Journal of the Korean Statistical Society (2001), **30**: *3, pp 481–498*

Empirical Bayes Nonparametric Estimation with Beta Processes Based on Censored Observations

Jee Chang Hong[1], Yongdai Kim[2] and Inha Jung[3]

ABSTRACT

Empirical Bayes procedure of nonparametric estimation of cumulative hazard rates based on censored data is considered using the beta process priors of Hjort(1990). Beta process priors with unknown parameters are used for cumulative hazard rates. Empirical Bayes estimators are suggested and asymptotic optimality is proved. Our result generalizes that of Susarla and Van Ryzin (1978) in the sense that (i) the cumulative hazard rate induced by a Dirichlet process is a beta process, (ii) our empirical Bayes estimator does not depend on the censoring distribution while that of Susarla and Van Ryzin (1978) does, (iii) a class of estimators of the hyperparameters is suggested in the prior distribution which is assumed known in advance in Susarla and Van Ryzin (1978). This extension makes the proposed empirical Bayes procedure more applicable to real data sets.

Keywords: Empirical Bayes nonparametric estimation; Censoring; Beta process; Lévy process; Dirichlet process; Martingale

1. INTRODUCTION

Empirical Bayes procedures by virtue of their using information from other similar data sets are useful when the sample sizes are small and data sets are sparse, in which case the maximum likelihood estimators may behave poorly. However, most of researches in the empirical Bayes problem have treated parametric models. Much fewer results treating nonparametric models are now available. This might be mainly due to the difficulties in finding suitable prior distributions. Since Ferguson's(1973) fundamental paper on Dirichlet process priors on

[1]Department of Liberal Arts, Dong Hae University, Donghae, Kangwon, 240-713, Korea

[2]Department of Statistics, Hankuk University of Foreign Studies, Yongin, Kyonggi, 447-791, Korea

[3]Division of Information and Computer Engineering, Ajou University, Suwon, Kyonggi, 442-749, Korea

the space of cumulative distribution functions(cdf), much of the works on nonparametric empirical Bayes decision problems have focused on the problems in which prior probabilities are placed on the space of cdf's.

In particular, Susarla and Van Ryzin (1978) provides a solution to an empirical Bayes problem of survival functions when the observations are censored on the right using the notion of Dirichlet process $\mathcal{D}(\nu)$, where ν is a finite measure on $\mathbf{R} = (-\infty, \infty)$. Their results, however, have the following limitations. Firstly, Susarla and Van Ryzin (1978) assumed that $\nu(\mathbf{R})$ is known in advance, which is hardly observed in practice. Secondly, their empirical Bayes procedure depends on the unknown censoring distribution, which should be estimated. However if the number of censored observations are small, estimation of the censoring distribution is not an easy task.

Deficiencies mentioned above are resolved to some extent in our work. We consider in this paper the nonparametric empirical Bayes decision problem of estimating the cumulative hazard rate (chr) with censored data based on the beta process priors. Hjort(1990) has introduced the beta process as a prior for the chr A. Note that since the chr induced by a Dirichlet process is a beta process(Hjort, 1990) our model extends the previous models based on the Dirichlet process prior.

By considering priors on the space of chrs, several advantages are obtained. Firstly, a natural empirical Bayes estimator can be constructed in the counting process framework. In particular, the proposed estimator does not depend on the unknown censoring distribution. Secondly, martingale techniques can be employed to prove the asymptotic optimality, and so many of the proofs become much simpler and better organized.

Before proceeding, we review several salient features of the beta process. Hjort(1990) introduced a nonparametric Bayesian model by placing a beta process on the space of the chr defined by

$$A(t) = \int_{[0,t]} \frac{dF(s)}{F[s,\infty)},$$

where F is a distribution function. Note that the cdf F is recovered from chr A by the relation

$$F(t) = 1 - \prod_{[0,t]} \{1 - dA(s)\}.$$

The definition of the beta process follows.

Definition 1.1. (Hjort(1990)) *Let α be a chr with a finite number of jumps taking place at $t_1, t_2, \cdots$, and let $c(\cdot)$ be a piecewise continuous, nonnegative function* *on* $[0, \infty)$. *Say that the Lévy process* A *is a beta process with parameters* $c(\cdot)$, $\alpha(\cdot)$, *and write as*

$$A \sim \text{beta}\{c(\cdot), \alpha(\cdot)\} \tag{1.1}$$

to indicate this, if the following holds : A has Lévy representation,

$$\mathrm{E}\left[\exp\{-\xi A(t)\}\right] = \left[\prod_{j:t_j \le t} \mathrm{E}\left\{\exp(-\xi S_j)\right\}\right] \exp\left\{-\int_0^1 (1 - e^{-\xi s}) dL_t(s)\right\} \tag{1.2}$$

with

$$S_j = A\{t_j\} \sim \text{beta}\{c(t_j)\alpha\{t_j\}, c(t_j)(1 - \alpha\{t_j\})\} \tag{1.3}$$

and

$$dL_t(s) = \left(\int_0^t c(z) s^{-1} (1-s)^{c(z)-1} d\alpha_c(z)\right) ds, \ t \ge 0, \ 0 < s < 1 \tag{1.4}$$

in which $\alpha_c(t) = \alpha(t) - \sum_{t_j \le t} \alpha\{t_j\}$ *is* α *with its jumps removed.*

The beta process has large support, its parameters have a nice interpretation and explicit Bayes estimators can be derived, both for A and for related quantities. Also, a particular transformation of a given Dirichlet process results in a special case of the beta process, and hence the class of beta processes is a much larger and more flexible class than that of Dirichlet processes. Hjort(1990) proved that the class of beta processes is also closed under possibly censored sampling. This is extremely useful in solving a Bayesian nonparametric estimation problem in which the Bayes estimator is given by its posterior mean. Given a chr A, let $X_1, \cdots, X_n$ be iid with the chr A and assume that $(T_1, \delta_1), \cdots, (T_n, \delta_n)$ are observed, where $T_i = \min\{X_i, C_i\}$, $\delta_i = I(X_i \le C_i)$ and $C_1, \cdots, C_n$ are censoring times. Define the counting process N and the left continuous process Y by

$$N(t) = \sum_{i=1}^n I\{T_i \le t, \delta_i = 1\}, \ Y(t) = \sum_{i=1}^n I\{T_i \ge t\}.$$

Assuming that the censoring times are either fixed or independent of the survival times X_i, we have

Theorem 1.1. (Hjort's(1990) corollary 4.1) *Let* $A \sim \text{beta}(c, \alpha)$. *Then*

$$A|(T_1, \delta_1), \cdots, (T_n, \delta_n) \sim \text{beta}\left\{c(\cdot) + Y(\cdot), \int_0^{(\cdot)} \frac{c d\alpha + dN}{c + Y}\right\}.$$

Remark 1.1. Since, by Theorem 1.1 posterior of a beta process A is also a beta process

$$\widetilde{A}(t) = \mathrm{E}(A(t)|(T_1, \delta_1), \cdots, (T_n, \delta_n)) = \int_0^t \frac{cd\alpha + dN}{c+Y}$$

defines a Bayes estimator of A and

$$\widetilde{F}(t) = 1 - \prod_{[0,t]}\{1 - d\widetilde{A}\}$$

is a posterior mean of $F(t)$ and therefore defines Bayes estimator of F under squared error losses.

