Ann. Inst. Statist. Math.
35 (1983), Part A, 389-400

NONPARAMETRIC BAYESIAN ESTIMATION OF A SURVIVAL CURVE WITH DEPENDENT CENSORING MECHANISM

E. G. PHADIA* AND V. SUSARLA**

(Received Oct. 19, 1981; revised Apr. 12, 1983)

Summary

Let $(X_1, Y_1), \cdots, (X_N, Y_N)$ be a random sample from a bivariate distribution function F. Based on observing only (δ_i, Z_i) where $\delta_i=1$ if $X_i \leqq Y_i$ and $=0$ otherwise and $Z_i=\min\{X_i, Y_i\}$ for $i=1, \cdots, n$, we obtain the Bayes estimator of F when F is a Dirichlet process under the usual integrated squared error loss function. It should be pointed out here that X_i and Y_i need not be independent which is the usual assumption in survival analysis models. The effect of this dependence can be seen clearly in the estimators obtained and also in the given example which illustrates the estimator when Freunds' bivariate exponential distribution is taken as the parameter of the Dirichlet process.

1. Introduction

The problem of estimating a distribution function F_1 (or the survival function $\bar{F}_1=1-F_1$) has been considered by several authors (for example, see the references in Miller [9]) when a random sample $X_1, \cdots, X_N$ from F_1 has been randomly right censored by the random sample $Y_1, \cdots, Y_n$ respectively. Most of the papers dealt with estimation of $\bar{F}_1$ under the assumption that $(X_1, \cdots, X_n)$ and $(Y_1, \cdots, Y_n)$ are independent of each other. Since this kind of censoring models occur mainly in situations involving life time data, we assume from now on that $\bar{F}_1(0)=P(X_i>0)=1$ and that $P(Y_i>0)=1$ where, here and elsewhere, i runs from 1 through N. There has been a lot of development in the estimation of $\bar{F}_1$ and its functionals like mean when one gets to observe only the minimum of X_i and Y_i and whether $X_i \leqq Y_i$ or $X_i > Y_i$. So let

* The research of this author is supported in part by an NSF Grant MCS80-03244.
** The research of this author is supported by the NIH Grant NO: 1 R01 GM 28405.

(1.1) $\delta_i=1$ if $X_i \leqq Y_i$ and $=0$ if $X_i > Y_i$, and $Z_i = X_i \wedge Y_i$.

However in most of the papers the results have been derived under the assumption

(A1) X_i and Y_i are independent of each other.

Some papers where this has been relaxed are Lagakos and Williams [7], and Langeberg, Proschan and Quinzi [8]. In this paper, we also do not assume (A1). The first paper does not assume a bivariate structure for their model but rather require assumptions about the marginal distributions of the survival times, the conditional probability of observing failure given actual survival time, and the odds of observing a failure given the observed portion of survival time. The second paper recommends converting first the dependent model to an independent component model, provided certain conditions are satisfied, and then uses the usual estimation procedures. These approaches are shown to be useful in certain applications. However, neither of these papers discuss the Bayesian approach. We present here this alternative approach which is useful in certain circumstances where a bivariate structure may be appropriate but the functional form is unknown.

Here, we take into consideration the dependence of X_i and Y_i through a bivariate model in a Bayesian context which is described in the next section. In this Bayes set up, we obtain the Bayes estimator of $\bar{F}(s,t)=P(X_i>s, Y_i>t)$ for $s, t \in (0, \infty)$ and by specializing this to $t=0$, the Bayes estimator of $\bar{F}_1(s)=P(X_i>s)$ can be obtained. This is done in Section 3 which also gives needed lemmas.

Section 2 provides some comments and important special cases such as Ferguson's [2] Bayes estimator of a distribution function. Section 4 deals with examples employing the Bayes estimator obtained in Section 2 along with its numerical application to the data reported in Kaplan and Meier [6] and compares the resulting estimator with that reported in Susarla and Van Ryzin [10]. Section 5 provides comments regarding the large sample behaviour of the Bayes estimator in the i.i.d. set up that is, when $(X_1, Y_1), \cdots, (X_N, Y_N)$ are i.i.d. with $P(X_1>s, Y_1>t)=\bar{F}_0(s,t)$.

