International Journal of Reliability and Application
Vol. 2, No. 1, pp. 49-56, 2001

Parametric Empirical Bayes Estimation of A Constant Hazard with Right Censored Data

Mostafa Mashayekhi
Department of Finance
University of Nebraska-Lincoln, Lincoln, Nebraska 68588-0426, USA

Abstract. In this paper we consider empirical Bayes estimation of the hazard rate and survival probabilities with right censored data under the assumption that the hazard function is constant over the period of observation and the prior distribution is gamma. We provide an estimator of the first derivative of the prior moment generating function that converges at each point to the true value in L_2 and use it to obtain, easy to compute, asymptotically optimal estimators under the squared error loss function.

Key words : *Empirical Bayes, Parametric empirical Bayes, Asymptotically optimal, Survival probabilities, Constant hazard rate, Right censored data.*

1. INTRODUCTION

Since the introduction of the empirical Bayes approach by Robbins (1955) statisticians have employed the method in many areas of statistics (see for example Morris (1983) for a remarkable list of such areas) to utilize similar information in constructing estimators.

In an empirical Bayes decision problem there are m independent random pairs $(\theta_1, X_i),...,(\theta_m, X_m)$ such that $X_1,...,X_m$ are observable and the distribution of X_i depends on the parameter θ_i. The parameters $\theta_1,...,\theta_n$ are assumed to be *i.i.d.* with an unknown common *a priori* distribution G. There is a non-negative loss function L and the task is to find decision rules $t_m(.) = t_m(X_1,...,X_m,.)$ that are asymptotically optimal in the sense that with E denoting the overall expectation

$$E[L(t_m(X_m),\theta_m)] - \min_t E[L(t(X_m),\theta_m)] \longrightarrow 0 \text{ as } m \longrightarrow \infty .$$

Robbins (1955) introduced the empirical Bayes method in a nonparametric framework in the sense that in his approach G is completely unspecified. In the parametric empirical Bayes approach, that was later explored by Efron and Morris (1973, 1975), the prior distribution G belongs to a parametric family of distributions.

In this paper we consider parametric empirical Bayes estimation of the hazard rate and survival probabilities, with data obtained by observing each cohort for one unit of time, under the assumption that each hazard rate is constant over the period of observation.

In Section 2 we formalize our model and verify some preliminary results. In Section 3 we present our estimators and prove their asymptotic optimality under the squared error loss function.

2. ASSUMPTION, NOTATION AND PRELIMINARIES

The main characteristics of our model are formalized in the following assumption.

Assumption 1. For each $i \in \{1,...,m\}$, $(X_{i1},...,X_{in_i},\theta_i)$ are independent random vectors such that

(i) conditional on θ_i, $X_{i1},...X_{in_i}$ are *i.i.d.* with survival function

$$\overline{F}(x) = 1_{[x\le 0]} + e^{-\theta_i(x)}1_{[0<x<1]},$$

(ii) θ_i's are identically distributed with gamma density

$$g(\theta) = (\Gamma(\alpha))^{-1}\beta^{\alpha}\theta^{\alpha-1}e^{-\beta.\theta}1_{[\theta>0]}$$

where $\alpha \in (0,N_1]$ and $\beta \in (0,N_2]$ for some pair of integers N_1 and N_2.

(iii) $\sup_{i\ge 1} n_i < \infty$.

We sometimes use $1_{[\ ]}$ to denote an indicator function and sometimes use it to denote the value of an indicator function at a given point, with the distinction being clear from the context. For each $u \in [-1,0]$, we use $M(u)$ to denote the moment generating function of the prior distribution evaluated at u. We use $\delta_{ij}(u)$ to denote $1_{[X_{ij}\ge -u]}$, and use $\overline{\delta}_i(u)$ to denote $n_i^{-1}\sum_{j=1}^{n_i}\delta_{ij}(u)$. The symbol $\overline{\delta}_{..}(u)$ denotes $n_T^{-1}\sum_{i=1}^{m}n_i\overline{\delta}_i(u)$ where

$$n_T = \sum_{i=1}^{m} n_i .$$

Note that $E(\delta_{ij}(u)\mid\theta_i) = e^{u\theta_i}$ and hence $E(\delta_{ij}(u)) = M(u) = (\beta/(\beta-u))^{\alpha}$. We also have

$$E(\overline{\delta}_{..}(u)) = n_T^{-1}\sum_{i=1}^{m}n_iM(u) = M(u) = (\beta/(\beta-u))^{\alpha}.$$

