Journal of the Korean Statistical Society (2004). 33: 2, pp 191–202

# EMPIRICAL BAYES ESTIMATION OF RESIDUAL SURVIVAL FUNCTION AT AGE $t$

TaChen Liang[1]


## Abstract

The paper considers nonparametric empirical Bayes estimation of residual survival function at age $t$ using a Dirichlet process prior $\mathcal{D}(\alpha)$. Empirical Bayes estimators are proposed for the case where both the function $\alpha(0, x]$ and the size $\alpha(R^+)$ are unknown. It is shown that the proposed empirical Bayes estimators are asymptotically optimal at a rate $n^{-1}$, where $n$ is the number of past data available for the present estimation problem. Therefore, the result of Lahiri and Park (1988) in which $\alpha(R^+)$ is assumed to be known and a rate $n^{-1}$ is achieved, is extended to $\alpha(R^+)$ unknown case.


*AMS 2000 subject classifications.* Primary 62G05.
*Keywords.* Asymptotically optimal, Dirichlet process, empirical Bayes, rate of convergence, regret, residual life.

## 1. Introduction

Consider a unit with a life time random variable $X$ and a distribution function $F$. Given that the unit is at age $t$, the remaining life of the unit after time $t$ is random. Study of the residual life is of importance in reliability. For a fixed age $t \geq 0$, the residual survival function at age $t$ is

$$R(x|t) = \frac{P(X > x + t)}{P(X > t)}. \tag{1.1}$$

Note that $R(x|t = 0) = 1 - F(x)$ is the survival function of a new unit at age 0. Our goal is to estimate $R(x|t)$. Let $\widetilde{R}(x|t)$ denote an estimator of $R(x|t)$. The following integrated squared error loss is applied:

$$L(\widetilde{R}, R) = \int_0^\infty \left\{\widetilde{R}(x|t) - R(x|t)\right\}^2 w(x)dx \tag{1.2}$$

Received October 2002; accepted February 2004.
[1] Department of Mathematics, Wayne State University, Detroit, MI 48202, U.S.A.

where $w(x)$ is a known, finite, positive integrable function on $(0, \infty)$ and $\int_0^\infty w(x)dx < \infty$.

Suppose that $F$ is a random distribution, distributed according to the Dirichlet process with parameter $\alpha$, denoted by $\mathcal{D}(\alpha)$. Here $\alpha$ is a finite non-null measure defined on the Borel $\sigma$-field $\mathcal{B}$ of the half line $(0, \infty)$. Let $\mathbf{X} = (X_1, \ldots, X_m)$ be a sample of size $m$ arising from the random distribution $F$. Using the nonparametric Bayesian approach of Ferguson (1973) and the integrated squared error loss (1.2), Lahiri and Park (1988) derive a nonparametric Bayes estimator of $R(x|t)$ as:

$$\varphi_\alpha(x) = D_\alpha(t, \mathbf{X})R_\alpha(x|t) + \{1 - D_\alpha(t, \mathbf{X})\}\widetilde{R}(x|t, \mathbf{X}), \tag{1.3}$$

where

$$R_\alpha(x|t) = \frac{\alpha(x+t, \infty)}{\alpha(t, \infty)}, \quad \widetilde{R}(x|t, \mathbf{X}) = \frac{\overline{F}_{\mathrm{em}}(t+x, \mathbf{X})}{\overline{F}_{\mathrm{em}}(t, \mathbf{X})},$$

$\overline{F}_{\mathrm{em}}(y, \mathbf{X}) = m^{-1}\sum_{j=1}^m I_{[y,\infty)}(X_j)$ is the empirical survival function, $\alpha(y, \infty)$ is the size of $\alpha$ over the interval $(y, \infty)$, and

$$D_\alpha(t, \mathbf{X}) = \frac{\alpha(t, \infty)}{\alpha(t, \infty) + m\overline{F}_{\mathrm{em}}(t, \mathbf{X})}.$$

The minimum Bayes risk is

$$r(\alpha, \varphi_\alpha) = E_{\mathbf{X}}\left[\int_0^\infty E_{\alpha|\mathbf{X}}\{\varphi_\alpha(x) - R(x|t)\}^2 w(x)dx\right]. \tag{1.4}$$

Note that $R_\alpha(x|t)$ is the nonparametric Bayes estimator of $R(x|t)$ without samples and $\widetilde{R}(x|t, \mathbf{X})$ is the empirical residual survival function at time $t$. Thus, the nonparametric Bayes estimator $\varphi_\alpha(x)$ of $R(x|t)$ based on the sample $\mathbf{X}$ is a weighted average of the two estimators $R_\alpha(x|t)$ and $\widetilde{R}(x|t, \mathbf{X})$.