As an example where our Bayes estimator might be applicable, consider a situation in which there are two causes of death labelled 1 and 2. With each cause of death r, associate a random variable T_r representing the time to death if r were the only cause of death. Then, in practice, one gets to observe only $\min(T_1, T_2)$ and whether cause of death is 1 or 2. For our model, we need to assume that the cause of death is either 1 or 2 but not both simultaneously. Also, live withdrawal from the study could be considered as a cause of death.

2. The Bayes estimator of $\bar{F}$

In this section we describe the Bayesian approach, introduce some notation and state the main result.

Let P be a Dirichlet process on $R_2^+=\{(0,\infty)\times(0,\infty)\}$ with parameter α, where α is a nonnegative finite measure on $(R_2^+, \mathcal{B}_2^+)$ and $\mathcal{B}_2^+$ is the Borel field defined on R_2^+. Let $(X, Y)\sim P$, and F be the bivariate distribution function corresponding to P. Further let $(X_1, Y_1), \cdots, (X_N, Y_N)$ be a radom sample from F, i.e. given F, (X_i, Y_i) are independently and identically distributed according to F. Based on the observable random variables $(\boldsymbol{\delta}, \boldsymbol{Z})=\{(\delta_i, Z_i)\}_{i=1}^N$ defined in Section 1, we would like to estimate $\bar{F}$. Since the Bayes estimator of $\bar{F}$ under integrated squared error loss function, $\int_{R_2^+}(\bar{F}-\hat{\bar{F}})^2 dW$, where W is a weight function, is the posterior mean, we give here an explicit expression for $\mathrm{E}\,[\bar{F}(s,t)|(\boldsymbol{\delta}, \boldsymbol{Z})]$.

Suppose that among N observations, we have n distinct z_i's and without loss of generality assume them to be ordered so that $0<z_1<z_2<\cdots<z_n<\infty$ and let $\delta_1, \cdots, \delta_n$ correspond to these distinct z_i's. Further, let λ_i and μ_i be the number of censored and uncensored observations at z_i, i.e. $\lambda_i=\#\{j|X_j\wedge Y_j=z_i \text{ and } X_j>Y_j\}$, $\mu_i=\#\{j|X_j\wedge Y_j=z_i \text{ and } X_j\leqq Y_j\}$. Then the following theorem holds.

THEOREM 2.1. *The Bayes estimator under squared error loss is given by*

$$\hat{\bar{F}}_\alpha(s,t)=\mathrm{E}\,[(\bar{F}(s,t)|(\boldsymbol{\delta}, \boldsymbol{Z})]$$

$$=\frac{1}{\alpha(R_2^+)+N}\left\{\alpha((s,\infty)\times(t,\infty))+N^+(\max(s,t))+\sum_r\theta_r\right\} \tag{2.1}$$

where the summation ranges over all r such that $\min(s,t)<z_r<\max(s,t)$,

$$N^+(u)=\sum_{i\ni z_i>u}(\lambda_i+\mu_i)=\#\ \textit{of observations}>u\,, \tag{2.2}$$

$$\theta_r=\begin{cases}\lambda_r\alpha_s'(Z_r) & \text{if } s>t\\ \mu_r\alpha_t'(Z_r) & \text{if } s<t\\ 0 & \text{if } s=t\,,\end{cases} \tag{2.3}$$

and

$$\begin{aligned}\alpha_s'(Z_r)&=\lim_{\varepsilon\downarrow 0}\frac{\alpha(\{X>s,\ Z_r-\varepsilon<Y\leqq Z_r\})}{\alpha(\{X>Y,\ Z_r-\varepsilon<Y\leqq Z_r\})}\,,\\ \alpha_t'(Z_r)&=\lim_{\varepsilon\downarrow 0}\frac{\alpha(\{Y>t,\ Z_r-\varepsilon<X\leqq Z_r\})}{\alpha(\{X\leqq Y,\ Z_r-\varepsilon<X\leqq Z_r\})}\end{aligned} \tag{2.4}$$

whenever the limits exist.

