ON CONSISTENCY OF SOME NONPARAMETRIC BAYES ESTIMATORS WITH RESPECT TO A BETA PROCESS BASED ON INCOMPLETE DATA

JEE CHANG HONG AND INHA JUNG

ABSTRACT. Let F and G denote the distribution functions of the failure times and the censoring variables in a random censorship model. Susarla and Van Ryzin(1978) verified consistency of \hat{F}_{α} , the NPBE of F with respect to the Dirichlet process prior $D(\alpha)$, in which they assumed F and G are continuous. Assuming that A, the cumulative hazard function, is distributed according to a beta process with parameters c, α , Hjort(1990) obtained the Bayes estimator $\hat{A}_{c,\alpha}$ of A under a squared error loss function. By the theory of product-integral developed by Gill and Johansen(1990), the Bayes estimator $\hat{F}_{c,\alpha}$ is recovered from $\hat{A}_{c,\alpha}$. Continuity assumption on F and G is removed in our proof of the consistency of $\hat{A}_{c,\alpha}$ and $\hat{F}_{c,\alpha}$. Our result extends Susarla and Van Ryzin(1978) since a particular transform of a beta process is a Dirichlet process and the class of beta processes forms a much larger class than the class of Dirichlet processes.

1. Introduction and Summary.

Let X_1, \dots, X_n be independent and identically distributed (i.i.d.) random variables from a distribution F on $[0, \infty)$ having F(0) = 0 and let C_1, \dots, C_n be i.i.d. random variables with cumulative distribution function (cdf) G on $[0, \infty)$. Assume that the X_i are independent of the C_i . Let $T_i = \min\{X_i, C_i\}$, $\delta_i = 1_{\{X_i \leq C_i\}}$ for each $i = 1, \dots, n$, and let H be the cdf of the i.i.d. random variables T_i, T_2, \dots . Then 1 - H = (1 - F)(1 - G). In the usual random censorship model one observes only $(T_1, \delta_1), \dots, (T_n, \delta_n)$.

The problem of constructing nonparametric Bayes estimators (NPBE) for F based on the censored data $(T_1, \delta_1), \dots, (T_n, \delta_n)$ has been considered by many authors by

Received by the editors Sep. 20, 1998 and, in revised form Oct. 22, 1998.

¹⁹⁹¹ Mathematics Subject Classification. 62G20

Key words and phrases. Nonparametric Bayes Estimators, Lévy processes, censored data, cumulative hazard function, martingales, product-integrals.

This research was supported in part by the Daewoo Foundation in 1997.

placing a prior distribution for F on the space \mathcal{F} of all cdf's on $[0, \infty)$. Using the Dirichlet process introduced by Ferguson(1973), an NPBE for F based on the censored data has been considered by Susarla and Van Ryzin(1978). Ferguson and Phadia(1979) obtained an NPBE for F with respect to the prior process neutral to the right introduced by Doksum(1974).

Definition 1.1. (Ferguson and Phadia(1979)) A process F(t) is said to be a random distribution function neutral to the right if it can be written in the form $F(t) = 1 - e^{-B(t)}$ where $B(\cdot)$ is a Lévy process with independent increments such that (a) $B(\cdot)$ is nondecreasing a.s., (b) $B(\cdot)$ is right continuous a.s., (c) $\lim_{t\to-\infty} B(t) = 0$ a.s., $\lim_{t\to\infty} B(t) = \infty$ a.s.

Ferguson and Phadia(1979) extends the result of Susarla and Van Ryzin(1978) in that a Dirichlet process is a process neutral to the right.

Given a cdf F on $[0,\infty)$, the cumulative hazard function(chf) A is defined by

$$A(t) = \int_{[0,t]} \frac{dF(s)}{F[s,\infty)}, \quad t \ge 0.$$
 (1.1)

The formula (1.1) yields

$$F(t) = \int_{[0,t]} F[s,\infty) dA(s) \tag{1.2}$$

which is well known as the Volterra integral equation. The unique solution of F determined by A in equation (1.2) is given in terms of the product-integral by

$$F(t) = 1 - \prod_{[0,t]} (1 - dA), \quad t \ge 0.$$
 (1.3)

See Gill and Johansen (1990).