When the parameter $\alpha(\cdot)$ is unknown, it is not possible to apply the nonparametric Bayes estimator $\varphi_\alpha$ for the concerned estimation problem. In such a situation and when a sequence of past data is available, Lahiri and Park (1988) study this estimation problem *via* the empirical Bays approach. When the size $\alpha(R^+)$ is known (but the measure $\alpha(\cdot)$ is in general unknown), Lahiri and Park (1988) study an empirical Bayes estimator for $R(x|t)$ based on equal sample sizes of past data. They prove that their proposed empirical Bayes estimator is asymptotically optimal at a rate $n^{-1}$, where $n$ is the number of past data at hand for the current estimation problem. When the size $\alpha(R^+)$ is unknown, they also propose an empirical Bayes estimator for $R(x|t)$ and prove that its corresponding regret converges to 0 as $n \to \infty$. However, the associated rate of convergence was not studied.

In the literature, the problem regarding the nonparametric empirical Bayes estimation using a Dirichlet process prior $\mathcal{D}(\alpha)$ has been extensively studied. We mentioned a few examples here. Lahiri and Park (1988, 1991) have studied nonparametric empirical Bayes estimation problems for $\alpha(R^+)$ unknown case. However, they only show that their proposed empirical Bayes estimators possess the asymptotic optimality; the associated rates of convergence are not studied. Susarla and Van Ryzin (1978) and Phadia (1980) have studied nonparametric empirical Bayes estimation of a distribution function based on right censored observations under the assumption that $\alpha(R)$ is known. When $\alpha(R)$ is unknown, Liang (2001) shows that his proposed nonparametric empirical Bayes estimators have a rate of convergence $n^{-1}$.

In this paper, it is assumed that both the measurable function $\alpha(\cdot)$ and the size $\alpha(R^+)$ are unknown. We consider the situation where the sample sizes at different stages may be unequal. We propose empirical Bayes estimators for the residual survival function $R(x|t)$ and investigate the associated asymptotic optimality. It is shown that under certain regularity conditions, the proposed empirical Bayes estimators are asymptotically optimal at a rate $n^{-1}$. Therefore, the result of Lahiri and Park (1988) in which $\alpha(R^+)$ is assumed to be known and a rate $n^{-1}$ is achieved, is extendet to $\alpha(R^+)$ unknown case.

## 2. EMPIRICAL BAYES ESTIMATION

In the empirical Bayes framework, let $(\mathbf{X}_i, F_i)$, $i = 1, 2, \ldots$ be a sequence of independent pairs of random elements, where for each $i = 1, 2, \ldots$, $\mathbf{X}_i = (X_{i1}, \ldots, X_{im_i})$ is a sample of size $m_i$ arising, at stage $i$, from a random distribution $F_i$, and $F_1, F_2, \ldots$ are *iid* random distributions, distributed according to the common Dirichlet process $\mathcal{D}(\alpha)$. At the present stage $n+1$, let $\mathbf{X}(n) = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ denote the $n$ past data and $\mathbf{X}_{n+1}$ stand for the present sample. We are interested in the estimation of the current residual survival function $R_{n+1}(x|t)$ at age $t$. Based on the present sample $\mathbf{X}_{n+1}$, according to (1.3), the nonparametric Bayes estimator of $R_{n+1}(x|t)$ is

$$\varphi_{n+1,\alpha}(x) = D_\alpha(t, \mathbf{X}_{n+1}) R_\alpha(x|t) + \left\{1 - D_\alpha(t, \mathbf{X}_{n+1})\right\} \widetilde{R}_{n+1}(x|t, \mathbf{X}_{n+1}). \quad (2.1)$$

where

$$D_\alpha(t, \mathbf{X}_{n+1}) = \frac{\alpha(t,\infty)}{\alpha(t,\infty) + m_{n+1}\overline{F}_{\text{em}}(t, \mathbf{X}_{n+1})},$$
$$\overline{F}_{\text{em}}(t, \mathbf{X}_{n+1}) = \frac{1}{m_{n+1}} \sum_{j=1}^{m_{n+1}} I_{[t,\infty)}(X_{n+1,j}). \tag{2.2}$$

The minimum Bayes risk for the $(n+1)^{th}$ component problem is

$$r(\alpha, \varphi_{n+1,\alpha}) = E_{\mathbf{X}_{n+1}} \left[ \int_0^\infty E_{\alpha|\mathbf{X}_{n+1}} \{\varphi_{n+1,\alpha}(x) - R_{n+1}(x|t)\}^2 w(x)dx \right]. \tag{2.3}$$

Note that the Bayes estimator $\varphi_{n+1,\alpha}(x)$ heavily depends on the Dirichlet parameter $\alpha$. Since $\alpha$ is unknown, it is not possible to apply $\varphi_{n+1,\alpha}$ for the present estimation problem. In such a situation, we may combine useful information from the past data $\mathbf{X}(n)$ and the present sample $\mathbf{X}_{n+1}$ to form a plausible estimator for $R_{n+1}(x|t)$. Such an estimator is called an empirical Bayes estimator and is denoted by $\varphi_{n+1,n}(x)$. Note that the data $\mathbf{X}_{n+1}$ and $\mathbf{X}(n)$ are implicitly contained in the subindices $n+1$ and $n$.