Remarks.

1. It can be checked that the Bayes estimator $\hat{\bar{F}}_\alpha$ is a proper distribution function.
2. It is obvious that the formula (2.1) will hold for R_2 (instead of R_2^+) also.
3. The numerator in the Bayes formula (2.1) represents the α-measure of the set $(s, \infty)\times(t, \infty)$, plus the number of observations greater than $\max(s, t)$, plus a quantity which may be considered as a sum of "conditional" probabilities each weighted by the number of ties at the point of conditioning.
4. If we take $s=t$, then the 'conditional' density part vanishes and the expression reduces to an analogue in 2-dimension of the Bayes estimator obtained by Ferguson [2]. In this case we can express the estimator as a convex combination of the prior guess and the empirical distribution function in two dimensions. This gives the Bayes estimator of the probability that the life time of the component is at least s units of time.
5. If we set $t=0$ in the formula (2.1), we get the Bayes estimator of the marginal $F(s, 0)$ $(=\bar{F}_1(s)$ in the notation of Section 1).
6. If α is taken to be such that $\alpha(A\times B)=\alpha_1(A)\alpha_2(B)$ where α_1 and α_2 are two measures on $(R_1, \mathscr{B}_1)$, $A, B\in\mathscr{B}_1$, and the σ-field in R_2 is the σ-field generated by the rectangles $\{A\times B: A, B\in\mathscr{B}_1\}$ then Ferguson's Bayes estimator [2] can be obtained as a special case by taking all the mass under α_2 to be at ∞. This means that Y_j's are degenerate at ∞ so that there is no censoring with probability one.
7. It should be noted that these results can be generalized to the Bayes dependent competing risk models where there are three or more competing (dependent) causes of failure and we observe only the life time of the component and the cause of failure.

Example. We take α to be continuous with the following density which is a special case of Freund's bivariate exponetial distribution (Johnson and Kotz [5], p. 265).

$$\alpha(x, y)=\begin{cases}\beta(\beta+\gamma)e^{-(\beta+\gamma)y} & (0\leqq x<y)\\ \gamma(\beta+\gamma)e^{-(\beta+\gamma)x} & (0\leqq y<x)\end{cases} \tag{2.5}$$

and $\beta, \gamma>0$. Then it is easy to see that, for $t<z_r<s$,

$$\begin{aligned}&\alpha\{(X>s,\ z_r-\varepsilon<Y\leqq z_r\})=\varepsilon\gamma e^{-(\beta+\gamma)s}\quad\text{and}\\ &\alpha(\{X>Y,\ z_r-\varepsilon<Y\leqq z_r)=\frac{\gamma}{\beta+\gamma}e^{-(\beta+\gamma)z_r}(e^{(\beta+\gamma)\varepsilon}-1)\end{aligned} \tag{2.6}$$

and hence by taking the limit of the ratio as $\varepsilon \to 0$, we have

$$\alpha'_s(z_r) = e^{(\beta+\gamma)(z_r-s)} . \tag{2.7}$$

Similarly we get

$$\alpha'_t(z_r) = e^{(\beta+\gamma)(z_r-t)} , \qquad s<z_r<t . \tag{2.8}$$

Also, a straight forward calculation yields,

$$\alpha((s, \infty)\times(t, \infty)) = e^{-\beta \max(s,t)}\{1+\beta(t-s)I[t>s]+\gamma(s-t)I[t<s]\} \tag{2.9}$$

with $I[A]$ the indicator function of the set A. Plugging these quantities in the formula (2.1) we get, for $s>t$,

$$\hat{\bar{F}}(s, t) = \frac{1}{N+1}\left\{e^{-(\beta+\gamma)s}[1+\gamma(s-t)]+N^+(s)+\sum_{t<z_r<s}\lambda_r e^{(\beta+\gamma)(z_r-s)}\right\} . \tag{2.10}$$

A similar expression can be obtained for the case $s<t$. For $s=t$, $\hat{\bar{F}}(s, \infty)=(N+1)^{-1}(\alpha((s, \infty)\times(s, \infty))+N^+(s))$.

