Bayesian Analysis of Progressively Censored Competing Risks Data

Debasis Kundu

Department of Mathematics & Statistics
Indian Institute of Technology Kanpur

This is a joint work with Dr. Biswabrata Pradhan, and part of this work is going to appear in Sankhya, B.

December 30, 2011

Outline

Outline

What is a competing risk model?

In many life-testing studies often the failure of items may be associated to more than one cause. These risks factors in some sense compete with each other for the failure of the experimental unit. The models which can be used to analyze such data are known as the competing risk models.

Associated Issues

1. Typically competing risk data looks like (T, Δ). Here T denotes the lifetime of the item and Δ denotes the cause of failure.
2. The cause of failure may be assumed to be dependent or independent.
3. Although the assumption of dependence seems more reasonable, there is some identifiability issue.
4. It is not possible to test the assumption of independence of the failure time distributions without the presence of covariates.

Latent failure time modeling

Cox (1959, JRSS B) proposed the latent failure time modeling to analyze competing risk data. The latent failure time model has the following form:

$$T_i = \min\{X_{i1}, \cdots, X_{iM}\}$$

here T_i denotes the lifetime of the i-th individual. $X_{i1}, \cdots, X_{iM}$ are the latent failure times of the M different causes for the i-th individual, and they are assumed to be independent.

Outline

Progressive Censoring

Progressive censoring can be described as follows:

- Put n items on test.
- m and $R_1, \cdots, R_m$ are pre-fixed integers such that

$$R_1 + \cdots + R_m = n - m.$$

- At the time of the first failure, R_1 of the remaining units are randomly chosen and removed. Similarly, at the time of the second failure R_2 of the remaining units are chosen and removed. Finally at the time of the m-th failure rest of the $R_m = n - m - R_1 - \cdots - R_{m-1}$ units are removed.

Outline

Progressively Censored Competing Risks Data

It is assumed we have a progressively censored competing risks data. Therefore, a progressively censored competing risks data will be as follows:

$$(t_1, \delta_1, R_1), \cdots, (t_m, \delta_m, R_m)$$

$$\{(t_i, 1); i \in I_1\}, \cdots, \{(t_i, M); i \in I_M\},$$

here

$$I_1 = \{i; \delta_i = 1\}, \cdots, I_M = \{i; \delta_i = M\}.$$

Lifetime distribution

Here it is assumed that $M = 2$. The latent failure times have distributions with the same shape parameter but different scale parameters, *i.e.* $X_1 \sim \text{WE}(\alpha, \lambda_1)$ and $X_2 \sim \text{WE}(\alpha, \lambda_2)$ and they are independent. $\text{WE}(\alpha, \lambda)$ has the PDF

$$f(x; \alpha, \lambda) = \alpha \lambda x^{\alpha-1} e^{-\lambda x^{\alpha}}; \quad x > 0.$$

It is well known that Weibull distribution is a very flexible lifetime distribution. Moreover, $\min\{X_1, X_2\} \sim \text{WE}(\alpha, \lambda)$, $\lambda = \lambda_1 + \lambda_2$.

Prior Assumptions on λ_1 and λ_2

$$\pi_0(\lambda|a_0, b_0) = \frac{b_0^{a_0}}{\Gamma(a_0)} \lambda^{a_0-1} e^{-b_0\lambda}$$

$$\pi(\lambda_1/\lambda|a_1, a_2) = \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)} \left(\frac{\lambda_1}{\lambda}\right)^{a_1-1} \left(1-\frac{\lambda_1}{\lambda}\right)^{a_2-1}$$

$$\begin{aligned}\pi(\lambda_1, \lambda_2|a_0, b_0, a_1, a_2) \quad = \quad & \frac{\Gamma(a_1+a_2)}{\Gamma(a_0)} (b_0\lambda)^{a_0-a_1-a_2} \times \\ & \frac{b_0^{a_1}}{\Gamma(a_1)} \lambda_1^{a_1-1} e^{-b_0\lambda_1} \\ & \times \frac{b_0^{a_2}}{\Gamma(a_2)} \lambda_2^{a_2-1} e^{-b_0\lambda_2}\end{aligned}$$