For an investigation of the survival phenomena chf A is an object as basic as the survival function F. Let \mathcal{A} be the space of all chf's. Hjort(1990) introduced a beta process for A with parameter functions $c(\cdot)$ and $\alpha(\cdot)$ denoted by $A \sim Be(c, \alpha)$, where $c(\cdot)$ is a piecewise continuous and nonnegative function on $[0, \infty)$ and $\alpha(\cdot)$ is a chf. A beta process is an \mathcal{A} -valued Lévy process with independent increments. (See the definition of a beta process in Hjort(1990).)

The NPBE $\hat{A}_{c,\alpha}$ of A with respect to the beta process $A \sim Be(c,\alpha)$ based on $(T_1, \delta_1), \dots, (T_n, \delta_n)$ obtained by Hjort(1990) is given by

$$\hat{A}_{c,\alpha}(t) = \int_{[0,t]} \frac{cd\alpha + dN}{c + Y},\tag{1.4}$$

where

$$N(t) = \sum_{i=1}^{n} 1_{\{T_i \le t, \delta_i = 1\}},$$

$$Y(t) = \sum_{i=1}^{n} 1_{\{T_i \ge t\}}.$$
(1.5)

Viewing definition 1.1, if A is a beta process, then the random distribution F given by (1.3) is a process neutral to the right. By a substitution of (1.4) into the right hand side of (1.3) we obtain an estimator $\hat{F}_{c,\alpha}$ of F given by

$$\hat{F}_{c,\alpha}(t) = 1 - \prod_{[0,t]} (1 - d\hat{A}_{c,\alpha}), \quad t \ge 0.$$
(1.6)

Using the fact that the posterior of a beta process given data is also a beta process and a beta process has independent increments, one can easily see that $\hat{F}_{c,\alpha}$ is a conditional expectation of F given data. Therefore we see that the estimator $\hat{F}_{c,\alpha}$ given in (1.6) is an NPBE of F with respect to a process neutral to the right under a squared error loss function.

Let (Ω, \mathcal{F}, P) be the underlying probability space for this model and take filtration as

$$\mathcal{F}_t = \sigma\{1_{\{T_i < s, \delta_i = 1\}}, \ 1_{\{T_i > s\}} : 0 \le s \le t, \ i = 1, \dots, n\}, \ t \ge 0. \tag{1.7}$$

Now, $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$ is the stochastic basis for this model. Thus, the estimators $\hat{A}_{c,\alpha}$ and $\hat{F}_{c,\alpha}$ can be written as the conditional expectations

$$\hat{A}_{c,\alpha}(t) = E(A(t)|\mathcal{F}_s, s \ge 0)$$

$$\hat{F}_{c,\alpha}(t) = E(F(t)|\mathcal{F}_s, s \ge 0). \tag{1.8}$$

Our goal is verifying consistency of the NPBE's $\hat{A}_{c,\alpha}$ and $\hat{F}_{c,\alpha}$ in the frequentist's view by assuming that

(A1) X_1, \dots, X_n are i.i.d. random variables with a fixed unknown distribution F_0 on $[0, \infty)$,

(A2) C_1, \dots, C_n are i.i.d. random variables with an unknown distribution G on $[0, \infty)$.

On this assumption the random variable $T_i = X_i \wedge C_i$ have cdf H_0 given by

$$1 - H_0 = (1 - F_0)(1 - G). (1.9)$$

Let A_0 be the chf corresponding to F_0 , i.e.,

$$A_0(t) = \int_{[0,t]} \frac{dF_0(s)}{F_0[s,\infty)}, \quad t \ge 0.$$
 (1.10)

Consider the process M on $[0, \infty)$ given by

$$M(t) = N(t) - \int_0^t Y dA_0,$$
 (1.11)

where A_0 is given by (1.10) and the processes N and Y are given by (1.5). It is well-known that the process M is a square-integrable zero mean martingale with respect to the filtration in (1.7) and it has the predictable variation process $\langle M, M \rangle$ given by

$$\langle M, M \rangle(t) = \int_0^t Y(1 - \Delta A_0) dA_0. \tag{1.12}$$

This is the unique, nondecreasing, predictable process such that $M^2 - \langle M, M \rangle$ is again a martingale.