3. Proofs

To prove the main result, we need the following lemmas.

LEMMA 3.1. *For any positive integer p, let $x^{(p)}=x(x+1)\cdots(x+p-1)$, the ascending factorial and $x^{(0)}=1$. Then we have*

$$\sum_{r=0}^{p}\binom{p}{r}a^{(r)}b^{(p-r)} = (a+b)^{(p)} \tag{3.1}$$

PROOF. We prove this lemma by the principle of mathematical induction. For $p=1$, the relationship is obviously true. So let us assume that it is true for k, that is,

$$\sum_{r=0}^{k}\binom{k}{r}a^{(r)}b^{(k-r)} = (a+b)^{(k)} . \tag{3.2}$$

Multiplying the ith summand on the left hand side of (3.2) by $(a+b+k)$ and using the fact that $x^{(l)}(x+l)=x^{(l+1)}$, it may be written as

$$\begin{aligned}\binom{k}{i}a^{(i)}b^{(k-i)}(a+b+k) &= \binom{k}{i}[a^{(i)}b^{(k-i)}(a+i)+a^{(i)}b^{(k-i)}(b+k-i)] \\ &= \binom{k}{i}[a^{(i+1)}b^{(k-i)}+a^{(i)}b^{(k)}] .\end{aligned} \tag{3.3}$$

Collecting the coefficients of like terms and noting that $\binom{k}{i}+\binom{k}{i-1}$

$=\binom{k+1}{i}$, we obtain

$$(3.4)\quad \sum_{r=0}^{k}\binom{k}{r}a^{(r)}b^{(k-r)}(a+b+k)=\binom{k+1}{0}a^{(0)}b^{(k+1)}+\cdots+\binom{k+1}{k+1}a^{(k+1)}b^{(0)}$$
$$=\sum_{r=0}^{k+1}\binom{k+1}{r}a^{(r)}b^{(k+1-r)}\,.$$

Since $(a+b)^k(a+b+k)=(a+b)^{(k+1)}$ on the r.h.s., the lemma is proved.

LEMMA 3.2. *With the notations of Lemma* 3.1, *we have*

$$(3.5)\qquad \text{(i)}\quad \sum_{r=0}^{p}(p-r)\binom{p}{r}a^{(r)}b^{(p-r)}=pb(a+b+1)^{(p-1)}$$

and

$$\text{(ii)}\quad \frac{1}{(a+b)^{(p)}}\sum_{r=0}^{p}(p-r)\binom{p}{r}a^{(r)}b^{(p-r)}=pb/(a+b)\,.$$

PROOF. Since $b^{(l)}=b(b+1)^{(l-1)}$, the l.h.s. of (i) is equal to

$$(3.6)\qquad \sum_{r=0}^{p-1}\frac{p(p-1)!}{r!(p-1-r)!}a^{(r)}b(b+1)^{(p-r-1)}=pb(a+b+1)^{(p-1)}$$

by Lemma 3.1. (ii) is obvious.

LEMMA 3.3. *Let* $(X_1,\cdots,X_m)$ *be jointly distributed as Dirichlet distribution with parameters* $(q_1,q_2,\cdots,q_m)$. *Then for any nonnegative integers* $\beta,\theta_1,\cdots,\theta_m$, *we have*

$$(3.7)\qquad \mathrm{E}\left\{\left(1-\sum_{i=1}^{j}X_i\right)^{\beta}X_1^{\theta_1}\cdots X_m^{\theta_m}\right\}$$
$$=\left\{\prod_{i=1}^{m-1}q_i^{(\theta_i)}\right\}\frac{\Gamma\sum_{i=1}^{m}q_i}{\Gamma\left[\sum_{i=1}^{m}(q_i+\theta_i)+\beta\right]}\left(\sum_{i=j+1}^{m}(q_i+\theta_i)\right)^{(\beta)}\,.$$

PROOF. The proof is straight forward.