Observe that since each $\bar{\delta}_i(u) \le 1$, we have $Var\bar{\delta}_i(u) \le E\bar{\delta}_i^2 \le 1$. Therefore

$$E(\bar{\delta}_{..}(u) - M(u))^2 = Var(\bar{\delta}_{..}(u) = n_T^{-2} \sum_{i=1}^{m} n_i^2 Var(\bar{\delta}_i(u) \le m^{-1} \sup_{i\ge 1} n_i \quad (1)$$

Let ϑ be the Lebesgue measure on (0,1] and let ν be the probability measure degenerate at 1. When X is a random variable with survival function $\bar{F}(x) = 1_{[x\le 0]} + e^{-\theta(x)} 1_{[0<x<1]}$, a density (Radon-Nikodym derivative) of the distribution of X with respect to $\mu = \vartheta + \nu$ is given by

$$f_\theta(x) = \theta e^{-\theta.x} 1_{[0<x<1]} + e^{-\theta} 1_{[x=1]} = \theta^{1_{[x<1]}} e^{-\theta.x} 1_{[0<x\le 1]}. \quad (2)$$

Suppose θ has a gamma distribution with parameters α and β, and conditional on θ, the random variables $X_1,...,X_n$ are *i.i.d.* with density given in (2). Then a posterior density of θ given $X_1 = x_1,..., X_n = x_n$ is given by

$$g(\theta \mid x_1,...x_n) = C(x_1,...x_n)\theta^{(\alpha + \sum 1_{[x_i<1]})-1} e^{-(\beta+\sum x_i)\theta} 1_{[\theta>0]} \quad (3)$$

which is a gamma density. Hence the Bayes estimator of θ under the squared error loss is

$$E[\theta \mid X_1,..., X_n] = (\beta + \sum X_i)^{-1}(\alpha + \sum 1_{[X_i<1]}) \quad (4)$$

and for each $t \in (0,1)$ the Bayes estimator, under the squared error loss, of the survival probability $e^{-\theta.t}$ is given by

$$E[e^{-\theta.t} \mid X_1,..., X_n] = \left(\frac{\beta + \sum X_i}{\beta + t + \sum X_i} \right)^{\alpha + \sum 1_{[X_i<1]}}. \quad (5)$$

3. ASYMPTOTICALLY OPTIMAL EMPIRICAL BAYES ESTIMATOR

In the remaining of the paper we use $\hat{\theta}_m^B$ to denote Bayes estimator of θ_m, and use $\hat{p}_m^B(t)$ to denote Bayes estimator of $e^{-\theta_m t}$. The corresponding empirical Bayes estimators are denoted by $\hat{\theta}_m^{EB}$ and $\hat{p}_m^{EB}(t)$.

In the following theorem we give sufficient conditions for asymptotic optimality of empirical Bayes estimators of θ_m and $e^{-\theta_m t}$ based on estimators of the prior parameters α and β and then provide estimators that satisfy the given sufficient conditions.

Note that all of the following results are under Assumption 1. Therefore we will not mention this assumption in the statement of every theorem or lemma. All incompletely described limits in this paper are as $m \longrightarrow \infty$ through positive integers. The symbol □ denotes end of proof.

Theorem 1. *Let $0 < \hat{\alpha} \le N_1$ and $0 < \hat{\beta} \le N_2$ be estimators of α and β such that $\hat{\alpha} \xrightarrow{P} \alpha$ and $\hat{\beta} \xrightarrow{P} \beta$, Let*

$$\hat{\theta}_m^{EB} = (\hat{\beta} + \sum_{j=1}^{n_m} X_{mj})^{-1}(\hat{\alpha} + \sum_{j=1}^{n_m} 1_{[X_{mj} < 1]}) \tag{6}$$

and for $t \in (0,1)$ let

$$\hat{p}_m^{EB}(t) = \left(\left(\hat{\beta} + t + \sum_{j=1}^{n_m} X_{mj} \right)^{-1} \left(\hat{\beta} + \sum_{j=1}^{n_m} X_{mj} \right) \right)^{\hat{\alpha} + \sum_{j=1}^{n_m} 1_{[X_{mj}<1]}} . \tag{7}$$

Then $\hat{\theta}_m^{EB}$ and $\hat{p}_m^{EB}(t)$ are asymptotically optimal estimators of θ_m and $p_m(t) = e^{-\theta_m t}$ respectively.