This paper is organized as follows. In section 2, we describe the empirical Bayes decision problem of estimating the chr based on censored data. In section 3, various martingales are introduced and their properties are investigated. In section 4, with the aid of the martingale results developed in section 3, asymptotic optimality of the empirical Bayes estimators that can be obtained by a suitable class of estimated priors is proved. Finally, conclusion with some remarks are given in section 5.

2. THE EMPIRICAL BAYES DECISION PROBLEM

Let $\mathcal{A}$ be the space of all chr's on $[0, \infty)$ equipped with the σ-algebra $\Sigma_{\mathcal{A}}$ generated by Borel cylinder sets. If the squared error loss is used in the decision problem, $\widetilde{A}$, the Bayes estimator of A with respect to beta$\{c(\cdot), \alpha(\cdot)\}$ is the posterior mean of A, given data. If the parameters $c(\cdot)$ and $\alpha(\cdot)$ are known, the minimum Bayes risk at beta$\{c(\cdot), \alpha(\cdot)\}$ is attained by the Bayes estimator $\widetilde{A}(\cdot)$. When the parameters $c(\cdot)$ and $\alpha(\cdot)$ are unknown but there are independent repetitions of the component problem of the same type, one may apply the empirical Bayes approach of Robbins(1955) to have a sequence of estimators with risks converging to the minimum Bayes risk.

For each $j = 1, 2, \cdots$, let $\underline{X}_j = (X_{j1}, \cdots, X_{jn_j})$, $\underline{C}_j = (C_{j1}, \cdots, C_{jn_j})$ and A_j denote the n_j failure times, n_j censoring times and the chr, respectively, in the jth problem. Let $A_1, A_2, \cdots$ be iid beta$\{c(\cdot), \alpha(\cdot)\}$. Let $X_{j1}, \cdots, X_{jn_j}$ be iid F_j, given A_j and let $C_{j1}, \cdots, C_{jn_j}$ be iid G, a nonrandom (sub)cdf on $[0, \infty)$, where F_j is the random cdf on $[0, \infty)$ determined uniquely by the relation

$$F_j(t) = 1 - \prod_{[0,t]}\{1 - dA_j(s)\}. \tag{2.1}$$

Similarly, the cdf Φ_α corresponding to the chr $\alpha(\cdot)$ is given by

$$\Phi_\alpha(t) = 1 - \prod_{[0,t]}\{1 - d\alpha(s)\}. \tag{2.2}$$

We assume that the parameters $c(\cdot)$ and $\alpha(\cdot)$ satisfy the following conditions;
(C1) $\alpha(\cdot)$ is absolutely continuous and satisfies $\int_0^\infty \alpha(t)dw(t) < \infty$.
(C2) $c(\cdot) = k\Phi_\alpha[\cdot, \infty)$ for some positive constant k in $[K^{-1}, K]$.

Remark 2.1. (a) Under **(C1)**, Φ_α can be written as $\Phi_\alpha(t) = 1 - \exp(-\alpha(t))$ and Φ_α has density $\phi_\alpha(t) = \alpha'(t)\exp(-\alpha(t))$.

(b) Under **(C1)**, each A_j has no fixed points of discontinuity and therefore $\Delta A_j(t) = A_j(t) - A_j(t-) = 0$, a.s. for all $t \geq 0$.

Remark 2.2. (a) Under **(C2)**, parameters of the prior are essentially k and $\alpha(\cdot)$. Henceforth, we frequently write as $A_1, A_2, \cdots$ are iid with beta$\{k, \alpha\}$.

(b) Since in our model, $A_1, A_2, \cdots$ are assumed iid with beta$\{k, \alpha\}$, $F_1, F_2, \cdots$ are iid with $\mathcal{D}(k\Phi_\alpha)$, where both k and $\alpha(\cdot)$ are unknown(see Hjort(1990)). In Susarla and Van Ryzin(1978), $F_1, F_2, \cdots$ are iid with $\mathcal{D}(\nu)$, where $\nu([0, \infty))$ is assumed known. Our model removes this assumption since the parameter k in our model is unknown. Also $\mathcal{D}(k\Phi_\alpha)$ can be more flexible in practice than $\mathcal{D}(\nu)$.

Assume that $\{\underline{X}_j; j = 1, 2, \cdots\}$ and $\{\underline{C}_j; j = 1, 2, \cdots\}$ are independent and the data are subject to censoring. Thus by letting $T_{ji} = \min\{X_{ji}, C_{ji}\}$, $\delta_{ji} = I(X_{ji} \leq C_{ji})$ and $Z_{ji} = (T_{ji}, \delta_{ji})$ one observes only

$$\underline{Z}_j = (Z_{j1}, \cdots, Z_{jn_j}), \quad j = 1, 2, \cdots .$$

Our empirical Bayes model is concerned with the sequence of independent but non-identically distributed stochastic processes $\{(A_j, \underline{Z}_j); j = 1, 2, \cdots\}$, where $\underline{Z}_j$ is the sample of size n_j from the jth problem. The processes $A_1, A_2, \cdots$ are not observable, but iid with $Q_{k,\alpha}(\cdot)$ on $(\mathcal{A}, \Sigma_{\mathcal{A}})$, the distribution of a Lévy process $A \sim$ beta$\{k, \alpha\}$. For each $j = 1, 2, \cdots$, $i = 1, \cdots, n_j$ and $t \geq 0$, let

$$\begin{aligned} N_{ji}(t) &= I\{T_{ji} \leq t, \delta_{ji} = 1\}, \\ Y_{ji}(t) &= I\{T_{ji} \geq t\}. \end{aligned} \tag{2.3}$$

Then the number of failures by time t and the number at risk at time t observed from the jth component problem are respectively,

$$\begin{aligned} N_j(t) &= \sum_{i=1}^{n_j} N_{ji}(t), \\ Y_j(t) &= \sum_{i=1}^{n_j} Y_{ji}(t), \end{aligned} \tag{2.4}$$

and the total number of failures by time t and the total number at risk at time t in the first m component problems are given by

$$\begin{aligned} \overline{N}_m(t) &= \sum_{j=1}^{m} N_j(t), \\ \overline{Y}_m(t) &= \sum_{j=1}^{m} Y_j(t). \end{aligned} \tag{2.5}$$

Let $\rho_m(i) = \#\{1 \leq j \leq m; n_j = i\}$ denote the number of the component problems with sample size i among the first m problems. In addition to **(C1)** and **(C2)** we assume that the sequence $\{n_j\}$ satisfies

(C3) $\lim_{m\to\infty} \frac{1}{m}\rho_m(i) = \lambda_i$, $2 \leq i \leq M$.