To prove Theorem 2.1, we need the following representation of $\mathrm{E}[\bar{F}(s,t)|(\boldsymbol{\delta},\boldsymbol{Z})]$. Assume $0<t<s<\infty$ and for any $\varepsilon>0$, divide the x-axis as follows:

$$0<z_1-\varepsilon<z_1<\cdots<z_l-\varepsilon<z_l<t<z_{l+1}-\varepsilon<z_{l+1}<\cdots$$
$$<z_{l+k}<s<z_{l+k+1}-\varepsilon<z_{l+k+1}<\cdots<z_n<\infty\,.$$

Define for $i=1,\cdots,n$,

$$B_i=\{(x,y):\ x\leqq y,\ z_i-\varepsilon<x\leqq z_i\},\qquad \tilde{B}_i=\{j:\ (X_j,Y_j)\in B_i\}$$

$$C_i=\{(x, y):\ x>y,\ z_i-\varepsilon<y\leqq z_i\}, \qquad \tilde{C}_i=\{j:\ (X_j, Y_j)\in C_i\}$$

and $\quad D=\{(x, y):\ x>s,\ y>t\}$.

The sets B_i and C_i represent vertical and horizontal strips above and to the right of the line $x=y$, respectively. Then we have the following lemma. In the following lemma (and elsewhere), $(P(A))^0$ and $(I_A)^0$ are taken to be unity for any event A. Let $\prod_{j\in\tilde{B}_i}(I[(X_j, Y_j)\in B_i])=1$ if $\mu_i=0$, and $\prod_{j\in\tilde{C}_i}(I[X_j, Y_j)\in C_i])=1$ if $\lambda_i=0$.

LEMMA 3.4.

$$\mathrm{E}\,[\bar{F}(s, t)|(\boldsymbol{\delta}, \boldsymbol{Z})]=\lim_{\varepsilon\downarrow 0}\frac{\mathrm{E}\left\{P(D)\prod_{i=1}^{n}P^{\mu_i}(B_i)P^{\lambda_i}(C_i)\right\}}{\mathrm{E}\left\{\prod_{i=1}^{n}P^{\mu_i}(B_i)P^{\lambda_i}(C_i)\right\}}. \tag{3.8}$$

PROOF. By Fubini's theorem and the definition of conditional probability we have, if the limit below exists,

$$\begin{aligned}(3.9)\quad &\mathrm{E}\,\{\bar{F}(s, t)|(\boldsymbol{\delta}, \boldsymbol{Z})\}\\ &=\int_0^1\mathrm{P}\,(\bar{F}(s, t)>a|(\boldsymbol{\delta}, \boldsymbol{Z}))da\\ &=\lim_{\varepsilon\downarrow 0}\frac{\mathrm{E}\left\{\bar{F}(s, t)\prod_{i=1}^{n}[I[(X_i, Y_i)\in B_i]]^{\mu_i}[I[(X_i, Y_i)\in C_i]]^{\lambda_i}\right\}}{\mathrm{E}\left\{\prod_{i=1}^{n}[I[(X_i, Y_i)\in B_i]]^{\mu_i}[I[(X_i, Y_i)\in C_i]]^{\lambda_i}\right.},\end{aligned}$$

where the second equality follows since the conditional probability is bounded by 1 for all a, and by interchanging the limit and the integral. But the numerator of the limit in (3.9) is equal to (see Ferguson [2], pp. 216).

$$\begin{aligned}(3.10)\quad &\mathrm{E}\left\{\mathrm{E}\left[\bar{F}(s, t)\prod_{i=1}^{n}[I[(X_i, Y_i)\in B_i]]^{\mu_i}[I[(X_i, Y_i)\in C_i]]^{\lambda_i}|P(D),\right.\right.\\ &\qquad\qquad \left.\left.P(B_1), \cdots, P(B_n),\ P(C_1), \cdots, P(C_n)\right]\right\}\\ &=\mathrm{E}\left\{P(D)\prod_{i=1}^{n}P^{\mu_i}(B_i)P^{\lambda_i}(C_i)\right\}.\end{aligned}$$

Hence (3.8) follows.