Proof. Let $A = \sum_{j=1}^{n_m} 1_{[X_{mj} < 1]}$ and $B = \sum_{j=1}^{n_m} X_{mj}$. Let $\| \ \|$ denote the L_2-norm defined by $\| Y \| = (E[Y^2])^{1/2}$. Observe that both A and B are non-negative and less than or equal to $\sup_{i \ge 1} n_i$ with probability 1 since for each j, $P[0 < X_{mj} \le 1] = 1$. Therefore

$$\| \hat{\theta}_m^{EB} - \hat{\theta}_m^B \| \le \beta^{-1}[(\beta + B) \| \hat{\alpha} - \alpha \| + (\alpha + A) \| \hat{\beta} - \beta \|] \longrightarrow 0 \tag{8}$$

by the triangle inequality and the Bounded Convergence Theorem since $| \hat{\alpha} - \alpha | < N_1$ and $| \hat{\beta} - \beta | < N_2$.

Since $E(\theta_m^2) < \infty$, by Lemma 2.1 of Singh (1979)

$$E[\hat{\theta}_m^{EB} - \theta_m]^2 - E[\hat{\theta}_m^B - \theta_m]^2 = \| \hat{\theta}_m^{EB} - \hat{\theta}_m^B \|^2 .$$

Therefore asymptotic optimality of $\hat{\theta}_m^{EB}$ follows from (8).

To prove asymptotic optimality of $\hat{p}_m^{EB}(t)$ observe that

$$|(\beta+B)^{-1}(\hat{\beta}+B)-1| \le |\beta^{-1}(\hat{\beta}-\beta)| \xrightarrow{P} 0 \tag{9}$$

by the assumed property of $\hat{\beta}$. Therefore by continuity of log

$$\log((\hat{\beta}+B)/(\beta+B)) \xrightarrow{P} 0. \tag{10}$$

Similarly

$$\log((\hat{\beta}+B+t)/(\beta+B+t)) \xrightarrow{P} 0. \tag{11}$$

We have

$$\begin{aligned}\log(\hat{p}_m^{EB}(t)/(\hat{p}_m^{B}(t)) = {} & [(\hat{\alpha}-\alpha)+(\alpha+A)][\log((\hat{\beta}+B)/(\beta+B)) \\ & -\log((\hat{\beta}+B+t)/(\beta+B+t))] \\ & +(\hat{\alpha}-\alpha)\log((\beta+B)/(\beta+B+t)).\end{aligned} \tag{12}$$

Since $(\beta+B)/(\beta+B+t)$ is increasing in B and its log is negative,

$$|\log((\beta+B)/\beta+B+t))| \le |\log(\beta/(\beta+t))|. \tag{13}$$

Therefore by (10) and (11) and (13) the lhs of (12) $\xrightarrow{P} 0$. Thus by continuity of $f(x)=e^x$ we have $\hat{p}_m^{EB}(t)/\hat{p}_m^{B}(t) \xrightarrow{P} 1$. It follows that

$$|\hat{p}_m^{EB}(t)-\hat{p}_m^{B}(t)| \le |(\hat{p}_m^{EB}(t)/\hat{p}_m^{B}(t))-1| \xrightarrow{P} 0.$$

Since $\hat{p}_m^{EB}(t)$ and $\hat{p}_m^{B}(t)$ are both positive and less than 1, by the Bounded Convergence Theorem it follows that

$$E[\hat{p}_m^{EB}(t)-\hat{p}_m^{B}(t)]^2 \longrightarrow 0.$$

Since $p_m(t)$ is bounded by Lemma 2.1 of Singh (1979)

$$E[\hat{p}_m^{EB}(t)-p_m(t)]^2 - E[\hat{p}_m^{B}(t)-p_m(t)]^2 = E[\hat{p}_m^{EB}(t)-\hat{p}_m^{B}(t)]^2 \longrightarrow 0. \quad \square$$

The following theorem provides an estimator of the derivative of the prior moment generating function that converges at each point to the true value in L_2.