Susarla and Van Ryzin(1978) verified consistency of \hat{F}_{α} , the NPBE of F with respect to the Dirichlet process prior $D(\alpha)$, where $\alpha(\cdot)$ is a finite non-null measure representing the parameter of the Dirichlet process, in which they assumed F and G are continuous. Continuity assumption on F and G is removed in our consistency proof. Our result extends Susarla and Van Ryzin(1978) since a particular transform of a beta process is a Dirichlet process and the class of beta processes forms a much larger class than the class of Dirichlet processes. (See Hjort(1990).) Section 2 treats the mean square consistency using some martingale techniques. Section 3 treats almost sure consistency.

2. Mean square consistency

Assume the conditions (A1) and (A2) given in section 1. Let $\hat{A}_{c,\alpha}$ and $\hat{F}_{c,\alpha}$ be given by (1.4) and (1.6), respectively. In this section we prove mean square consistency of $\hat{A}_{c,\alpha}$ and $\hat{F}_{c,\alpha}$ by assuming further that the parameter function $c(\cdot)$ of the beta process $Be\{c,\alpha\}$ is bounded by a positive constant K>0 so that

(A3)
$$0 \le c(t) \le K, \ t \ge 0.$$

The function $c(\cdot)$ plays the role of prior number at risk and the assumption (A3) seems not unreasonable. Using (1.11), $\hat{A}_{c,\alpha}$ can be written as

$$\hat{A}_{c,\alpha}(t) = \int_0^t \frac{dM}{c+Y} + \int_0^t \frac{c}{c+Y} d\alpha + \int_0^t \frac{Y}{c+Y} dA_0. \tag{2.1}$$

Since $(c+Y)^{-1}$ is bounded and predictable, the first term in the right hand side defines a square-integrable, zero mean martingale with predictable variation process

$$\left\langle \int \frac{dM}{c+Y}, \int \frac{dM}{c+Y} \right\rangle = \int_0^t \left(\frac{1}{c+Y}\right)^2 d\langle M, M \rangle$$
$$= \int_0^t \left(\frac{1}{c+Y}\right)^2 Y(1 - \Delta A_0) dA_0. \tag{2.2}$$

Thus, we have

$$E\left(\int_0^t \frac{dM}{c+Y}\right) = 0, (2.3)$$

$$E\left(\int_0^t \frac{dM}{c+Y}\right)^2 = \int_0^t E\frac{Y}{(c+Y)^2} (1 - \Delta A_0) dA_0.$$
 (2.4)

Since $\hat{A}_{c,\alpha}(t)$ is a conditional expectation given data, it can be easily seen that

$$E(\hat{A}_{c,\alpha}(t) - E\hat{A}_{c,\alpha}(t))(E\hat{A}_{c,\alpha}(t) - A_0(t)) = 0$$
(2.5)

and from which together with (2.1), (2.3) and (2.4) we see that

$$E(\hat{A}_{c,\alpha}(t) - A_{0}(t))^{2} = E(\hat{A}_{c,\alpha}(t) - E\hat{A}_{c,\alpha}(t))^{2} + (E\hat{A}_{c,\alpha}(t) - A_{0}(t))^{2}$$

$$= E\left\{\int_{0}^{t} \frac{dM}{c+Y} - \int_{0}^{t} \left(\frac{c}{c+Y} - E\frac{c}{c+Y}\right) dA_{0} + \int_{0}^{t} \left(\frac{c}{c+Y} - E\frac{c}{c+Y}\right) d\alpha\right\}^{2} + (E\hat{A}_{c,\alpha}(t) - A_{0}(t))^{2}$$

$$\leq 3E\left[\int_{0}^{t} \frac{dM}{c+Y}\right]^{2} + 3E\left[\int_{0}^{t} \left(\frac{c}{c+Y} - E\frac{c}{c+Y}\right) dA_{0}\right]^{2}$$

$$+ 3E\left[\int_{0}^{t} \left(\frac{c}{c+Y} - E\frac{c}{c+Y}\right) d\alpha\right]^{2} + (E\hat{A}_{c,\alpha}(t) - A_{0}(t))^{2}. \tag{2.6}$$

The last inequality in (2.6) uses the fact that $(a+b+c)^2 \le 3(a^2+b^2+c^2)$ for all real numbers a, b, c.

