The Korean Communications
in Statistics Vol. 4, No. 2, 1997
pp. 491 - 496

Discount Survival Models for No Covariate Case

Joo Yong Shim[1]

Abstract

For the survival data analysis of no covariate the discount survival model is proposed to estimate the time-varying hazard rate and the survival function recursively. In comparison with the covariate case it provide the distributionally explicit evolution of hazard rate between time intervals under the assumption of a conjugate gamma distribution. Also forecasting of the hazard rate in the next time interval is suggested, which leads to the forecasted survival function.

1. Introduction

West, Harrison and Migon(1985) developed the dynamic generalized linear model which allows the use of one dimensional exponential family distribution. The prior distribution is chosen to be a conjugate family member having same first two moments as the linearized function of a parameter vector. Gamerman(1991) proposed the dynamic Bayesian model for the analysis of survival data with the hazard rate which is a function of a time-varying parameter vector modeling effects of covariates. It is assumed that the time-variation of a parameter vector is determined through the evolution of a parameter vector between time intervals in the linear pattern, where its distribution is partially specified in terms of the mean vector and the variance-covariance matrix. Ameen and Harrison(1985) developed the normal discount Bayesian model to overcome some practical disadvantages of dynamic linear model, where the discount factor $d_i\,(0 < d_i < 1)$ is introduced for an increase of $100(1-d_i)/d_i$ percent in the variance of the time interval I_i . Bolstad(1995) proposed the multiprocess dynamic Poisson model for the no covariate case by applying the multiprocess dynamic approach to the time-varying Poisson parameter.

In this article, under the assumption of a conjugate gamma distribution of the hazard rate in each time interval, the time-variation of the hazard rate between time intervals is determined through the multiplication of a gamma parameters by discount factors.

1) Lecturer, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

The recursive estimation of the hazard rate and the survival function based on the information gathered until the end of each time interval is provided in Section 2. Forecasting of the hazard rate and the survival function in the next time interval is provided in Section 3. The performance of the discount survival model for no covariate case is illustrated via the gastric carcinoma data in Section 4.

2. Model descriptions

Here we assume that the survival time has a piecewise exponential distribution which has the constant hazard rate in each time interval, $\lambda(t)=\lambda_i$ for $t\in I_i=(\tau_{i-1},\tau_i]$, where τ_0 is usually set to 0 and $I_s=(\tau_{s-1},\infty)$. The survival function and the probability function for the current time interval I_i given survival up to the end of previous time interval can be easily calculated due to the lack of memory property of exponential distributions. Thus we obtain the likelihood for λ_i under the assumption that the random censoring time has no relation with the survival time,

$$l(\lambda_i \mid t)=\lambda_i^{\delta}\exp(-\lambda_i(t-\tau_{i-1})) \quad \text{for } t\in I_i$$

where δ is the indicator function of a death of the individual in the time interval I_i. We denote the hazard rate for the time interval I_i by λ_i and the number of individuals alive at the beginning of the time interval I_i by n_i. Let T_{ij} be the survival time of the j-th individual alive at the beginning of the time interval I_i. Also let t_{ij} be an observed failure time corresponding to T_{ij} , which can be interpreted as the minimum of the survival time of the j-th individual and the corresponding censoring time. Let D_i be a set of the information from all observations of each time interval, $I_1,\cdots,I_i$, and let $D_{i-1(j)}$ be a set containing D_{i-1} and information from first j observations of the time interval I_i so that $D_{i-1(0)}=D_{i-1}$ and $D_{i-1(n_i)}=D_i$.

Then the discount survival model for the no covariate case in the time interval I_i can be defined as follows.

i) observation equation ;

$$T_{ij}\sim \exp(\lambda_i) \quad \text{for } i=1,\cdots,s.$$

ii) evolution equation;

$$(\lambda_i \mid D_{i-1})\sim Ga(d_i b_{i-1}, d_i r_{i-1}),$$

where b_{i-1} and r_{i-1} are parameters of the posterior distribution of λ_{i-1} given D_{i-1} , d_i $(0 < d_i < 1)$ is the discount factor of the time interval I_i .

3. Inference Procedures

The process is started with the initial distribution of λ_0 , $Ga(b_0, r_0)$, where b_0 and r_0 are given prior to the time interval I_1. At the end of each time interval I_{i-1} the posterior distribution of λ_{i-1} is obtained as

$$(\lambda_{i-1} | D_{i-1}) \sim Ga(b_{i-1}, r_{i-1}),$$

which leads to the prior distributions of λ_i at the beginning of each time interval I_i ,

$$(\lambda_i | D_{i-1}) \sim Ga(b_{i0}, r_{i0}),$$

where

$$b_{i0} = d_i b_{i-1}, \quad r_{i0} = d_i r_{i-1}$$

with the discount factor of the time interval I_i, $d_i^{\cdot}$ $(0 < d_i < 1)$.