In order to construct empirical Bayes estimator for $R_{n+1}(x|t)$, we consider an alternative form of the Bayes estimator $\varphi_{n+1,\alpha}(x)$. From Zehnwirth (1977), it is known that

$$\alpha(R^+) = \frac{E_\alpha\{\text{Var}_F(X)\}}{\text{Var}_\alpha\{E_F(X)\}} \equiv \frac{E}{V}. \tag{2.4}$$

Let $\bar{p}(t) = \alpha(t,\infty)/\alpha(R^+)$. Then,

$$D_\alpha(t, \mathbf{X}_{n+1}) = \frac{\alpha(t,\infty)/\alpha(R^+)}{\alpha(t,\infty)/\alpha(R^+) + m_{n+1}\overline{F}_{\text{em}}(t, \mathbf{X}_{n+1})/\alpha(R^+)}$$
$$= \frac{E\bar{p}(t)}{E\bar{p}(t) + Vm_{n+1}\overline{F}_{\text{em}}(t, \mathbf{X}_{n+1})}, \tag{2.5}$$

$$R_\alpha(x,t) = \frac{\alpha(x+t,\infty)/\alpha(R^+)}{\alpha(t,\infty)/\alpha(R^+)} = \frac{\bar{p}(x+t)}{\bar{p}(t)}. \tag{2.6}$$

Thus, the Bayes estimator $\varphi_{n+1,\alpha}(x)$ can be written as

$$\varphi_{n+1,\alpha}(x) = \frac{E\bar{p}(t)}{E\bar{p}(t) + Vm_{n+1}\overline{F}_{\text{em}}(t, \mathbf{X}_{n+1})} \times \frac{\bar{p}(x+t)}{\bar{p}(t)}$$
$$+ \frac{Vm_{n+1}\overline{F}_{\text{em}}(t, \mathbf{X}_{n+1})}{E\bar{p}(t) + Vm_{n+1}\overline{F}_{\text{em}}(t, \mathbf{X}_{n+1})} \widetilde{R}_{n+1}(x|t, \mathbf{X}_{n+1}) \tag{2.7}$$
$$= D_\alpha(t, \mathbf{X}_{n+1}) \frac{\bar{p}(x+t)}{\bar{p}(t)} + \left\{1 - D_\alpha(t, \mathbf{X}_{n+1})\right\} \frac{\overline{F}_{\text{em}}(x+t, \mathbf{X}_{n+1})}{\overline{F}_{\text{em}}(t, \mathbf{X}_{n+1})}.$$

We need to estimate $E$, $V$ and $\bar{p}(y)$, $y > 0$. Let

$$\begin{aligned} \overline{X}_{i.} &= \frac{1}{m_i}\sum_{l=1}^{m_i} X_{il}, \\ \overline{X}_{..} &= \frac{\sum_{i=1}^n m_i \overline{X}_{i.}}{\sum_{i=1}^n m_i}, \\ \mathrm{MS_W} &= \frac{\sum_{i=1}^n \sum_{l=1}^{m_i} \left(X_{il} - \overline{X}_{i.}\right)^2}{\sum_{i=1}^n m_i - n}, \\ \mathrm{MS_B} &= \frac{\sum_{i=1}^n m_i \left(\overline{X}_{i.} - \overline{X}_{..}\right)^2}{n-1}. \end{aligned} \tag{2.8}$$

Straightforward computations yield that

$$\begin{aligned} E_{\mathbf{X}(n)}\left(\mathrm{MS_W}\right) &= E_\alpha\left\{\mathrm{Var}_F(X)\right\} \equiv E, \\ E_{\mathbf{X}(n)}\left(\mathrm{MS_B}\right) &= E_\alpha\left\{\mathrm{Var}_F(X)\right\} + \frac{g(m_1,\ldots,m_n)}{n-1}\mathrm{Var}_\alpha\left\{E_F(X)\right\} \\ &= E + \frac{g(m_1,\ldots,m_n)}{n-1}V, \end{aligned} \tag{2.9}$$

where

$$g(m_1,\ldots,m_n) = \sum_{i=1}^n m_i - \frac{\sum_{i=1}^n m_i^2}{\sum_{i=1}^n m_i}.$$

Therefore,

$$E_{\mathbf{X}(n)}\left\{\frac{n-1}{g(m_1,\ldots,m_n)}\left(\mathrm{MS_B} - \mathrm{MS_W}\right)\right\} = \mathrm{Var}_\alpha\left\{E_F(X)\right\} \equiv V. \tag{2.10}$$

We may use $\widetilde{E} \equiv \mathrm{MS_W}$ to estimate $E$. Since $V > 0$, while $\mathrm{MS_B} - \mathrm{MS_W}$ may be negative, thus, we use

$$\widetilde{V} = \frac{n-1}{g(m_1,\ldots,m_n)}\left(\mathrm{MS_B} - \mathrm{MS_W}\right)^+$$

to estimate $V$, where $y^+ = \max(0, y)$.

