ACTUARIAL RESEARCH CLEARING HOUSE
1993 VOL. 1

BAYESIAN ESTIMATION OF TABULAR SURVIVAL MODELS FROM COMPLETE SAMPLES

A.K.Singh, M.A.Ananda, R.Dalpatadu

Departmment of Mathematics
University of Nevada, Las Vegas

ABSTRACT

The problem of estimation of a tabular survival model from complete samples is considered from a Bayesian approach. If a life test involves a large cohort group, then data is typically given as a grouped data set. The survival function S(t) is usually estimated by the moment estimator. We approximate the joint prior distribution of the unknown unconditional failure probabilities by a Dirichlet distribution, and then use a linear relationship between S(t) and the conditional survival probabilities to obtain a Bayes estimator of S(t).

INTRODUCTION

There are two basic study designs used in a clinical trial or by the actuary in a life insurance situation (london, 1988):

I: <u>COMPLETE SAMPLE</u>

A study unit of size n comes under observation at a well-defined time t=0, and are observed over time until all have died.

II: <u>CENSORED OR INCOMPLETE SAMPLE</u>

A sampling unit is allowed to come under observation at time t>0, and not all units will die at the end of the study.

One of the goals of the clinical trial is to estimate the survival function

S(t) = P(a sampling unit survives till time t).

Maximum likelihood estimator of S(t) is given in London(1988).
In this paper, we consider Bayesian estimation of S(t) from a grouped (tabular) complete sample in a Bayesian framework.

NOTATIONS AND STATEMENT OF PROBLEM

Let n be the initial size of the cohort under study. Divide the time axis $[0,\infty)$ into an indefinite number of equal intervals: $(0,1]$, $(1,2]$, ..., $(t,t+1]$,...,$(k-1,k]$ such that each observed time to death falls in one of k intervals (where k is a sufficiently large positive integer).

Let

d_u = # of deaths in the (t+1)st interval $[t,t+1)$, $0 \le t \le k-1$

n_u = # of survivors at time t

Then

$$n = \sum_{t=0}^{k-1} d_t$$

and

$$n_{t+1} = n_t - d_t,\ t=0,1,\dots,k-1.$$

The joint pdf of the random variables d_u, t=0,1,...,k-1 is:

$$f(d_0,d_1,\dots,d_{k-1}|\theta_0,\theta_1,\dots,\theta_{k-1}) = \frac{n!}{\prod_{t=0}^{k-1} d_t} \prod_{t=0}^{k-1} \theta_t \qquad (1)$$

where

θ_t = P(death between time t and t+1)

= P(death between time t and t+1 | unit alive at t) P(unit alive at time t)

= $q_u\, S(t)$, t=0,1,...,k-1 (2)

and

$$\sum_{t=0}^{k-1} \theta_t = 1 \ .$$

In Bayesian framework, the vector $(\theta_0,\theta_1,\dots,\theta_{k-1})$ itself is random with prior distribution $g(\theta_0,\theta_1,...,\theta_{k-1})$. We assume that the prior joint pdf of $(\theta_0,\theta_1,...,\theta_{k-1})$ is the natural conjugate

Dirichlet distribution $D(\alpha_0,\alpha_1,...,\alpha_{k-1})$

$$g(\theta_0,\theta_1,...,\theta_{k-1}) = \frac{\Gamma(\sum_{i=0}^{k-1}\alpha_i)}{\prod_{i=0}^{k-1}\Gamma(\alpha_i)}\prod_{i=0}^{k-1}\theta_i \ , \ \sum_{i=0}^{k-1}\theta_i = 0 \qquad (3)$$

The following results are needed:

(1) The first two moments of the Dirichlet distribution are (Johnson, 1979):

$$E(\theta_i) = \frac{\alpha_i}{\alpha_0} \ , \qquad (4)$$

$$Var(\theta_i) = \frac{\alpha_i(\alpha_0-\alpha_i)}{\alpha_0^2(\alpha_0+1)} \ , \qquad (5)$$

and

$$Cov(\theta_i,\theta_j) = -\frac{\alpha_i\alpha_j}{\alpha_0^2(\alpha_0+1)} \qquad (6)$$

where

$$A = \sum_{i=0}^{k-1}\alpha_i \ .$$

(2) The posterior jpdf of $(\theta_0,\theta_1,...,\theta_{k-1})$ given $(d_1,d_2,...,d_{1.2})$ is easily shown to be also Dirichlet:

$$D(\alpha_0+d_0,\alpha_1+d_1,...,\alpha_{k-1}+d_{k-1}) \ .$$

We will now use the above results to obtain:

(i) exact Bayes estimates of the survival function S(t),

and

(ii) adaptive Bayes estimates of q_u.

We will assume the following loss function:

$$L(\theta,\hat{\theta}) = \sum_{t=0}^{k-1}(\theta_t-\hat{\theta}_t)^2 \quad . \tag{7}$$

It follows from (1) and (2) above and a result in Lehmann (1983) that the Bayes estimate of θ_t is

$$\hat{\theta}_{tB} = E(\theta_t|d_0,...,d_t) = \frac{\alpha_t+d_t}{A+n} \tag{8}$$

with posterior risk

$$Var(\hat{\theta}_t|d_0,...,d_{k-1}) = \frac{(\alpha_t+d_t)(A+n-\alpha_t-d_t)}{(A+n)^2(A+n+1)} \quad . \tag{9}$$

BAYES ESTIMATE OF S(t)

We know that

$$S(t) = \prod_{i=0}^{t-1} p_i \tag{10}$$

where

$p_t = 1 - q_t$ = conditional survival probability.