Here M is an upper bound of $\{n_j\}$ and the lower bound 2 is related to the identifiability problem which arises in common in empirical Bayes. This will be discussed in section 5.

Let $H = [0,\infty) \times \{0,1\}$ and let $P_A^n(\cdot) = P^n(\cdot|A)$ denote the conditional distribution of $(Z_{11}, \cdots, Z_{1n})$ on H^n given chr A and let

$$\begin{aligned} P_{k,\alpha}^{(n)}(\cdot) &= P^{(n)}(\cdot|k,\alpha) \\ &= \int_{\mathcal{A}} P_A^n(\cdot) Q_{k,\alpha}(dA) \\ &= \text{the marginal distribution of } (Z_{11}, \cdots, Z_{1n}) \text{ on } H^n \\ &\quad\ \text{which is the } Q_{k,\alpha}\text{−mixture of } P_A^n(\cdot). \end{aligned}$$

At the $(m+1)$th problem, estimation of A_{m+1} is desired. Let the loss function at the $(m+1)$th problem be

$$L(\widehat{A}, A_{m+1}) = \int_0^\infty (\widehat{A}(t) - A_{m+1}(t))^2 dw(t), \tag{2.6}$$

where $w(\cdot)$ is a given weight function(a finite measure) on $[0, \infty)$ and $\widehat{A}$ is an estimator of A_{m+1} based on $\underline{Z}_{m+1}$. Bayes risk of $\widehat{A}$ at the prior $Q_{k,\alpha}(\cdot)$ is

$$R_{m+1}(\widehat{A}, (k, \alpha)) = \int_{\mathcal{A}} \int_{H_{m+1}} L(\widehat{A}, A) P^{n_{m+1}}(d\underline{z}_{m+1}|A) Q_{k,\alpha}(dA), \tag{2.7}$$

where $H_{m+1} = ([0, \infty) \times \{0, 1\})^{n_{m+1}}$ is the observation space of $\underline{Z}_{m+1}$. Minimum Bayes risk at the $(m+1)$th problem is

$$r_{m+1}(k, \alpha) = \inf_{\widehat{A}} R_{m+1}(\widehat{A}, (k, \alpha)), \tag{2.8}$$

where the infimum is taken over all possible estimators of A_{m+1}. Under the loss function (2.6), Bayes estimator of $A_{m+1}(t)$ is the posterior mean of $A_{m+1}(t)$, given $\underline{Z}_{m+1}$. By Theorem 1.1 and **(C2)**, Bayes estimator of A_{m+1} at beta$\{k, \alpha\}$ is given by

$$\begin{aligned} \widetilde{A}_{m+1}(t) &= \widetilde{A}_{m+1}(t; k, \alpha) \\ &= \int_{\mathcal{A}} A(t) Q_{k,\alpha}(dA|\underline{Z}_{m+1}) \\ &= \int_{[0,t]} \frac{k\Phi_\alpha[s, \infty) d\alpha(s) + dN_{m+1}(s)}{k\Phi_\alpha[s, \infty) + Y_{m+1}(s)}. \end{aligned} \tag{2.9}$$

In the case where (k, α) is known, one merely employs $\widetilde{A}_{m+1}$ and thereby incurs the minimum Bayes risk, i.e.,

$$r_{m+1}(k, \alpha) = R_{m+1}(\widetilde{A}_{m+1}, (k, \alpha)). \tag{2.10}$$

Suppose (k, α) is unknown and let $(\widehat{k}_m, \widehat{\alpha}_m)$ be an estimator based on $\underline{Z}_1, \cdots, \underline{Z}_m$. Bayes estimator $\widehat{A}_{m+1}(t)$ of $A_{m+1}(t)$ at the estimated prior beta$\{\widehat{k}_m, \widehat{\alpha}_m\}$ is taken as empirical Bayes estimator of $A_{m+1}(t)$. Thus, viewing (2.9) we have

$$\begin{aligned} \widehat{A}_{m+1}(t) &= \widetilde{A}_{m+1}(t; \widehat{k}_m, \widehat{\alpha}_m) \\ &= \int_{\mathcal{A}} A(t) Q_{\widehat{k}_m, \widehat{\alpha}_m}(dA|\underline{Z}_{m+1}) \\ &= \int_{[0,t]} \frac{\widehat{k}_m \widehat{\Phi}_m[s, \infty) d\widehat{\alpha}_m(s) + dN_{m+1}(s)}{\widehat{k}_m \widehat{\Phi}_m[s, \infty) + Y_{m+1}(s)}, \end{aligned} \tag{2.11}$$

where $\widehat{\Phi}_m(t) = \Phi_{\widehat{\alpha}_m}(t) = 1 - \prod_{[0,t]}\{1 - d\widehat{\alpha}_m(s)\}$.

We see from (2.7), that the Bayes risk of $\widehat{A}_{m+1}$ conditional on $(\underline{Z}_1, \cdots, \underline{Z}_m)$ is

$$\begin{aligned} &R_{m+1}(\widehat{A}_{m+1}, (k, \alpha)) \\ &= \int_{\mathcal{A}} \int_{H_{m+1}} L(\widehat{A}_{m+1}, A) P^{n_{m+1}}(d\underline{z}_{m+1}|A) Q_{k,\alpha}(dA), \end{aligned} \tag{2.12}$$

where the dependence on $(\underline{Z}_1, \cdots, \underline{Z}_m)$ is not displayed.