Lemma 2.1. Assume that conditions (A1), (A2) and (A3) hold. Then for each $t \geq 0$ we have

$$E\left[\frac{c(t)}{c(t)+Y(t)}\right] \le \frac{1}{n} \frac{1+K}{1-H_0(t-)},\tag{2.7}$$

$$E\left[\frac{Y(t)}{(c(t)+Y(t))^2}\right] \le \frac{2}{n} \frac{1}{1-H_0(t-)},\tag{2.8}$$

$$E\left[\frac{c(t)}{c(t)+Y(t)}\right]^{2} \le \frac{2}{n^{2}} \left(\frac{1+K}{1-H_{0}(t-)}\right)^{2}.$$
 (2.9)

Proof. The following computations will be based on the fact that Y(t) is a binomial random variable with parameters n and $1 - H_0(t-)$, where H_0 is the cdf of the random variables $T_i = X_i \wedge C_i$, $i = 1, \dots, n$.

$$E\left[\frac{c(t)}{c(t)+Y(t)}\right] = \sum_{j=0}^{n} \binom{n}{j} \frac{c(t)}{c(t)+j} (1-H_0(t-))^j H_0(t-)^{n-j}$$

$$\leq \sum_{j=0}^{n} \binom{n}{j} \frac{c(t)+1}{j+1} (1-H_0(t-))^j H_0(t-)^{n-j}$$

$$= \frac{c(t)+1}{(1-H_0(t-))(n+1)} \sum_{j=0}^{n} \binom{n+1}{j+1} (1-H_0(t-))^{j+1} H_0(t-)^{n+1-(j+1)}$$

$$= \frac{1}{n+1} \frac{c(t)+1}{1-H_0(t-)} \sum_{j=1}^{n+1} \binom{n+1}{j} (1-H_0(t-))^j H_0(t-)^{n+1-j}$$

$$\leq \frac{1}{n} \frac{1+K}{1-H_0(t-)},$$

which proves (2.7). Similarly (2.8), (2.9) can proved easily.

Lemma 2.2. Assume that conditions (A1), (A2) and (A3) hold. Then, for each $t \geq 0$ we have

$$E\left[\int_0^t \frac{dM}{c+Y}\right]^2 \le \frac{2}{n} \frac{A_0(t)}{1 - H_0(t-)},\tag{2.10}$$

$$E\left[\int_0^t \left(\frac{c}{c+Y} - E\frac{c}{c+Y}\right) dA_0\right]^2 \le \frac{2}{n^2} \left\{\frac{(1+K)A_0(t)}{1 - H_0(t-)}\right\}^2,\tag{2.11}$$

$$E\left[\int_{0}^{t} \left(\frac{c}{c+Y} - E\frac{c}{c+Y}\right) d\alpha\right]^{2} \le \frac{2}{n^{2}} \left\{\frac{(1+K)\alpha(t)}{1-H_{0}(t-)}\right\}^{2}, \tag{2.12}$$

$$[E\hat{A}_{c,\alpha}(t) - A_0(t)]^2 \le \frac{2}{n^2} \left(\frac{1+K}{1-H_0(t-)}\right)^2 [\alpha^2(t) + A_0^2(t)]. \tag{2.13}$$

Proof. Since $0 \le \Delta A_0 \le 1$, it follows from (2.4) and (2.8) that

$$E\left[\int_{0}^{t} \frac{dM}{c+Y}\right]^{2} \leq \int_{0}^{t} E \frac{Y}{(c+Y)^{2}} dA_{0}$$

$$\leq \frac{2}{n} \int_{0}^{t} \frac{1}{1-H_{0}(s-)} dA_{0}$$

$$\leq \frac{2}{n} \frac{A_{0}(t)}{1-H_{0}(t-)},$$

which proves (2.10). From (2.11) observe that

$$\left\{ \int_0^t \left(\frac{c}{c+Y} - E \frac{c}{c+Y} \right) dA_0 \right\}^2 \le A_0(t) \int_0^t \left(\frac{c}{c+Y} - E \frac{c}{c+Y} \right)^2 dA_0 \quad (2.14)$$