Beta-Dirichlet Prior

1. $$\pi(\lambda_1, \lambda_2 | a_0, b_0, a_1, a_2) \sim \text{BD}(b_0, a_0, a_1, a_2)$$

2. If $a_0 = a_1 + a_2$, λ_1 and λ_2 are independent

3. λ_1 and λ_2 positively or negatively correlated if $a_0 > a_1 + a_2$ or $a_0 < a_1 + a_2$ respectively.

Mean and Variance of the Beta-Dirichlet Prior

If $(\lambda_1, \lambda_2) \sim \text{BD}(b_0, a_0, a_1, a_2)$, then for $i = 1, 2$

$$E(\lambda_i) = \frac{a_0 a_i}{b_0(a_1 + a_2)}$$

$$V(\lambda_i) = \frac{a_0 a_i}{b_0^2(a_1 + a_2)} \times \left\{ \frac{(a_i + 1)(a_0 + 1)}{a_1 + a_2 + 1} - \frac{a_0 a_i}{a_1 + a_2} \right\}$$

Prior Assumption on α

No specific prior has been assumed on α. It is assumed that the prior on α has a support on $(0, \infty)$, the PDF is log-concave. It is independent of the prior on (λ_1, λ_2).
Note that several standard distribution functions have log-concave PDF, for example (i) log-normal, (ii) gamma (shape parameter greater than 1), (iii) Weibull (shape parameter greater than 1), (iii) generalized exponential distribution (shape parameter greater than 1) etc.

Outline

Shape Parameter Known

In this case

$$\pi(\lambda_1, \lambda_2 | Data) \sim \text{BD}\left(B_0, A_0, A_1, A_2\right),$$

here

$$\begin{aligned} B_0 &= b_0 + \sum_{i=1}^{m} (R_i + 1) t_i^{\alpha} \\ A_0 &= a_0 + m_1 + m_2 \\ A_1 &= a_1 + m_1 \\ A_2 &= a_2 + m_2. \end{aligned}$$

Bayes Estimates

The Bayes estimate under the squared error loss function is the posterior mean. Therefore, the Bayes estimates of λ_1 and λ_2 can be obtained as

$$\widehat{\lambda}_{1B} = \frac{A_0 A_1}{B_0(A_1 + A_2)}$$

$$\widehat{\lambda}_{2B} = \frac{A_0 A_2}{B_0(A_1 + A_2)}$$

The corresponding posterior variances also can be easily obtained in terms of B_0, A_0, A_1, A_2.

Credible Set

A set C_β is called a level 100(1-β)% credible set if

$$P\left((\lambda_1, \lambda_2) \in C_\beta | (\lambda_1, \lambda_2) \sim \pi(B_0, A_0, A_1, A_2)\right) = 1 - \beta$$

$(\lambda_1, \lambda_2 | Data) \sim \text{BD}(B_0, A_0, A_1, A_2)$, therefore

$$\lambda_1 + \lambda_2 | Data \sim \text{Gamma}(A_0, B_0)$$

and

$$\left. \frac{\lambda_1}{\lambda_1 + \lambda_2} \right| Data \sim \text{Beta}(A_1, A_2)$$

and they are independently distributed.

Credible Set

Choose β_1 and β_2 such that (1-β) = (1-β_1)(1-β_2). Further choose (l_1, u_1) and (l_2, u_2) such that

$$P(l_1 \leq \lambda_1 + \lambda_2 \leq u_1) = 1 - \beta_1$$

$$P(l_2 \leq \frac{\lambda_1}{\lambda_1 + \lambda_2} \leq u_2) = 1 - \beta_2.$$

$$C_\beta = \left\{ (\lambda_1, \lambda_2); l_1 \leq \lambda_1 \leq u_1, l_2 \leq \frac{\lambda_1}{\lambda_1 + \lambda_2} \leq u_2 \right\}$$