With the information from first j observations, the joint prior distribution of the updated distribution of λ_i is obtained as

$$(\lambda_i | D_{i-1(j)}) \sim Ga(b_{ij}, r_{ij}),$$

where

$$b_{ij} = b_{i,j-1} + \delta_{ij}, \quad r_{ij} = r_{i,j-1} + t_{ij} - \tau_{i-1}$$

and δ_{ij} is the indicator function of a death of the j-th individual alive at the beginning of the time interval I_i .

When all individuals in the time interval I_i are observed, the posterior distribution is obtained as

$$(\lambda_i | D_i) \sim Ga(b_i, r_i),$$

where

$$D_i = D_{i-1(n_i)}, \quad b_i = b_{i,n_i} \text{ and } r_i = r_{i,n_i} .$$

At the end of each time interval the posterior distribution of λ_i given D_i is used to obtain the survival function under the assumption of a conjugate gamma distribution of the hazard rate λ_i.

Using that

$$\begin{aligned} P(T > t | T > \tau_{i-1}, D_i) &= \int P(T > t - \tau_{i-1}, \lambda_i | D_i) d\lambda_i \\ &= \int P(T > t - \tau_{i-1} | \lambda_i, D_i) p(\lambda_i | D_i) d\lambda_i \\ &= \left(1 + \frac{t - \tau_{i-1}}{r_i}\right)^{-b_i} \quad \text{for } t \in I_i, i = 1, \cdots, s, \end{aligned}$$

at the end of the time interval I_i the survival function is obtained as

$$P(T > t | D_i) = \left(1 + \frac{t - \tau_{i-1}}{r_i}\right)^{-b_i} \prod_{k=1}^{i-1} \left(1 + \frac{\tau_k - \tau_{k-1}}{r_k}\right)^{-b_k} \quad \text{for } t \in I_i.$$

At the beginning of the time interval I_{i+1} the prior distribution of λ_{i+1} given D_i is obtained, which is used to obtain the forecasted survival function of the time interval I_i given D_i,

$$(\lambda_{i+1} | D_i) \sim Ga(d_{i+1} b_i, d_{i+1} r_i).$$

Thus the forecasted survival function of the time interval I_{i+1} based on the information gathered until the end of the time interval I_t is obtained as

$$P(T>t|D_i) = (1+\frac{t-\tau_i}{d_{i+1} r_i})^{d_{i+1} b_i} \prod_{k=1}^{i} (1+\frac{\tau_k-\tau_{k-1}}{r_k})^{-b_k} \quad \text{for } t \in I_{i+1}.$$

4. Illustrations

In this section, the performance of the estimation proposed in previous sections is illustrated via the data set in Table 1, which consists of survival times of 45 gastric carcinoma patients of treatment of the chemotherapy.

Table 1. Survival Times in Days

1 63 105 129 182 216 250 262 301 301 342 354 356 358 380 381c 383 383 388 394 408
460 489 499 524 529c 535 562 675 676 748 748 778 786 797 945c 955 968 1180c 1245
1271 1277c 1397c 1512c 1519c

c: censored, source: Stablein et al.(1981)

In the discount survival model, the survival time of the individuals alive at the beginning of time interval I_i is assumed to have the exponential distribution with the hazard rate λ_i. The end point of each time interval is taken as the multiple of 30 days. We start the analysis with the initial distribution of $(\lambda_0 | D_0) \sim Ga(1, 1362)$, where values of parameters are obtained by equating the PL-estimate(Kaplan and Meier,1958) of the time interval I_1 to the survival function of the time interval I_1, where the PL-estimate is introduced not to be compared with but to be used as a criterion of the estimtes under the discount survival model. The value of the discount factor is tentatively obtained by comparing estimated survival functions from different values of the discount factor - 0.85, 0.9, 0.95. The value d_i = 0.9 for $i=1,\cdots,51$, which gives the closest estimated survival function to the PL-estimate, is chosen. Figure 1 shows the cumulative hazard rate obtained under the discount survival model and the cumulative hazard rate estimated from the PL-estimate - the cumulative hazard rate at time i is the negative value of the logarithm of the PL-estimate. In the figure the estimates show the similar pattern in the increase of the cumulative hazard rate. Figure 2 shows the 1-step ahead forecasted survival function and the PL-estimate. In the figure the 1-step ahead forecasted survival function under the discount survival model looks close to the PL-estimate.