holds by the Hölder inequality. Then by (2.9) and (2.14) we have

$$E\left[\int_{0}^{t} \left(\frac{c}{c+Y} - E\frac{c}{c+Y}\right) dA_{0}\right]^{2} \leq A_{0}(t) \int_{0}^{t} var\left(\frac{c}{c+Y}\right) dA_{0}$$

$$\leq A_{0}(t) \int_{0}^{t} E\left(\frac{c}{c+Y}\right)^{2} dA_{0}$$

$$\leq A_{0}(t) \frac{2}{n^{2}} (1+K)^{2} \int_{0}^{t} \frac{1}{(1-H_{0}(s-))^{2}} dA_{0}$$

$$\leq \frac{2}{n^{2}} \left\{\frac{(1+K)A_{0}(t)}{1-H_{0}(t-)}\right\}^{2},$$

which proves (2.11). (2.12) can be proved similarly. Finally, using (2.7) we see that

$$\begin{split} [E\hat{A}_{c,\alpha}(t) - A_0(t)]^2 &= \left[\int_0^t E \frac{c}{c+Y} d(\alpha - A_0) \right]^2 \\ &\leq 2 \left[\int_0^t E \frac{c}{c+Y} d\alpha \right]^2 + 2 \left[\int_0^t E \frac{c}{c+Y} dA_0 \right]^2 \\ &\leq 2 \left[\int_0^t \frac{1}{n} \frac{1+K}{1-H_0(s-)} d\alpha \right]^2 + 2 \left[\int_0^t \frac{1}{n} \frac{1+K}{1-H_0(s-)} dA_0 \right]^2 \\ &\leq \frac{2}{n^2} \left(\frac{1+K}{1-H_0(t-)} \right)^2 [\alpha^2(t) + A_0^2(t)], \end{split}$$

which proves (2.13).

Applying lemma 2.2 to the right hand side of (2.6) we have

Theorem 2.3. Assume that conditions (A1), (A2) and (A3) hold. Then, for each $t \geq 0$

$$E[\hat{A}_{c,\alpha}(t) - A_0(t)]^2 \le \frac{\gamma_t}{n},\tag{2.15}$$

where γ_t is a constant depending on t.

Corollary 2.4. Assume that (A1), (A2) and (A3) hold. Then for each $t \geq 0$

$$E[\hat{F}_{c,\alpha}(t) - F_0(t)]^2 \longrightarrow 0 \quad as \ n \to \infty.$$
 (2.16)

Proof. Since $\hat{F}_{c,\alpha}(t)$ and $F_0(t)$ are cdf's, $\{(\hat{F}_{c,\alpha}(t) - F_0(t))^2\}$ forms a uniformly integrable sequence of random variables. Since $\hat{F}_{c,\alpha}(t)$ and $F_0(t)$ are defined by the product-integrals of $\hat{A}_{c,\alpha}(t)$ and $A_0(t)$ and the product-integral operator is a continuous mapping, it follows from theorem 2.3 that

$$\hat{F}_{c,\alpha}(t) \xrightarrow{P} F_0(t) \text{ as } n \to \infty.$$
 (2.17)

Uniform integrability of $\{(\hat{F}_{c,\alpha}(t) - F_0(t))^2\}$ together with (2.17) implies (2.16).

3. Almost sure consistency.

Let $A \sim Be\{c, \alpha\}$ be the beta process and let $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$ be the stochastic basis as given in section 1. Consider the sequence of increasing σ -fields,

$$\mathcal{F}_n = \sigma\{1_{\{T_i < t, \delta_i = 1\}}, \ 1_{\{T_i > t\}} : t \ge 0, \ i = 1, \dots, n\}, \ n = 1, 2, \dots$$
(3.1)

Then the NPBE $\hat{A}_{c,\alpha}$ given in (1.4) and (1.8) can be rewritten as $E(A(t)|\mathcal{F}_n)$. In order to clarify the dependence in the sample size n of the NPBE $\hat{A}_{c,\alpha}$ we write as

$$\hat{A}_n(t) = \hat{A}_{c,\alpha}(t) = E(A(t)|\mathcal{F}_n). \tag{3.2}$$

Lemma 3.1. Assume that conditions (A1), (A2) and (A3) hold. Then for each fixed $t \geq 0$, sequence of random variables $\{\hat{A}_n(t) : n = 1, 2, \dots\}$ is a uniformly integrable martingale with respect to the σ -fields $\{\mathcal{F}_n : n = 1, 2, \dots\}$.