Note that $E_{\mathbf{X}(n)} I_{[y,\infty)}(X_{il}) = \bar{p}(y)$. Thus, $E_{\mathbf{X}(n)}\overline{F}_{\mathrm{em}}(y, \mathbf{X}_i) = \bar{p}(y)$. Define

$$\widetilde{\bar{p}}_n(y) = \frac{1}{n}\sum_{i=1}^n \overline{F}_{em}(y, \mathbf{X}_i), \quad \bar{p}_n^*(y) = \frac{\sum_{i=1}^n m_i \overline{F}_{\mathrm{em}}(y, \mathbf{X}_i)}{\sum_{i=1}^n m_i}. \tag{2.11}$$

We see that $E_{\mathbf{X}(n)}\{\widetilde{\overline{p}}_n(y)\} = \bar{p}(y)$ and $E_{\mathbf{X}(n)}\{\bar{p}^*(y)\} = \bar{p}(y)$. Let

$$\begin{aligned}\widetilde{D}_n(t, \mathbf{X}_{n+1}) &= \frac{\widetilde{E}\widetilde{\overline{p}}_n(t)}{\widetilde{E}\widetilde{\overline{p}}_n(t) + \widetilde{V}m_{n+1}\overline{F}_{\text{em}}(t, \mathbf{X}_{n+1})},\\ D_n^*(t, \mathbf{X}_{n+1}) &= \frac{\widetilde{E}\bar{p}_n^*(t)}{\widetilde{E}\bar{p}_n^*(t) + \widetilde{V}m_{n+1}\overline{F}_{\text{em}}(t, \mathbf{X}_{n+1})}.\end{aligned} \tag{2.12}$$

We now propose two empirical Bayes estimators $\widetilde{\varphi}_{n+1,n}$ and $\varphi^*_{n+1,n}$ as follows:

$$\begin{aligned}\widetilde{\varphi}_{n+1,n}(x) &= \widetilde{D}_n(t, \mathbf{X}_{n+1})\frac{\widetilde{\overline{p}}_n(x+t)}{\widetilde{\overline{p}}_n(t)}\\ &\quad+\left\{1 - \widetilde{D}_n(t, \mathbf{X}_{n+1})\right\}\widetilde{R}_{n+1}(x|t, \mathbf{X}_{n+1}),\\ \varphi^*_{n+1,n}(x) &= D_n^*(t, \mathbf{X}_{n+1})\frac{\bar{p}_n^*(x+t)}{\bar{p}_n^*(t)}\\ &\quad+\left\{1 - D_n^*(t, \mathbf{X}_{n+1})\right\}\widetilde{R}_{n+1}(x|t, \mathbf{X}_{n+1}).\end{aligned} \tag{2.13}$$

Note that all the values of $\widetilde{D}_n(t, \mathbf{X}_{n+1})$, $D_n^*(t, \mathbf{X}_{n+1})$, $\widetilde{\overline{p}}_n(x+t)/\widetilde{\overline{p}}_n(t)$, $\bar{p}_n^*(x+t)/\bar{p}_n^*(t)$, $\bar{p}(x+t)/\bar{p}(t)$ and $\widetilde{R}_{n+1}(x|t, \mathbf{X}_{n+1})$ are between (including) 0 and 1.

The Bayes risk of $\widetilde{\varphi}_{n+1,n}$ and $\varphi^*_{n+1,n}$ are, respectively, given as follows:

$$\begin{aligned} &r\left(\alpha, \widetilde{\varphi}_{n+1,n}\right)\\ &= E_{\mathbf{X}_{n+1}}\left[\int_0^\infty E_{\mathbf{X}(n)}E_{\alpha|\mathbf{X}_{n+1}}\left\{\widetilde{\varphi}_{n+1,n}(x) - R_{n+1}(x|t)\right\}^2 w(x)dx\right],\end{aligned} \tag{2.14}$$

$$\begin{aligned} &r(\alpha, \varphi^*_{n+1,n})\\ &= E_{\mathbf{X}_{n+1}}\left[\int_0^\infty E_{\mathbf{X}(n)}E_{\alpha|\mathbf{X}_{n+1}}\left\{\varphi^*_{n+1,n}(x) - R_{n+1}(x|t)\right\}^2 w(x)dx\right].\end{aligned} \tag{2.15}$$

## 3. ASYMPTOTIC OPTIMALITY

Let $\varphi_{n+1,n}$ be any empirical Bayes estimator of $R_{n+1}(x|t)$. Since $\varphi_{n+1,\alpha}(x)$ is the Bayes estimator of $R_{n+1}(x|t)$, $r(\alpha, \varphi_{n+1,n}) - r(\alpha, \varphi_{n+1,\alpha}) \geq 0$ for all $n \geq 1$. This nonnegative regret is used as a measure of performance of the empirical Bayes estimator $\varphi_{n+1,n}$. An empirical Bayes estimator $\varphi_{n+1,n}$ is said to be asymptotically optimal at a rate $\varepsilon_n$ relative to the Dirichlet process $\mathcal{D}(\alpha)$ if $r(\alpha, \varphi_{n+1,n}) - r(\alpha, \varphi_{n+1,\alpha}) = O(\varepsilon_n)$, where $\{\varepsilon_n\}$ is a sequence of positive, decreasing numbers such that $\lim_{n\to\infty}\varepsilon_n = 0$. In the following, we shall investigate the asymptotic optimality of the two empirical Bayes estimators $\widetilde{\varphi}_{n+1,n}$ and $\varphi^*_{n+1,n}$.