Let $\widetilde{P}^m_{k,\alpha}(\cdot)$ be the joint distribution of $(\underline{Z}_1, \cdots, \underline{Z}_m)$, i.e.,

$$\widetilde{P}^m_{k,\alpha}(\cdot) = P^{(n_1)}_{k,\alpha} \times \cdots \times P^{(n_m)}_{k,\alpha}(\cdot).$$

Then the overall Bayes risk of the empirical Bayes estimator $\widehat{A}_{m+1}(\cdot)$ of A_{m+1} is

$$\begin{aligned} &r^m_{m+1}(k, \alpha) \\ &= \int_{H_1 \times \cdots \times H_m} R_{m+1}(\widehat{A}_{m+1}, (k, \alpha)) d\widetilde{P}^m_{k,\alpha}(\underline{z}_1, \cdots, \underline{z}_m). \end{aligned} \tag{2.13}$$

We see that $r^m_{m+1}(k, \alpha) \geq r_{m+1}(k, \alpha)$. It is required that the empirical Bayes procedure $\{\widehat{A}_m\}$ satisfy the asymptotic optimality ;

$$\lim_{m \to \infty} \{r^m_{m+1}(k, \alpha) - r_{m+1}(k, \alpha)\} = 0 \tag{2.14}$$

holds for all (k, α).

The following three lemmas will be useful in proving the asymptotic optimality. The first lemma is a consequence of the L_2-orthogonality of $\mathrm{E}(A_{m+1}(t)|\underline{Z}_{m+1}) - A_{m+1}(t)$ and $\widehat{A}_{m+1}(t) - \mathrm{E}(A_{m+1}(t)|\underline{Z}_{m+1})$.

Lemma 2.1.

$$r^m_{m+1}(k, \alpha) = r_{m+1}(k, \alpha) + \int_0^\infty \mathrm{E}(\widehat{A}_{m+1}(t) - \widetilde{A}_{m+1}(t))^2 dw(t), \tag{2.15}$$

where E *is the expectation operation with respect to the joint distribution of* A_{m+1} *and* $(\underline{Z}_1, \cdots, \underline{Z}_{m+1})$.

Lemma 2.2. *For each* $t \geq 0$

$$P(\overline{Y}_m(t) > 0 \text{ for all sufficiently large } m) = 1 \tag{2.16}$$

and

$$P(\lim_{m \to \infty} \overline{Y}_m(t) = \infty) = 1. \tag{2.17}$$

Proof. For each $j \geq 1$ and $t \geq 0$, $P(Y_j(t) > 0) \geq P(Y_{11}(t) > 0) = \Phi_\alpha[t, \infty)G[t, \infty) > 0$ and therefore, $\sum_{j=1}^{\infty} P(Y_j(t) > 0) = \infty$. Since $Y_1(t), Y_2(t), \cdots$ are independent we have by the second Borel-Cantelli lemma $P(Y_j(t) > 0 \text{ i.o.}) = 1$, which completes the proof. □

Lemma 2.3. *Let $\widehat{c}_m(\cdot)$ be a bounded and mean-square consistent estimator of $c(\cdot)$. Then under* **(C1)**

$$\lim_{m \to \infty} \int_0^{\infty} \int_0^t \mathrm{E}(\widehat{c}_m(s) - c(s))^2 d\alpha(s) dw(t) = 0. \tag{2.18}$$

Proof. Let $D_m(t) = \int_0^t \mathrm{E}(\widehat{c}_m(s) - c(s))^2 d\alpha(s)$. Then $D_m(t)$ is dominated by a constant times $\alpha(t)$ and $D_m(t) \to 0$ as $m \to \infty$. Since **(C1)** is assumed, $\int_0^{\infty} D_m(s) dw(t) \to 0$ by dominated convergence theorem and the lemma is proved. □

3. RELATED MARTINGALES AND THEIR PROPERTIES

In this section we introduce some martingales and investigate their properties that will be used in estimating the prior distribution and proving the asymptotic optimality.

Let $\underline{Z}_1, \underline{Z}_2, \cdots$ be the sequence of data obtained by the independent repetitions of the component problem. Let N_{ji}, Y_{ji}, N_j, Y_j, $\overline{N}_m$, $\overline{Y}_m$ be the stochastic processes based on $\underline{Z}_1, \cdots, \underline{Z}_m$ which are given by (2.3), (2.4) and (2.5).

For $m \geq 1$, $1 \leq j \leq m$, $1 \leq i \leq n_j$ we define the following filtrations,

$$\begin{aligned} \mathcal{A}_{jt} &= \sigma\{A_j(u) : 0 \leq u \leq t\}, \\ \overline{\mathcal{A}}_{mt} &= \mathcal{A}_{1t} \otimes \cdots \otimes \mathcal{A}_{mt}, \end{aligned} \tag{3.1}$$

$$\begin{aligned} \mathcal{F}_{jit} &= \sigma\{N_{ji}(u), Y_{ji}(u) : 0 \leq u \leq t\}, \\ \mathcal{F}_{jt} &= \bigvee_{i=1}^{n_j} \mathcal{F}_{jit}, \\ \overline{\mathcal{F}}_{mt} &= \mathcal{F}_{1t} \otimes \cdots \otimes \mathcal{F}_{mt}, \end{aligned} \tag{3.2}$$

$$\begin{aligned} \mathcal{G}_{jit} &= \mathcal{F}_{jit} \otimes \mathcal{A}_{jt}, \\ \mathcal{G}_{jt} &= \mathcal{F}_{jt} \otimes \mathcal{A}_{jt}, \\ \overline{\mathcal{G}}_{mt} &= \overline{\mathcal{F}}_{mt} \otimes \overline{\mathcal{A}}_{mt}. \end{aligned} \tag{3.3}$$

We also define the processes

$$\begin{aligned} M_{ji}(t) &= N_{ji}(t) - \int_0^t Y_{ji} dA_j, \\ M_j(t) &= \sum_{i=1}^{n} M_{ji}(t) = N_j(t) - \int_0^t Y_j dA_j, \\ \overline{M}_m(t) &= \sum_{j=1}^{m} M_j(t) = \overline{N}_m(t) - \sum_{j=1}^{m} \int_0^t Y_j dA_j. \end{aligned} \tag{3.4}$$

By (1.2)-(1.4) in Definition 1.1 we see that

$$\mathrm{E} A_j(t) = \alpha(t) \tag{3.5}$$

and

$$\begin{aligned} \mathrm{Var}(A_j(t)) &= \int_0^t \frac{d\alpha(u)}{c(u)+1} \\ &= \int_0^t \frac{d\alpha(u)}{k\Phi_\alpha[u,\infty)+1}. \end{aligned} \tag{3.6}$$

It is well-known that the process M_{ji} given in (3.4) is a mean-zero , square-integrable $\{\mathcal{G}_{jit}\}$-martingale and M_{ji}^2 is a $\{\mathcal{G}_{jit}\}$-submartingale with the compensator

$$\langle M_{ji}, M_{ji}\rangle(t) = \int_0^t Y_{ji}(1-\Delta A_j) dA_j. \tag{3.7}$$

See Theorem 2.6.1 of Fleming and Harrington(1990).