Proof. It is obvious that $\hat{A}_n(t)$ is integrable and \mathcal{F}_n -measurable for each $n = 1, 2, \cdots$. Since

$$E(\hat{A}_{n+1}(t)|\mathcal{F}_n) = E[E(A(t)|\mathcal{F}_{n+1})|\mathcal{F}_n] = E(A(t)|\mathcal{F}_n) = \hat{A}_n(t)$$

we conclude that $\{\hat{A}_n(t): n=1,2,\cdots\}$ is a martingale with respect to the σ -fields $\{\mathcal{F}_n: n=1,2,\cdots\}$. Uniform integrability of $\{\hat{A}_n(t): n=1,2,\cdots\}$ follows from observing

$$E|\hat{A}_n(t) - A_0(t)| \le (E[\hat{A}_n(t) - A_0(t)]^2)^{1/2} \to 0 \text{ as } n \to \infty.$$

(See theorem 10.3.6 of Dudley(1989).)

Theorem 3.2. Assume that conditions (A1), (A2) and (A3) hold. Let $\{\hat{A}_n(t): n=1,2,\cdots\}$ be the $\{\mathcal{F}_n\}$ -martingale given by (3.1) and (3.2). Then for each fixed $t \geq 0$, $\{\hat{A}_n(t)\}$ converges almost surely to $A_0(t)$.

Proof. Since $\{\hat{A}_n(t)\}$ is uniformly integrable, $\sup_n E|\hat{A}_n(t)| < \infty$. (See theorem 7.5.4 of Ash(1972).) By the martingale convergence theorem, $\{\hat{A}_n(t)\}$ converges almost surely to a random variable, say $A_\infty(t)$, as $n \to \infty$. Since $\{\hat{A}_n(t)\}$ converges in probability to $A_0(t)$, $\{\hat{A}_n(t)\}$ converges almost surely to $A_0(t)$ through a subsequence. Therefore the limits $A_\infty(t)$ and $A_0(t)$ must coincide almost surely.

Let $\hat{F}_{c,\alpha}$ be the NPBE given by (1.6) and (1.8). Paralleling with (3.2) write

$$\hat{F}_n(t) = \hat{F}_{c,\alpha}(t) = E(F(t)|\mathcal{F}_n), \tag{3.3}$$

where \mathcal{F}_n is the σ -field given by (3.1). Since $\hat{F}_n(t)$ and $F_0(t)$ are the continuous images of $\hat{A}_n(t)$ and $A_0(t)$ under the product-integral operator, we have the following corollary to theorem 3.2.

Corollary 3.3. Assume that conditions (A1), (A2) and (A3) hold. Then for each fixed $t \geq 0$, $\{\hat{F}_n(t)\}$ converges almost surely to $F_0(t)$ as $n \to \infty$.

REFERENCES

- 1. Ash, R. B., Real analysis and probability, Academic press, New York, 1989.
- Doksum, K., Tailfree and neutral random probabilities and their posterior distributions, Ann. Probab. 2 (1974), 183-201.
- 3. Dudley, R. M., Real analysis and probability, Wadsworth, California., 1989.
- Ferguson, T. S., A Bayesian analysis of some nonparametric problems, Ann. Statist. 1 (1973), 209–230.
- 5. Ferguson, T. S. and Phadia, E. G., Bayesian nonparametric estimation based on censored data, Ann. Statist. 7 (1979), 163-186.
- Gill, R. D. and Johansen, S., A survey of product-integration with a view toward application in survival analysis, Ann. Statist. 18 (1990), 1501-1555.
- Hjort, N. L., Nonparametric Bayes estimators based on beta processes in models for life history data, Ann. Statist. 18 (1990), 1259–1294.
- 8. Susarla, V. and Van Ryzin, J., Large sample theory for a Bayesian nonparametric survival curve estimator based on censored samples, Ann. Statist. 6 (1978), 755-768.

DEPARTMENT OF MATHEMATICS, AJOU UNIVERSITY, SUWON, 442-749, KOREA.

DIVISION OF INFORMATION AND COMPUTER ENGINEERING, AJOU UNIVERSITY, SUWON, 442-749, KOREA.