*3.1. Asymptotic optimality of $\widetilde{\varphi}_{n+1,n}$*

The empirical Bayes estimator $\widetilde{\varphi}_{n+1,n}$ has the following asymptotic optimality.

THEOREM 3.1. *Suppose that*

(a) *The Dirichlet process $\mathcal{D}(\alpha)$ satisfies that*

$$\alpha(R^+) < \infty \text{ and } \int_0^\infty x^4 d\alpha(0,x] < \infty,$$

(b) $2 \le m_i \le m$ *for all* $i = 1, 2, \ldots,$ *where the value of* $m$ *is independent of* $i$.

*Then under the prior distribution $\mathcal{D}(\alpha)$, $\widetilde{\varphi}_{n+1,n}$ is asymptotically optimal in the sense that*

$$r\left(\alpha, \widetilde{\varphi}_{n+1,n}\right) - r\left(\alpha, \varphi_{n+1,\alpha}\right) \le \frac{W_1}{n\{\bar{p}(t)\}^2}\left\{\frac{16Q_\alpha\left(n, \mathbf{m}_{n+1}, M_4\right)}{E^2} + 40\right\} = O(n^{-1}),$$

*where*

$$0 < W_1 = \int_0^\infty w(x)dx < \infty,$$

$$M_4 = E_{\mathbf{X}(n)}(X_{11}^4) = \int_0^\infty x^4 \frac{d\alpha(0,x]}{\alpha(R^+)} < \infty,$$

*and*

$$\begin{aligned} Q_\alpha(n, \mathbf{m}_{n+1}, M_4) &= \frac{32nM_4}{\sum_{i=1}^n (m_i - 1)} + 8E^2 \\ &\quad + \frac{6m_{n+1}^2(n-1)^2}{g^2(m_1, \ldots, m_n)}\left\{\frac{2n\sum_{i=1}^n m_i^2}{(n-1)^2} + \frac{4n}{\sum_{i=1}^n (m_i - 1)}\right\} M_4. \end{aligned}$$

PROOF. Note that $0 < W_1 = \int_0^\infty w(x)dx < \infty$ since $w(x)$ is a positive integrable function on $(0, \infty)$. By assumption (a) of the theorem, $M_4 < \infty$.

From (2.3) and (2.14), the regret of $\widetilde{\varphi}_{n+1,n}$ can be written as

$$r\left(\alpha, \widetilde{\varphi}_{n+1,n}\right) - r(\alpha, \varphi_{n+1,\alpha}) = \int_0^\infty E_{\mathbf{X}(n+1)}\{\widetilde{\varphi}_{n+1,n}(x) - \varphi_{n+1,\alpha}(x)\}^2 w(x)dx. \tag{3.1}$$

From (2.7) and (2.13),

$$\begin{aligned} &\widetilde{\varphi}_{n+1,n}(x) - \varphi_{n+1,\alpha}(x) \\ &= \left\{\widetilde{D}_n(t, \mathbf{X}_{n+1}) - D_\alpha(t, \mathbf{X}_{n+1})\right\}\left\{\frac{\widetilde{\bar{p}}_n(x+t)}{\widetilde{\bar{p}}_n(t)} - \frac{\overline{F}_{\mathrm{em}}(x+t, \mathbf{X}_{n+1})}{\overline{F}_{\mathrm{em}}(t, \mathbf{X}_{n+1})}\right\} \\ &\quad + D_\alpha(t, \mathbf{X}_{n+1})\left\{\frac{\widetilde{\bar{p}}_n(x+t)}{\widetilde{\bar{p}}_n(t)} - \frac{\bar{p}(x+t)}{\bar{p}(t)}\right\}. \end{aligned}$$

Since

$$0 \le \widetilde{D}_n(t, \mathbf{X}_{n+1}),\ D_\alpha(t, \mathbf{X}_{n+1}),\ \frac{\tilde{\bar{p}}_n(x+t)}{\tilde{\bar{p}}_n(t)},\ \frac{\overline{F}_{\text{em}}(x+t, \mathbf{X}_{n+1})}{\overline{F}_{\text{em}}(t, \mathbf{X}_{n+1})} \le 1,$$

thus,

$$|\widetilde{\varphi}_{n+1,n}(x) - \varphi_{n+1,\alpha}(x)| \le \left|\widetilde{D}_n(t, \mathbf{X}_{n+1}) - D_\alpha(t, \mathbf{X}_{n+1})\right| + \left|\frac{\tilde{\bar{p}}_n(x+t)}{\tilde{\bar{p}}_n(t)} - \frac{\bar{p}(x+t)}{\bar{p}(t)}\right|.$$

Therefore,

$$\begin{aligned} &\{\widetilde{\varphi}_{n+1,n}(x) - \varphi_{n+1,\alpha}(x)\}^2 \\ &\le 2\left\{\widetilde{D}_n(t, \mathbf{X}_{n+1}) - D_\alpha(t, \mathbf{X}_{n+1})\right\}^2 + 2\left\{\frac{\tilde{\bar{p}}_n(x+t)}{\tilde{\bar{p}}_n(t)} - \frac{\bar{p}(x+t)}{\bar{p}(t)}\right\}^2 \end{aligned} \tag{3.2}$$