Theorem 2. *Let $t \in (-1,0]$ and let $t_m < t$ be such that with $h_m = t - t_m$, $h_m \longrightarrow 0$ and $h_m(m^{1/2}) \longrightarrow \infty$ as $m \longrightarrow \infty$. Let $\hat{M}'(t) = h_m^{-1}[\bar{\delta}_{..}(t) - \bar{\delta}_{..}(t_m)]$. Then with $M(t)$ denoting the moment generating function of θ evaluated at t,*

$$E[\hat{M}'(t) - M'(t)]^2 \longrightarrow 0 . \tag{14}$$

Proof. Let $\|\ \ \|$ denote the L_2-norm. Then by the triangle inequality

$$\begin{aligned}\| \hat{M}'(t) - M'(t) \| \le \| h_m^{-1}(\bar{\delta}_{..}(t_m) - M(t_m)) \| + \| h_m^{-1}(M(t) - \bar{\delta}_{..}(t)) \| \\ + | h_m^{-1}[M(t) - M(t_m)] - M'(t) | .\end{aligned} \tag{15}$$

By (1) the first and the second term on the rhs of (15) are both less than or equal to $(h_m m^{1/2})^{-1}(\sup_{i\ge 1} n_i)^{1/2}$ and the third term converges to zero, as $m \longrightarrow \infty$, by the definition of the derivative. □

The following lemma is used in the proof of Theorem 3.

Lemma 1. Suppose $\langle a_m \rangle, \langle b_m \rangle$, and $\langle \beta_m \rangle$ are sequences such that

(i) $a_m > 0$ for each m, and $a_m \longrightarrow a > 0$,

(ii) $b_m \longrightarrow ((\beta+1)^{-1}\beta)^{\beta.a}$ for some $\beta > 0$,

(iii) $\log((\beta_{m+1})\beta_m)^{\beta_m \cdot} = a_m^{-1} \log b_m$.

Then $\beta_m \longrightarrow \beta$.

Proof. Since $\log((\beta_m + 1)^{-1}\beta_m)^{\beta_m} = a_m^{-1}\log b_m \longrightarrow \log((\beta+1)^{-1}\beta)^{\beta}$ we have

$$\lim((\beta_m + 1)^{-1}\beta_m)^{\beta_m} = ((\beta+1)^{-1}\beta)^{\beta} . \tag{16}$$

Let $\beta_* = \underline{\lim}\beta_m$ and let $\beta^* = \overline{\lim}\beta_m$ and let $f(x) = ((x+1)^{-1}x)^x$. Then by (16) $f(\beta_*) = f(\beta^*)$.

Observe that

$$d \log f(x)/dx = \log x - \log(x+1) + (x+1)^{-1}$$

and

$$d^2 \log f(x)/dx^2 = x^{-1} - (x+1)^{-1} - (x+1)^{-2} = x^{-1}(x+1)^{-2} .$$

Since the second derivative of $\log(f(x)$ is positive on $(0,\infty)$ the first derivative of $\log(f(x)$ is strictly increasing on $(0,\infty)$. Since $\lim_{x\to\infty} d\log f(x)/dx = 0$, we have the first derivative of $\log(f(x)$ is negative on $(0,\infty)$. Hence $\log(f(x)$ is strictly decreasing and therefore is one-one. This means $f(x)$ is one-one on $(0,\infty)$ and therefore $\beta_* = \beta^*$. □

Let $f(x) = ((x+1)^{-1}x)^x$.
Observe that $\lim_{x\to 0} \log f(x) = 0$ and $\lim_{x\to\infty} \log f(x) = -1$. Since as shown in the proof of Lemma 1, $\log f(x)$ is strictly decreasing on $(0,\infty)$ the equation $\log f(x) = b$ has a solution in $(0,\infty)$ if and only if $-1 < b < 0$.