Since $\mathrm{E}\Delta A_j(t) = \Delta\alpha(t) = 0$, $\Delta A_j(t) = 0$ a.s. Therefore (3.7) becomes

$$\langle M_{ji}, M_{ji}\rangle(t) = \int_0^t Y_{ji} dA_j \quad \text{a.s.} \tag{3.8}$$

Using that $\mathrm{E}(M_{ji}(t)|\mathcal{G}_{js}) = \mathrm{E}(M_{ji}(t)|\mathcal{G}_{jis})$ and $\mathrm{E}(M_j(t)|\overline{\mathcal{G}}_{ms}) = \mathrm{E}(M_j(t)|\mathcal{G}_{js})$ for $s \le t$ we have the following theorem.

Theorem 3.1. *For each $j \ge 1$, $m \ge 1$,*

(a) M_j is a mean-zero , square-integrable $\{\mathcal{G}_{jt}\}$-martingale,

(b) $\overline{M}_m$ is a mean-zero , square-integrable $\{\overline{\mathcal{G}}_{mt}\}$-martingale.

Theorem 3.2.

(a) If $s < t$, then $A_j(t) - A_j(s)$ is independent of $\mathcal{G}_{js} = \mathcal{F}_{js} \otimes \mathcal{A}_{js}$.

(b) $A_j(t) - \alpha(t)$ is a mean-zero, square-integrable $(\mathcal{G}_{jt})$-martingale.

Proof. (a) Let $D_1 \times D_2 \in \mathcal{F}_{js} \otimes \mathcal{A}_{js}$. We see that $\mathrm{P}(D_1|\mathcal{A}_{jt})$ is $\mathcal{A}_{js}$-measurable and therefore $\mathrm{P}(D_1 \times D_2|\mathcal{A}_{jt}) = 1_{D_2}\mathrm{P}(D_1|\mathcal{A}_{jt})$ is $\mathcal{A}_{js}$-measurable. This leads to

$$\begin{aligned} &\mathrm{P}(D_1 \times D_2|A_j(t) - A_j(s)) \\ &= \mathrm{E}(\mathrm{P}(D_1 \times D_2|\mathcal{A}_{jt})|A_j(t) - A_j(s)) \\ &= \mathrm{EP}(D_1 \times D_2|\mathcal{A}_{jt}) \\ &= \mathrm{P}(D_1 \times D_2), \end{aligned} \tag{3.9}$$

where the second equality holds because of the independent increments of A_j. Since $D_1 \times D_2 \in \mathcal{F}_{js} \otimes \mathcal{A}_{js}$ is arbitrary, (3.9) implies that $A_j(t) - A_j(s)$ is independent of $\mathcal{G}_{js} = \mathcal{F}_{js} \otimes \mathcal{A}_{js}$.

(b) Adaptedness and square-integrability of the Lévy process A_j are obvious. It suffices to show that for $0 \le s \le t$,

$$\mathrm{E}(A_j(t) - \alpha(t) - (A_j(s) - \alpha(s))|\mathcal{G}_{js}) = 0 \quad \text{a.s.} \tag{3.10}$$

But (3.10) follows from (a). □

Now we observe that for any distinct integers $i_1, i_2, \cdots \in \{1, 2, \cdots\}$, $m \ge 1$, $\{\Delta N_{1i_1}, \cdots, \Delta N_{m,i_m}\}$ are independent $\{0,1\}$-random variables. By Lemma 2.6.1 in Fleming and Harrington(1990), we see that for each $(j,i) \ne (l,k)$

$$\langle M_{ji}, M_{lk}\rangle(t) = 0 \quad \text{a.s.} \tag{3.11}$$

This together with (3.8) yields

Theorem 3.3.

(a) $\langle M_j, M_k\rangle(t) = \begin{cases} \int_0^t Y_j dA_j, & j = k \\ 0, & j \ne k \end{cases}$

(b) $\langle \overline{M}_m, \overline{M}_m\rangle(t) = \sum_{j=1}^m \int_0^t Y_j dA_j$.

For quadratic variation and covariation processes related to the Lévy processes $\{A_j\}_{j=1}^\infty$, we need the following two lemmas.

Lemma 3.1. *Let* $0 \le s < t$. *Then,*

(a) $\mathrm{E}(A_j(t)|\mathcal{G}_{js}) = \alpha(t) - \alpha(s) + A_j(s)$.

(b) $\mathrm{E}(A_j^2(t)|\mathcal{G}_{js}) = (\alpha(t) - \alpha(s) + A_j(s))^2 + \int_s^t \frac{d\alpha(u)}{c(u)+1}$.

(c) $\mathrm{E}(A_j(t)A_k(t)|\mathcal{G}_{js}\otimes\mathcal{G}_{ks}) = (\alpha(t)-\alpha(s)+A_j(s))(\alpha(t)-\alpha(s)+A_k(s))$ for $j \ne k$.

Proof. Part (a) is immediate from Theorem 3.2.(b). By Theorem 3.2.(a), $A_j(t) - A_j(s)$ is independent of $\mathcal{G}_{js}$ for $s < t$. Using (3.6) we see that

$$\begin{aligned}
&\mathrm{E}[(A_j(t) - A_j(s))^2|\mathcal{G}_{js}] \\
&= \mathrm{E}(A_j(t) - A_j(s))^2 \\
&= \mathrm{E}A_j^2(t) - 2\mathrm{E}(A_j(t) - A_j(s))\mathrm{E}A_j(s) - \mathrm{E}A_j^2(s) \\
&= (\alpha(t) - \alpha(s))^2 + \int_s^t \frac{d\alpha(u)}{c(u)+1}.
\end{aligned}$$

Then,

$$\begin{aligned}
&\mathrm{E}(A_j^2(t)|\mathcal{G}_{js}) \\
&= \mathrm{E}[(A_j(t) - A_j(s))^2|\mathcal{G}_{js}] + 2\mathrm{E}[(A_j(t) - A_j(s))A_j(s)|\mathcal{G}_{js}] + A_j^2(s) \\
&= \mathrm{E}(A_j(t) - A_j(s))^2 + 2A_j(s)\mathrm{E}(A_j(t) - A_j(s)) + A_j^2(s) \\
&= (\alpha(t) - \alpha(s) + A_j(s))^2 + \int_s^t \frac{d\alpha(u)}{c(u)+1},
\end{aligned}$$

which proves (b).