Substituting (3.2) into (3.1), we obtain

$$\begin{aligned} &r(\alpha, \widetilde{\varphi}_{n+1,n}) - r(\alpha, \varphi_{n+1,\alpha}) \\ &\le 2\int_0^\infty E_{\mathbf{X}(n+1)}\{\widetilde{D}_n(t, \mathbf{X}_{n+1}) - D_\alpha(t, \mathbf{X}_{n+1})\}^2 w(x)dx \\ &\quad +2\int_0^\infty E_{\mathbf{X}(n+1)}\left\{\frac{\tilde{\bar{p}}_n(t+x)}{\tilde{\bar{p}}_n(t)} - \frac{\bar{p}(t+x)}{\bar{p}(t)}\right\}^2 w(x)dx. \end{aligned} \tag{3.3}$$

Note that

$$0 \le \frac{\tilde{\bar{p}}_n(t+x)}{\tilde{\bar{p}}_n(t)},\ \frac{\bar{p}(t+x)}{\bar{p}(t)} \le 1.$$

Then by Singh's Lemma (see Lemma 4.1) and Lemma 4.2(a),

$$\begin{aligned} &E_{\mathbf{X}(n+1)}\left\{\frac{\tilde{\bar{p}}_n(t+x)}{\tilde{\bar{p}}_n(t)} - \frac{\bar{p}(t+x)}{\bar{p}(t)}\right\}^2 \\ &\le \frac{8}{\{\bar{p}(t)\}^2}\left[E_{\mathbf{X}(n+1)}\left\{\tilde{\bar{p}}_n(t+x) - \bar{p}(t+x)\right\}^2 + \frac{3}{2}E_{\mathbf{X}(n+1)}\left\{\tilde{\bar{p}}_n(t) - \bar{p}(t)\right\}^2\right] \\ &\le \frac{8}{\{\bar{p}(t)\}^2}\left(\frac{1}{n} + \frac{3}{2n}\right) = \frac{20}{n\{\bar{p}(t)\}^2}. \end{aligned} \tag{3.4}$$

Note that the upper bound given at the RHS of (3.4) is independent of $x$ for all $x > 0$.

Also, by Singh's Lemma again. Lemma 4.3 and the fact that

$$E\bar{p}(t) + Vm_{n+1}\overline{F}_{\text{em}}(t,\mathbf{X}_{n+1}) \geq E\bar{p}(t) > 0,$$

we have

$$
\begin{aligned}
&E_{\mathbf{X}(n+1)}\Big\{\widetilde{D}_n(t,\mathbf{X}_{n+1}) - D_\alpha(t,\mathbf{X}_{n+1})\Big\}^2 \\
&= E_{\mathbf{X}_{n+1}}\Bigg[E_{\mathbf{X}(n)}\Bigg\{\Bigg(\frac{\widetilde{E}\widetilde{\bar{p}}_n(t)}{\widetilde{E}\widetilde{\bar{p}}_n(t) + \widetilde{V}m_{n+1}\overline{F}_{\text{em}}(t,\mathbf{X}_{n+1})} \\
&\qquad - \frac{E\bar{p}(t)}{E\bar{p}(t) + Vm_{n+1}\overline{F}_{\text{em}}(t,\mathbf{X}_{n+1})}\Bigg)^2 \Bigg| \mathbf{X}_{n+1}\Bigg\}\Bigg] \\
&\leq E_{\mathbf{X}_{n+1}}\Bigg[\frac{8}{\{E\bar{p}(t)\}^2}E_{\mathbf{X}(n)}\Big\{\widetilde{E}\widetilde{\bar{p}}_n(t) - E\bar{p}(t)\Big\}^2\Bigg] \\
&\quad + E_{\mathbf{X}_{n+1}}\Bigg[\frac{8}{\{E\bar{p}(t)\}^2} \times \frac{3}{2}E_{\mathbf{X}(n)}\Big\{\widetilde{E}\widetilde{\bar{p}}_n(t) + \widetilde{V}m_{n+1}\overline{F}_{\text{em}}(t,\mathbf{X}_{n+1}) \\
&\qquad - E\bar{p}(t) - Vm_{n+1}\overline{F}_{\text{em}}(t,\mathbf{X}_{n+1})\Big\}^2\Bigg] \\
&\leq \frac{8}{\{E\bar{p}(t)\}^2}\Bigg[4E_{\mathbf{X}(n+1)}\Big\{\widetilde{E}\widetilde{\bar{p}}_n(t) - E\bar{p}(t)\Big\}^2 \\
&\qquad + 3E_{\mathbf{X}(n+1)}\Big\{(\widetilde{V} - V)m_{n+1}\overline{F}_{\text{em}}(t,\mathbf{X}_{n+1})\Big\}^2\Bigg] \\
&\leq \frac{8}{\{E\bar{p}(t)\}^2}\Bigg[8E_{\mathbf{X}(n+1)}\big(\widetilde{E} - E\big)^2 + 8E^2E_{\mathbf{X}(n+1)}\Big\{\widetilde{\bar{p}}_n(t) - \bar{p}(t)\Big\}^2 \\
&\qquad + 3m_{n+1}^2E_{\mathbf{X}(n+1)}\big(\widetilde{V} - V\big)^2\Bigg] \\
&\leq \frac{8}{\{E\bar{p}(t)\}^2}\Bigg[\frac{32M_4}{\sum_{i=1}^n(m_i - 1)} + \frac{8E^2}{n} \\
&\qquad + \frac{6(n-1)^2m_{n+1}^2}{g^2(m_1,\ldots,m_n)}\Bigg\{\frac{2\sum_{i=1}^n m_i^2}{(n-1)^2} + \frac{4}{\sum_{i=1}^n(m_i - 1)}\Bigg\}M_4\Bigg] \\
&= \frac{8}{n\{E\bar{p}(t)\}^2}Q_\alpha(n,\mathbf{m}_{n+1},M_4).
\end{aligned} \tag{3.5}
$$