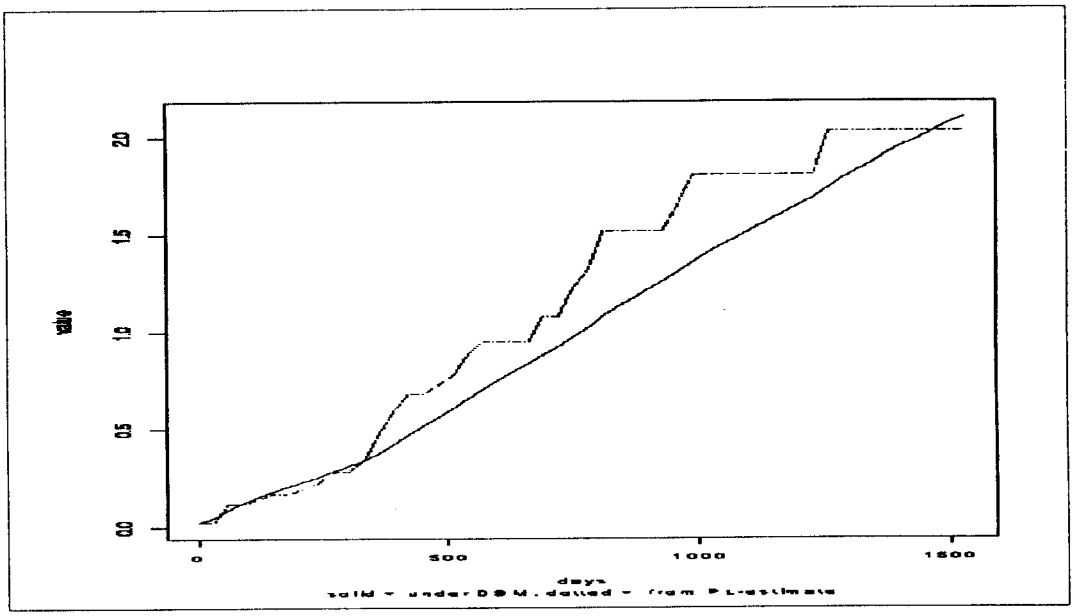

Figure 1. Estimated Cumulative Hazard Rate

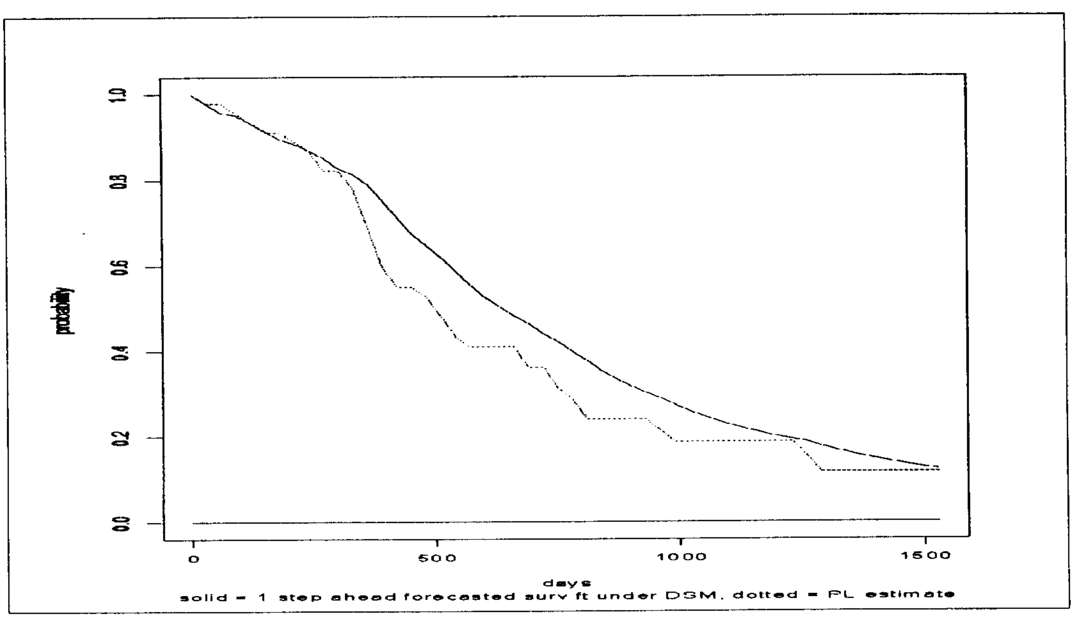

Figure 2. Forecasted Survival Function under DSM and PL-estimate

References

[1] Ameen, J. R. M. and Harrison, P. J. (1985), Normal Discount Bayesian Models, *Bayesian Statistics*, 2nd edition, 271-291.

[2] Bolstad, W. M. (1995), The Multiprocess Dynamic Poisson Model, *Journal of the American Statistical Association,* 90, 227-232.

[3] Gamerman, D. (1991), "Dynamic Bayesian Models for Survival Data", *Applied Statistics,* 40, 63-79.

[4] Kaplan, E. L. and Meier, P. (1958), "Nonparametric Estimation from incomplete Observations", *Journal of the American Statistical Association,* 53, 457-481.

[5] Stablein, D. M., Carter, W. H. and Novak, J. W. (1981). "Analysis of survival with nonproportional hazard functions", *Controlled Clinical Trials, 2,* 149-159.

[6] West, M., Harrison, P. J. and Migon, H. S. (1985), "Dynamic Generalized Linear Models and Bayesian Forecasting", *Journal of the American Statistical Association,* 80, 73-97.