For $j \neq k$, $s < t$, $A_j(t) - A_j(s)$ is independent of $\mathcal{G}_{js} \otimes \mathcal{G}_{ks}$. Therefore, $(A_j(t) - A_j(s))(A_k(t) - A_k(s))$ is also independent of $\mathcal{G}_{js} \otimes \mathcal{G}_{ks}$ and we have

$$\begin{aligned}
&\mathrm{E}(A_j(t)A_k(t)|\mathcal{G}_{js} \otimes \mathcal{G}_{ks}) \\
&= \mathrm{E}(A_j(t) - A_j(s))\mathrm{E}(A_k(t) - A_k(s)) + A_k(s)\mathrm{E}(A_j(t)|\mathcal{G}_{js}) \\
&+ A_j(s)\mathrm{E}(A_k(t)|\mathcal{G}_{ks}) - A_j(s)A_k(s) \\
&= (\alpha(t) - \alpha(s) + A_j(s))(\alpha(t) - \alpha(s) + A_k(s)),
\end{aligned}$$

which proves (c). □

Applying Lemma 3.1 we have

Lemma 3.2.

(a) If $j \neq k$, $(A_j(t) - \alpha(t))(A_k(t) - \alpha(t))$ is a $(\mathcal{G}_{jt} \otimes \mathcal{G}_{kt})$-martingale.
(b) $(A_j(t) - \alpha(t))^2 - \int_0^t \frac{d\alpha(u)}{c(u)+1}$ is a $(\mathcal{G}_{jt})$-martingale.

Theorem 3.4.

$$\langle A_j - \alpha, A_k - \alpha\rangle(t) = \begin{cases} \int_0^t \frac{d\alpha(u)}{c(u)+1}, & j = k \\ 0, & j \neq k \end{cases}.$$

Proof. This is an immediate consequence of Lemma 3.2. □

4. ASYMPTOTIC OPTIMALITY

Recall from (2.11) the empirical Bayes estimator $\widehat{A}_{m+1}(\cdot)$ of $A_{m+1}(\cdot)$;

$$\widehat{A}_{m+1}(t) = \int_0^t \frac{\widehat{c}_m(s)d\widehat{\alpha}_m(s) + dN_{m+1}(s)}{\widehat{c}_m(s) + Y_{m+1}(s)}, \tag{4.1}$$

where $\widehat{c}_m(s) = \widehat{k}_m\widehat{\Phi}_m[s,\infty)$ is an estimator of $c(s) = k\Phi_\alpha[s,\infty)$ and $\widehat{\alpha}_m$ is an estimator of α based on censored data $\underline{Z}_j = (Z_{j1},\cdots,Z_{jn_j})$, $j = 1,\cdots,m$. Assume that the sample size sequence $\{n_j\}$ satisfies the condition**(C3)**. Our goal in this section is to show that the empirical Bayes procedures $\{\widehat{A}_m(\cdot)\}$ provided by a suitable class of estimated priors is asymptotically optimal, i.e., (2.14) holds for all $(c(\cdot),\alpha(\cdot))$ satisfying conditions **(C1)** and **(C2)**.

Theorem 4.1. (asymptotic optimality) *The empirical Bayes procedure* $\{\widehat{A}_m(\cdot)\}$ *given by (4.1) is asymptotically optimal if*

(i) $\{\widehat{k}_m\}$ *is a bounded sequence and* $\lim_{m\to\infty}\widehat{k}_m = k$, *a.s.*

(ii) $\lim_{m\to\infty}\widehat{\alpha}_m = \alpha$, *uniformly, a.s. on compact intervals,*

(iii) $\int_0^\infty \mathrm{E}(\widehat{\alpha}_m(t) - \alpha(t))^2 dw(t) < \infty$ *for each* $t \geq 0$.

Corollary 4.1. $\widehat{c}_m(\cdot) = \widehat{k}_m\widehat{\Phi}_m[\cdot,\infty)$ *is a bounded and mean-square consistent estimator of* $c(\cdot) = k\Phi_\alpha[\cdot,\infty)$.

Proof. From the condition (i) and bounded convergence theorem we see that

$$\mathrm{E}(\widehat{k}_m - k)^2 \to 0 \quad \text{as } m \to \infty.$$

By Duhamel equation(Gill and Johansen(1990)),

$$\begin{aligned}|\widehat{\Phi}_m(t) - \Phi_\alpha(t)| &= \Big|\prod_{[0,t]}(1 - d\widehat{\alpha}_m(s)) - \prod_{[0,t]}(1 - d\alpha(s))\Big| \\ &\leq |\widehat{\alpha}_m(t) - \alpha(t)|.\end{aligned}$$

This together with the condition (ii) implies

$$\mathrm{E}(\widehat{\Phi}_m[t,\infty) - \Phi_\alpha[t,\infty))^2 \to 0 \quad \text{as } m \to \infty.$$

Now, proof follows by observing that

$$\begin{aligned}\mathrm{E}(\widehat{c}_m(t) - c(t))^2 &\leq 2\mathrm{E}(\widehat{k}_m - k)^2\widehat{\Phi}_m[t,\infty) + 2\mathrm{E}k^2(\widehat{\Phi}_m[t,\infty) - \Phi_\alpha[t,\infty))^2 \\ &\leq 2\mathrm{E}(\widehat{k}_m - k)^2 + 2K^2\mathrm{E}(\widehat{\Phi}_m[t,\infty) - \Phi_\alpha[t,\infty))^2,\end{aligned}$$

where K is the constant in **(C2)**. □

Let $\widetilde{A}_{m+1}(\cdot)$ be the Bayes estimator of $A_{m+1}(\cdot)$ given by (2.9). By a simple calculation we have

$$\begin{aligned}
&\widehat{A}_{m+1}(t) - \widetilde{A}_{m+1}(t) \\
&= \int_0^t \psi_m(s) d(\widehat{\alpha}_m - \alpha)(s) + \int_0^t \varphi_m(s) dM_{m+1}(s) \\
&+ \int_0^t \varphi_m(s) Y_{m+1}(s) d(A_{m+1} - \alpha)(s),
\end{aligned} \tag{4.2}$$

where

$$\begin{aligned}
\varphi_m(s) &= \frac{1}{\widehat{c}_m(s) + Y_{m+1}(s)} - \frac{1}{c(s) + Y_{m+1}(s)} \\
&= \frac{c(s) - \widehat{c}_m(s)}{(\widehat{c}_m(s) + Y_{m+1}(s))(c(s) + Y_{m+1}(s))}, \\
\psi_m(s) &= \frac{\widehat{c}_m(s)}{\widehat{c}_m(s) + Y_{m+1}(s)}.
\end{aligned} \tag{4.3}$$