Note that the upper bound given at the RHS of (3.5) is independent of $x$ for all

$x > 0$. Combining (3.3)-(3.5) yields

$$\begin{aligned} & r(\alpha, \varphi_{n+1,n}) - r(\alpha, \varphi_{n+1,\alpha}) \\ & \leq 2 \int_0^\infty \frac{8}{n \{E\bar{p}(t)\}^2} Q_\alpha(n, \mathbf{m}_{n+1}, M_4) w(x) dx + 2 \int_0^\infty \frac{20}{n \{\bar{p}(t)\}^2} w(x) dx \\ & = \frac{1}{n [E\bar{p}(t)]^2} \left\{16 Q_\alpha(n, \mathbf{m}_{n+1}, M_4) + 40 E^2\right\} W_1 = O(n^{-1}), \end{aligned} \tag{3.6}$$

where the last equality of (3.6) is guaranteed under the assumption (b) of the theorem. □

### *3.2. Asymptotic optimality of* $\varphi^*_{n+1,n}$

The empirical Bayes estimator $\varphi^*_{n+1,n}$ has the following asymptotic optimality.

THEOREM 3.2. *Suppose that*

(a) *The Dirichlet process* $\mathcal{D}(\alpha)$ *satisfies that*

$$\alpha(R^+) < \infty \text{ and } \int_0^\infty x^4 d\alpha(0, x] < \infty,$$

(b) $2 \leq m_i \leq m$ *for all* $i$, *where the value of* $m$ *is independent of* $i$.

*Then,* $\varphi^*_{n+1,n}$ *is asymptotically optimal in the sense that*

$$\begin{aligned} r\left(\alpha, \varphi^*_{n+1,n}\right) - r\left(\alpha, \varphi_{n+1,\alpha}\right) & \leq \frac{1}{n \{\bar{p}(t)\}^2} \left\{ \frac{16 Q_\alpha(n, \mathbf{m}_{n+1}, M_4)}{E^2} + 40 \right\} W_1 \\ & = O(n^{-1}). \end{aligned}$$

The proof of Theorem 3.2 is analogous to that of Theorem 3.1. Thus, the detail is omitted.

## 4. AUXILIARY RESULTS

The following lemmas are helpful for presenting a concise proof for Theorems 3.1 and 3.2. First, we introduce a lemma from Singh (1977).

LEMMA 4.1 (SINGH'S LEMMA). *Let* $Y$ *and* $Z$ *be random variables, and* $y$ *and* $z$ *be real values with* $z > 0$, *and* $L$ *be a finite positive value. Then,*

$$E\left(\left|\frac{Y}{Z} - \frac{y}{z}\right| \wedge L\right)^2 \leq \frac{8}{z^2} \left\{ E(Y - y)^2 + \left( \left|\frac{y}{z}\right|^2 + \frac{L^2}{2} \right) E(Z - z)^2 \right\}.$$

LEMMA 4.2.

(a) $E_{\mathbf{X}(n)}\left\{\tilde{\bar{p}}_n(y)\right\} = \bar{p}(y)$ *and* $\mathrm{Var}_{\mathbf{X}(n)}\left\{\tilde{\bar{p}}_n(y)\right\} \le n^{-1}$.

(b) $E_{\mathbf{X}(n)}\{\bar{p}_n^*(y)\} = \bar{p}(y)$ *and* $\mathrm{Var}_{\mathbf{X}(n)}\{\bar{p}_n^*(y)\} \le n^{-1}$.