Credible Set

C_β = The area between the four straight lines:

$$\begin{aligned} \lambda_1 &= u_1, \\ \lambda_1 &= l_1, \\ \lambda_1(1-u_2) &= u_2\lambda_2 \\ \lambda_1(1-l_2) &= l_2\lambda_2. \end{aligned}$$

Common Shape Parameter is Unknown

In this case the joint posterior distribution of α, λ_1 and λ_2 given the Data can be written as follows:

$$l(\alpha, \lambda_1, \lambda_2 | Data) = l(\lambda_1, \lambda_2 | \alpha, Data) \times l(\alpha | Data),$$

where

$$l(\lambda_1, \lambda_2 | \alpha, Data) \sim \text{BD}\left(B_0, A_0, A_1, A_2\right)$$

$$l(\alpha | Data) \propto \pi_2(\alpha) \alpha^m \prod_{i=1}^{m} t_i^{\alpha} \times \frac{1}{B_0^{A_0}}$$

Bayes Estimates

It is not possible to compute the Bayes estimates or the associated credible intervals in closed form for α, λ_1 and λ_2.
The following result becomes useful:

THEOREM The conditional PDF of α give the data is log-concave.

Algorithm

1. Step 1: Generate α using the algorithm suggested by Devroye (1984, Computing), or the approximation method suggested by Kundu (2008, Technometrics).
2. Step 2: Given α generate λ_1, λ_2 from a BD(B_0, A_0, A_1, A_2).
3. Step 3: Once the posterior samples are obtained, then simulation consistent Bayes estimates and the associated credible intervals can be constructed.

Outline

Optimal censoring scheme

What is an Optimal Censoring Scheme?
For a given n and m what should be the choice of $R_1, \cdots, R_m$ such

$$R_1 + \cdots + R_m = n - m$$

and it provides the maximum *Information* of the unknown parameters.

Information measure

How to define Information measure of a censoring scheme? If it is only one parameter, the variance of the estimators can be defined as an information measure. But for more than one parameters, the choice is not clear. Trace or determinat of the Fisher information matrix has been used, but they are not scale invariant.

Information measure

Variance of the p-th percentile estimator can be taken as a possible measure, Zhang and Meeker (2005, Metrika). In this case

$$C(R) = wVar(\ln \widehat{T}_{p,1}) + (1-w)Var(\ln \widehat{T}_{p,2}),$$

where $0 < w < 1$, and

$$T_{p,1} = \left[-\frac{1}{\lambda_1}\ln(1-p)\right]^{1/\alpha} \quad T_{p,2} = \left[-\frac{1}{\lambda_2}\ln(1-p)\right]^{1/\alpha}$$

Criterion

The proposed criterion takes the following form:

$$C^B(R) = wE_{data}\left[V_{posterior}(\ln \widehat{T}_{p,1})\right] + (1-w)E_{data}\left[V_{posterior}(\ln \widehat{T}_{p,2})\right]$$

It depends on the percentile point p. The following more general criterion has been proposed.

Criterion

$$
\begin{aligned}
C^B(R) \;&=\; wE_{data}\left[\int_0^1 V_{posterior}(\ln \widehat{T}_{p,1})\,dW_1(p)\right] \\
&\quad +(1-w)E_{data}\left[\int_0^1 V_{posterior}(\ln \widehat{T}_{p,2})\,dW_2(p)\right]
\end{aligned}
$$

Optimum censoring scheme

In all the cases the optimization has to be performed numerically. It is a discrete optimization problem. For a given n and m, the optimum censoring scheme with respect to a given criterion can be found by exhaustive search for all possible values of R_i's satisfying

$$R_1 + \cdots + R_m = n - m.$$

Outline

Open problems

1. $M > 2$.
2. Other censoring schemes, like Type-I progressive censoring scheme, (ii) Type-I hybrid censoring scheme, (iii) Type-II hybrid censoring scheme etc.
3. Other proportional hazard models.
4. Data analysis in presence of covariates.

Thank You