Let

$$\begin{aligned}
\mathrm{U}_m(t) &= \int_0^t \psi_m(s) d(\widehat{\alpha}_m - \alpha)(s), \\
\mathrm{V}_{m+1}(t) &= \int_0^t \varphi_m(s) dM_{m+1}(s), \\
\mathrm{W}_{m+1}(t) &= \int_0^t \varphi_m(s) Y_{m+1}(s) d(A_{m+1} - \alpha)(s).
\end{aligned} \tag{4.4}$$

Then, (4.2) is written as

$$\widehat{A}_{m+1}(t) - \widetilde{A}_{m+1}(t) = \mathrm{U}_m(t) + \mathrm{V}_{m+1}(t) + \mathrm{W}_{m+1}(t). \tag{4.5}$$

The following upper bounds of the second moments of the quantities in (4.4) will be useful in proving asymptotic optimality.

Lemma 4.1. *Let* $\mathrm{U}_m(\cdot)$, $\mathrm{V}_{m+1}(\cdot)$ *and* $\mathrm{W}_{m+1}(\cdot)$ *be given by (4.4). Then we have the following.*

(a) $\mathrm{E}(\mathrm{U}_m^2(t)) \leq \mathrm{E}(\widehat{\alpha}_m(t) - \alpha(t))^2$.

(b) $\mathrm{E}(\mathrm{V}_{m+1}^2(t)) \leq \int_0^t \mathrm{E}(\widehat{c}_m(s) - c(s))^2 d\alpha(s)$.

(c) $\mathrm{E}(\mathrm{W}_{m+1}^2(t)) \leq \int_0^t \mathrm{E}(\widehat{c}_m(s) - c(s))^2 d\alpha(s)$.

Proof. (a) Since $0 \leq \psi_m \leq 1$, by considering the positive and negative parts of the signed measure $\widehat{\alpha}_m - \alpha$ we have

$$\begin{aligned}
\mathrm{E}(\mathrm{U}_m^2(t)) &= \mathrm{E}\Big(\int_0^t \psi_m(s)d(\widehat{\alpha}_m - \alpha)(s)\Big)^2 \\
&= \mathrm{E}\Big(\int_0^t \psi_m(s)d(\widehat{\alpha}_m - \alpha)^+(s) - \int_0^t \psi_m(s)d(\widehat{\alpha}_m - \alpha)^-(s)\Big)^2 \\
&\leq \mathrm{E}\Big(\int_0^t \psi_m(s)d(\widehat{\alpha}_m - \alpha)^+(s) + \int_0^t \psi_m(s)d(\widehat{\alpha}_m - \alpha)^-(s)\Big)^2 \\
&\leq \mathrm{E}[(\widehat{\alpha}_m(t) - \alpha(t))^+ + (\widehat{\alpha}_m(t) - \alpha(t))^-]^2 \\
&= \mathrm{E}(\widehat{\alpha}_m(t) - \alpha(t))^2.
\end{aligned}$$

(b) Conditional on A_{m+1}, M_{m+1} is a mean-zero, square-integrable $\{\mathcal{F}_{m+1,t}\}$-martingale by Theorem 3.1.(a). Since $\varphi_m(\cdot)$ is $\{\mathcal{F}_{m+1,t}\}$-predictable, given $\underline{Z}_1, \cdots, \underline{Z}_m$ and bounded we see that $\mathrm{V}_{m+1}(\cdot)$ is a mean-zero, square-integrable $\{\mathcal{F}_{m+1,t}\}$-martingale, given $A_{m+1}, \underline{Z}_1, \cdots, \underline{Z}_m$. Therefore we have, by Theorem 3.3.(a),

$$\begin{aligned}
&\mathrm{E}(\mathrm{V}_{m+1}^2(t) | A_{m+1}, \underline{Z}_1, \cdots, \underline{Z}_m) \\
&= \mathrm{E}\Big(\int_0^t \varphi_m^2(s) d\langle M_{m+1}, M_{m+1}\rangle(s) \big| A_{m+1}, \underline{Z}_1, \cdots, \underline{Z}_m\Big) \\
&= \mathrm{E}\Big(\int_0^t \varphi_m^2(s) Y_{m+1}(s) dA_{m+1}(s) \big| A_{m+1}, \underline{Z}_1, \cdots, \underline{Z}_m\Big).
\end{aligned}$$

By taking expectation and using $Y_{m+1}^2(s) \geq Y_{m+1}(s)$, $s \geq 0$, we see that

$$\begin{aligned}
\mathrm{E}(\mathrm{V}_{m+1}^2(t)) &= \mathrm{E}\Big(\int_0^t \varphi_m^2(s) Y_{m+1}(s) dA_{m+1}(s)\Big) \\
&\leq \mathrm{E}\Big(\int_0^t (\varphi_m(s) Y_{m+1}(s))^2 dA_{m+1}(s)\Big) \\
&\leq \mathrm{E}\Big(\int_0^t (\widehat{c}_m(s) - c(s))^2 dA_{m+1}(s)\Big) \\
&= \int_0^t \mathrm{E}(\widehat{c}_m(s) - c(s))^2 d\alpha(s),
\end{aligned}$$

where the last equality holds since A_{m+1} is independent of $(\underline{Z}_1, \cdots, \underline{Z}_m)$.