PROOF. (a) From (2.11),

$$\begin{aligned} E_{\mathbf{X}(n)}\left\{\tilde{\bar{p}}_n(y)\right\} &= \frac{1}{n}\sum_{j=1}^{n} E_{\mathbf{X}(n)}\left\{\overline{F}_{\mathrm{em}}(y,\mathbf{X}_j)\right\} = \frac{1}{n}\sum_{j=1}^{n}\bar{p}(y) = \bar{p}(y), \\ \mathrm{Var}_{\mathbf{X}(n)}\left\{\tilde{\bar{p}}_n(y)\right\} &= \mathrm{Var}_{\mathbf{X}(n)}\left\{\frac{1}{n}\sum_{j=1}^{n}\overline{F}_{\mathrm{em}}(y,\mathbf{X}_j)\right\} \\ &= \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{Var}_{\mathbf{X}(n)}\left\{\overline{F}_{\mathrm{em}}(y,\mathbf{X}_j)\right\} \le n^{-1}. \end{aligned}$$

(b) Since

$$\bar{p}_n^*(y) = \frac{1}{\sum_{i=1}^{n} m_i}\sum_{j=1}^{n} m_j\overline{F}_{\mathrm{em}}(y,\mathbf{X}_j),$$

thus,

$$\begin{aligned} E_{\mathbf{X}(n)}\{\bar{p}_n^*(y)\} &= \frac{1}{\sum_{i=1}^{n} m_i}\sum_{j=1}^{n} m_j E_{\mathbf{X}(n)}\left\{\overline{F}_{\mathrm{em}}(y,\mathbf{X}_j)\right\} \\ &= \frac{1}{\sum_{j=1}^{n} m_i}\sum_{j=1}^{n} m_j\bar{p}(y) = \bar{p}(y), \\ \mathrm{Var}_{\mathbf{X}(n)}\{\bar{p}_n^*(y)\} &= \frac{1}{(\sum_{i=1}^{n} m_i)^2}\sum_{j=1}^{n} m_j^2\,\mathrm{Var}_{\mathbf{X}(n)}\left\{\overline{F}_{\mathrm{em}}(y,\mathbf{X}_j)\right\} \\ &\le \frac{1}{(\sum_{i=1}^{n} m_i)^2}\sum_{j=1}^{n} m_j^2 \times m_j^{-1} = \frac{1}{\sum_{i=1}^{n} m_i} \le n^{-1}. \end{aligned}$$

□

The following lemma is from Liang (2001).

LEMMA 4.3. *Let* $V = \mathrm{Var}_\alpha\{E_F(X)\}$ *and* $E = E_\alpha\{\mathrm{Var}_F(X)\}$. *Then,*

(a) $E_{\mathbf{X}(n)}(\mathrm{MS_W} - E)^2 \le \dfrac{4}{\sum_{i=1}^{n}(m_i-1)}E_{\mathbf{X}(n)}\left(X_{11}^4\right)$.

(b) $E_{\mathbf{X}(n)}\left\{\mathrm{MS_B} - E - \dfrac{g(m_1,\ldots,m_n)}{n-1}V\right\}^2 \le \dfrac{2\sum_{i=1}^{n} m_i^2}{(n-1)^2}E_{\mathbf{X}(n)}\left(X_{11}^4\right)$.

(c) $E_{\mathbf{X}(n)}\left(\widetilde{V}-V\right)^2$

$$= E_{\mathbf{X}(n)}\left\{\frac{n-1}{g(m_1,\ldots,m_n)}(\mathrm{MS_B}-\mathrm{MS_W})^+ - V\right\}^2$$

$$\le \frac{2(n-1)^2}{g^2(m_1,\ldots,m_n)}\left[E_{\mathbf{X}(n)}\left\{\mathrm{MS_B}-E-\frac{g(m_1,\ldots,m_n)}{n-1}V\right\}^2 \right.$$

$$\left. + E_{\mathbf{X}(n)}(\mathrm{MS_W}-E)^2\right]$$

$$\le \frac{2(n-1)^2}{g^2(m_1,\ldots,m_n)}\left\{\frac{2\sum_{i=1}^n m_i^2}{(n-1)^2}+\frac{4}{\sum_{i=1}^n(m_i-1)}\right\}E_{\mathbf{X}(n)}\left(X_{11}^4\right).$$

## REFERENCES

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

LAHIRI, P. AND PARK, D. H. (1988). "Nonparametric Bayes and empirical Bayes estimation of the residual survival function at age $t$", *Communications in Statistics-Theory and Methods*, **17**, 4085–4098.

LAHIRI, P. AND PARK, D. H. (1991). "Nonparametric Bayes and empirical Bayes estimators of mean residual life at age $t$", *Journal of Statistical Planning and Inference*, **29**, 125–136.

LIANG, T. C. (2001). "Rate of convergence for empirical Bayes estimation of a distribution function", In *Advances on Theoretical and Methodological Aspects of Probability and Statistics* (N. Balakrishnan, ed.), 331–344, Gordon and Breach Science Publishers, Amsterdam.

PHADIA, E. G. (1980). "A note on empirical Bayes estimation of a distribution function based on censored data", *The Annals of Statistics*, **8**, 226–229.

SINGH, R. S. (1977). "Applications of estimators of a density and its derivatives to certain statistical problems", *Journal of the Royal Statistical Society*, **B39**, 357–363.

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.

ZEHNWIRTH, B. (1977). "The mean credibility formula in a Bayes rule", *Scandinavian Acturial Journal*, **4**, 212–216.