Theorem 3. *Let $\hat{M}'(0)$ be as in Theorem 2. Let $\tilde{\beta}$ be the solution of the equation*

$$\log((x+1)^{-1}x)^x = (\hat{M}'(0))^{-1}\log\bar{\delta}_{..}(-1) \qquad (17)$$

if $-1 < \hat{M}'(0))^{-1}\log\bar{\delta}_{..}(-1) < 0$, and let $\tilde{\beta} = N_2$ otherwise. Let $\tilde{\alpha} = \hat{\beta}\hat{M}'(0)$ and let $\hat{\alpha} = \min(\tilde{\alpha}, N_1)$ and $\hat{\beta} = \min(\tilde{\beta}, N_2)$. Then $\hat{\beta} \xrightarrow{P} \beta$ and $\hat{\alpha} \xrightarrow{P} \alpha$.

Proof. By Theorem 1 $\hat{M}'(0) \xrightarrow{P} M'(0) = \alpha/\beta$
and by (1)

$$\bar{\delta}_{..}(-1) \xrightarrow{P} M(-1) = ((\beta+1)^{-1}\beta)^{\beta.M'(0)}.$$

Use the fact (see for example Billingsley (1986)) that a sequence x_m converges in probability to x if and only if every subsequence of x_m has a further subsequence that converges to x with probability 1. Let $\langle m_k \rangle$ be a subsequence of $\langle m \rangle$. Then there is a further subsequence $\langle m_{k_j} \rangle$ along which the right hand side of (17) converges to $\log((\beta+1)^{-1}\beta)^{\beta}$ with probability 1 and hence eventually becomes negative and more than -1. Therefore with probability 1 the estimator $\hat{\beta}$ eventually becomes the solution of (17) along $\langle m_{k_j} \rangle$ and hence converges to β by Lemma 1. This means that $\tilde{\beta} \xrightarrow{P} \beta$ and hence $\tilde{\alpha} \xrightarrow{P} \alpha$. It follows that $\hat{\beta} \xrightarrow{P} \beta$ and $\hat{\alpha} \xrightarrow{P} \alpha$ because $|\hat{\alpha} - \alpha| \le |\tilde{\alpha} - \alpha|$ and $|\hat{\beta} - \beta| \le || \tilde{\beta} - \beta|$. □

4. CONCLUDING REMARKS

If instead of assuming a gamma prior we assume the range of θ_i's is compact then our model will satisfy the assumptions of Datta (1991) which provides nonparametric admissible asymptotically optimal empirical Bayes estimators based on a hyperprior with full support. However the computation of Datta's estimators is not trivial.

The advantage of the estimators offered in this paper is that they are very easy to compute and therefore can be used very easily in practice. While it may be argued that this simplicity has been achieved at the expense of sacrificing some robustness by choosing a parametric model it is also worth mentioning that parametric empirical Bayes is considered by some statisticians (see for example Morris (1983)) as being closer to Bayesian and therefore more attractive. For more comments regarding the comparison of parametric and nonparametric empirical Bayes see the discussion of Morris (1983) by James Berger.

REFERENCES

Billingsley, P. (1986), *Probability and Measure.* Wiley, New York.

Datta, S. (1991), On the Consistency of Posterior Mixtures and its Applications. *Ann. Statist.*, **19**, 338-353.

Efron, B.,and Morris, C. (1973), Stein's Estimation Rule and Its Competitors-An Empirical Bayes Approach. *Journal of the American Statistical Association*, **68**, 117-130.

Efron, B., and Morris, C. (1975), Data Analysis Using Stein's Estimator and Its Generalizations. *Journal of the American Statistical Association*, **70**, 311-319.

Morris, C.(1983), Parametric Empirical Bayes Inference: Theory and Applications (with discussion), *Journal of the American Statistical Association*, **78**, 47-65.

Robbins, H. (1955), An Empirical Bayes Approach to Statistics. *Proc. Third Berkeley Symp. Math. Statist. Prob. 1*, Univ.of California Press, 157-163.

Robbins, Herbert. (1964), The Empirical Bayes Approach to Statistical Decision Problems. *Ann. Math Statist.*, **35**, 1-20

Singh, R.S. (1979), Empirical Bayes Estimation in Lebesgue-Exponential Families With Rates Near the Best Possible Rate, *Ann Statist.* **7(4)**, 890-902.