(c) By Theorem 3.2.(b), $A_{m+1}(\cdot) - \alpha(\cdot)$ is a mean-zero, square-integrable $\{\mathcal{G}_{m+1,t}\}$-martingale. And $\varphi_m(\cdot)Y_{m+1}(\cdot)$ is $\{\mathcal{G}_{m+1,t}\}$-predictable, given $\underline{Z}_1, \cdots, \underline{Z}_m$ and bounded. Since $A_{m+1}(\cdot) - \alpha(\cdot)$ is independent of $\underline{Z}_1, \cdots, \underline{Z}_m$, we see that

$\mathrm{W}_{m+1}(\cdot)$ is a mean-zero, square-integrable $\{\mathcal{G}_{m+1,t}\}$-martingale, given $\underline{Z}_1, \cdots, \underline{Z}_m$. Therefore by Theorem 3.4,

$$\begin{aligned}
&\mathrm{E}(\mathrm{W}^2_{m+1}(t)|\underline{Z}_1, \cdots, \underline{Z}_m) \\
&= \mathrm{E}\Big(\int_0^t \varphi^2_m(s) Y^2_{m+1}(s) d\langle A_{m+1} - \alpha, A_{m+1} - \alpha\rangle(s)\big|\underline{Z}_1, \cdots, \underline{Z}_m\Big) \\
&= \mathrm{E}\Big(\int_0^t \frac{\varphi^2_m(s) Y^2_{m+1}(s)}{c(s)+1} d\alpha(s)\big|\underline{Z}_1, \cdots, \underline{Z}_m\Big).
\end{aligned}$$

By taking expectation,

$$\begin{aligned}
\mathrm{E}(\mathrm{W}^2_{m+1}(t)) &= \int_0^t \frac{\varphi^2_m(s) Y^2_{m+1}(s)}{c(s)+1} d\alpha(s) \\
&\le \int_0^t \mathrm{E}(\widehat{c}_m(s) - c(s))^2 d\alpha(s). \quad \square
\end{aligned}$$

Proof of Theorem 4.1. By Lemma 2.1, it suffices to show

$$\lim_{m\to\infty} \int_0^\infty \mathrm{E}(\widehat{A}_{m+1}(t) - \widetilde{A}_{m+1}(t))^2 dw(t) = 0. \tag{4.6}$$

We see that

$$\begin{aligned}
&\frac{1}{3}\int_0^\infty \mathrm{E}(\widehat{A}_{m+1}(t) - \widetilde{A}_{m+1}(t))^2 dw(t) \\
&\le \int_0^\infty \mathrm{E}(\mathrm{U}^2_m(t)) dw(t) + \int_0^\infty \mathrm{E}(\mathrm{V}^2_{m+1}(t)) dw(t) \\
&+ \int_0^\infty \mathrm{E}(\mathrm{W}^2_{m+1}(t)) dw(t) \\
&= \text{(I)} + \text{(II)} + \text{(III)},
\end{aligned} \tag{4.7}$$

where

(I) $\le \int_0^\infty \mathrm{E}(\widehat{\alpha}_m(t) - \alpha(t))^2 dw(t)$, by Lemma 4.1.(a),
(II) $\le \int_0^\infty \int_0^t \mathrm{E}(\widehat{c}_m(s) - c(s))^2 d\alpha(s) dw(t)$, by Lemma 4.1.(b),
(III) $\le \int_0^\infty \int_0^t \mathrm{E}(\widehat{c}_m(s) - c(s))^2 d\alpha(s) dw(t)$, by Lemma 4.1.(c).
Therefore,

$$\begin{aligned}
&\int_0^\infty \mathrm{E}(\widehat{A}_{m+1}(t) - \widetilde{A}_{m+1}(t))^2 dw(t) \\
&\le 3\int_0^\infty \mathrm{E}(\widehat{\alpha}_m(t) - \alpha(t))^2 dw(t) \\
&+ 6\int_0^\infty \int_0^t \mathrm{E}(\widehat{c}_m(s) - c(s))^2 d\alpha(s) dw(t).
\end{aligned} \tag{4.8}$$

By condition (ii) of the theorem $\lim_m \int_0^t (\widehat{\alpha}_m(s) - \alpha(s))^2 dw(s) = 0$ a.s. for $t \geq 0$. By Fatou's lemma and Fubini's theorem $\lim_m \int_0^t \mathrm{E}(\widehat{\alpha}_m(s) - \alpha(s))^2 dw(s) = 0$ for $t \geq 0$. Since $t \geq 0$ is arbitrary and condition (iii) is assumed the first term of rhs of (4.8) converges to zero as $m \to \infty$. The second term converges to zero by Corollary 4.1 and Lemma 2.3 and the proof is completed. □

5. CONCLUDING REMARKS

Our empirical Bayes problem can be transformed into an empirical Bayes nonparametric estimation problem of survival functions based on the right censored data with Dirichlet process $\mathcal{D}(k\Phi_\alpha)$. The quantity k which is well-interpreted as the prior sample size in Ferguson(1973) would be generally unknown. In this sense the work of Susarla and Van Ryzin(1978) is incomplete in that k is assumed known.

It would be necessary to work on constructing a sequence of estimators of the hyperparameter $(k, \alpha(\cdot))$ with which the empirical Bayes estimation procedure (4.1) satisfies asymptotic optimality. We treat this problem in a separate paper in which martingale technique is used in the estimation of $\alpha(\cdot)$ and a closed form of the marginal distribution is obtained for a maximum likelihood estimator of the parameter k.

In Susarla and Van Ryzin(1978), without the assumption that k is known the class of marginal distributions based on single sample size need not be identifiable. A class of marginal distributions is said to be identifiable if there is a one-to-one correspondence between the class of marginal distributions and the class of parameters of the prior. Without the identifiability condition, estimating the prior distribution for the empirical Bayes problem would be meaningless. By taking two or more observations at each component in our empirical Bayes problem(See condition(**C3**)) we see easily that from the closed form of the marginal distribution the class of marginal distributions is identifiable.

Asymptotic optimality of the empirical Bayes estimators of the survival functions $\{\widehat{F}_m\}$,

$$\widehat{F}_m(t) = 1 - \prod_{[0,t]} \{1 - d\widehat{A}_m(s)\}$$

follows easily from the asymptotic optimality of $\{\widehat{A}_m\}$ using the analytic properties of the product-integral operation(Gill and Johansen(1990)).

REFERENCES

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics* **1**, 209-230.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis,* John Wiley, New York.

Gill, R. D. and Johansen, S. (1990). "A survey of product-integration with a view toward application in survival analysis," *The Annals of Statistics* **18**, 1501-1555.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application,* Academic Press, New York.

Hjort, N. L. (1990). "Nonparametric Bayes estimators based on beta processes in models for life history data," *Annals of Statistics* **18**, 1259-1294.

Pollard, D. (1984). *Convergence of Stochastic Processes,* Springer Verlag, New York.

Robbins, H. E. (1955). "An empirical Bayes approach to statistics," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **1**, 157-164.

Susarla, V. and Van Ryzin, J. (1978). "Empirical Bayes estimation of a distribution (survival) function from right-censored observations," *The Annals of Statistics* **6**, 740-754.