The above relationship can be used to obtain a plug-in estimate of S(t) using estimates of p_u. It is easy to see, however, that the estimate of S(t) obtained by plugging in Bayes estimate of p_u is not a Bayes estimate

We use the following relationship

$$\theta_t = S(t)q_t = S(t)(1-p_t) = S(t)-S(t+1) \tag{11}$$

to obtain a Bayes estimate of S(t), as follows:

$$\theta_0 = S(0)-S(1) \rightarrow S(1) = S(0)-\theta_0 = 1-\theta_0$$

and therefore

$$\hat{S}(1) = E[S(1)|d_0,\ldots,d_{k-1}] = E[1-\theta_0|d_0,\ldots,d_{k-1}] = 1-\hat{\theta}_{0B}$$

$$S(2) = \theta_1-S(1) \rightarrow \hat{S}(2) = E[\theta_1-S(1)|d_0,\ldots,d_{k-1}) = \hat{\theta}_{1B}-\hat{S}_B(1)$$

...
...

$$\theta_{k-1} = S(k-1)-S(k) = S(k-1) \rightarrow \hat{S}_B(k-1) = \hat{\theta}_{(k-1)B}$$

ADAPTIVE BAYES ESTIMATE OF q_u

Since $\theta_t = S(t)q_t$, an adaptive Bayes estimate of q_u is:

$$\hat{q}_t = \frac{\hat{\theta}_t}{\hat{S}(t)} \, .$$

Adaptive Bayes estimates of other parameters or functions of interest can be similarly obtained.

EXAMPLES

We will now present a few examples of our method for the purpose of illustration of our procedure.

EXAMPLE 1:

Following Monte Carlo simulation experiment was used to generate data sets for some of these examples:

INPUT REQUIRED: n, k, $\alpha_1, \ldots, \alpha_{1.2}$

STEP 1: Generate $(\theta_1, \ldots, \theta_{1.2})$ from the Dirichlet distribution

$D(\alpha_0,\ldots,\alpha_{k-1})$ given by (3) using the following result:

If $X_1, X_2, \ldots, X_{1.2}$ are independent Gamma rv's with shape parameters $\alpha_1, \alpha_2, \ldots, \alpha_{1.2}$, respectively, and common scale parameter 1, then

$$(Y_1, Y_2, \ldots, Y_{1.2}) \text{ is } D(\alpha_0, \alpha_1, \ldots, \alpha_{k-1}) .$$

The IMSL subroutine RNGAM was used to generate an independent Gamma random vector $(\theta_0, \theta_1, \ldots, \theta_{k-1})$.

STEP 2: Generate $(d_1, d_2, \ldots, d_{1.2})$ from the conditional multinomial distribution $M(n, \theta_0, \theta_1, \ldots, \theta_{k-1})$.

The IMSL subroutine RNMTN was used for this step.

In this example, k = 5, n = 20, and the joint pdf of $(\theta_1, \ldots \theta_1)$ is taken to be a Dirichlet distribution with parameters (1,1,...,1).

The value of $(\theta_1, \ldots \theta_6)$, generated from this Dirichlet pdf, is

(0.4455, 0.0154, 0.4516, 0.0398, 0.0477).

The value of the random vector $(x_2, \ldots, x_6)$ generated from the conditional jpdf (1) is

(10,1,7,2,0).

The results of our calculations are shown in Table 1.

EXAMPLE 2: In this example (London, 1988) a sample of 20 individuals exist at time t=0, and all fail within 5 weeks, with 2 failing in the first week, 3 in the second, 8 in the third, 6 in the fourth, and 1 in the last week:

$d_1 = 2,\ d_2 = 3,\ d_3 = 8,\ d_4 = 6,\ d_5 = 1.$

We will assume a Dirichlet(1,1,1,1,1) prior.

The MLE and Bayes estimates of θ and S(t) are given in Table 2:

TABLE 1: Bayes estimate of $(\theta_1, \ldots \theta_6)$ and $S(t)$ for Example 1

TRUE					
θ	0.4455	0.0154	0.4516	0.0398	0.0477
S(t) given by (11)	0.5545	0.5391	0.0875	0.0477	0
MLE					
θ	0.5	0.05	0.35	0.10	0
S(t)	0.5	0.45	0.10	0	0
BAYES ESTI-MATES					
θ	0.44	0.08	0.32	0.12	0.04
S(t)	0.56	0.48	0.16	0.04	0

Table 2: MLE and Bayes estimates of θ and S(t) for Example 2

MLE (from London, 1988)					
θ	0.1	0.15	0.4	0.3	0.05
S(t)	0.9	0.75	0.35	0.05	0
BAYES ESTI-MATES					
θ	0.12	0.16	0.36	0.28	0.08
S(t)	0.88	0.72	0.36	0.08	0

REFERENCES:

Dick London (1988). Survival Models and Their Estimation, ACTEX Publications, Winsted and New Britain, Connecticut, pp. 77-94.

N.L. Johnson (1979). Distributions in Statistics: Continuous Multivariate Distributions, John Wiley, New York, p. 233.

E.J. Lehmann (1983). Theory of Point Estimation, John Wiley, New York, p. 293.