interevent times, for n=50 and M=2000

Calculated Bias using $u_k^w(\alpha)$

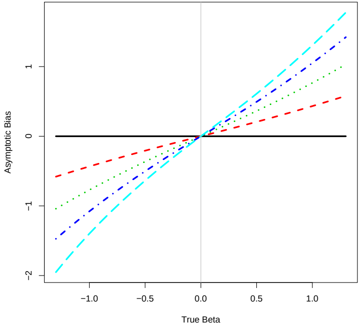

Figure: Calculated bias of the WLW method when the interevent times follow an Exponential (e^{β}).

Summary

- Models where the use of marginal analysis is justified need to be developed and studied (see Yang and Ying (2001).
- More theory for the effect of marginal analysis under an IID setting is needed.
- When combing coefficients, the resulting estimator is difficult to interpret.

Thank You

Thanks for reading about my work
Please contact me at alex.mclain@gmail.com with any questions!!