PROOF OF THE THEOREM. In view of Lemma 3.4, the theorem will be proved by evaluating the expectations on the r.h.s. of (3.8). To do so, we will first express the sets B_i, C_i, and D as unions of some disjoint sets. Let

$$A_j=B_j, \qquad j=1, 2, \cdots, l+k,$$

$$
\begin{aligned}
&A_{l+k+j}=C_j\,, \qquad j=1,2,\cdots,l\,,\\
&\left.\begin{aligned}
&A_{l+k+j}=\{(x,y)\colon\ s\geqq x>y,\ z_j-\varepsilon<y\leqq z_j\}\\
&A_{l+2k+j}=\{(x,y)\colon\ x>y,\ x>s,\ z_j-\varepsilon<y\leqq z_j\}\,,
\end{aligned}\right\}\ j=l+1,\cdots,l+k\\
&\qquad\qquad (\text{thus } A_{l+k+j}\cup A_{l+2k+j}=C_j\,,\ j=l+1,\cdots,l+k)\,,
\end{aligned}
$$

$$
\begin{aligned}
(3.11)\quad &A_{l+2k+j}=C_j\,, \qquad j=l+k+1,\cdots,n\,,\\
&A_{n+k+j}=B_j\,, \qquad j=l+k+1,\cdots,n\,,\\
&A_0=\{(x,y)\colon\ 0<x\leqq s,\ y>0\}\cup\{(x,y)\colon\ x>0,\ 0<y\leqq t\}\Big\backslash\bigcup_{j=1}^{2l+2k}A_j\\
&\quad=R_2^+\Big\backslash\Big(D\cup\bigcup_{j=1}^{2l+2k}A_j\Big)
\end{aligned}
$$

and finally,

$$
A_{2n+k+1}=D\Big\backslash\bigcup_{j=2l+3k+1}^{2n+k}A_j\,.
$$

The sets A_i's are disjoint and $\bigcup_{i=0}^{2n+k+1}A_i=R_2^+$ and $D=\bigcup_{j=2l+2k+j}^{2n+k+1}A_j$. Now by expressing the sets B_i and C_i in terms of A_i, the numerator of (3.8) becomes

$$
\begin{aligned}
(3.12)\qquad \mathrm{E}\Big\{&P(D)\prod_{i=1}^{l+k}P^{\mu_i}(A_i)\prod_{i=1}^{l}P^{\lambda_i}(A_{l+k+i})\prod_{j=l+1}^{l+k}P^{\lambda_j}(A_{l+k+j}\cup A_{l+2k+j})\\
&\times\prod_{i=l+k+1}^{n}[P^{\lambda_i}(A_{l+2k+i})P^{\mu_i}(A_{n+k+i})]\Big\}\,.
\end{aligned}
$$

Denoting by $X_i=P(A_i)$, $i=0,1,2,\cdots,2n+k+1$, we have $\sum_{i=0}^{2n+k+1}X_i=1$, and expanding the binomial term $(X_{l+k+j}+X_{l+2k+j})^{\lambda_j}$, (3.12) can be written as

$$
\begin{aligned}
(3.13)\qquad \mathrm{E}\Big\{&\Big(1-\sum_{i=0}^{2l+2k}X_i\Big)\prod_{i=1}^{l+k}X_i^{\mu_i}\prod_{i=1}^{l}X_{l+k+i}^{\lambda_i}\sum_{j=l+1}^{l+k}\Big[\sum_{\gamma_j=0}^{\lambda_j}\binom{\lambda_j}{\gamma_j}X_{l+k+j}^{\gamma_j}X_{l+2k+j}^{\lambda_j-\gamma_j}\Big]\\
&\times\prod_{i=l+k+1}^{n}(X_{l+2k+i}^{\lambda_i}X_{n+k+i}^{\mu_i})\Big\}\,.
\end{aligned}
$$

Since P is a Dirichlet process with parameter α, $(X_0, X_1,\cdots,X_{2n+k+1})$ is distributed according to the Dirichlet distribution with parameters $(\alpha_0,\alpha_1,\cdots,\alpha_{2n+k+1})$ where $\alpha_i=\alpha(A_i)$, $i=0,1,\cdots,2n+k+1$. Using this, pulling the sums and binomial coefficients outside the products, and applying Lemma 3.3, (3.13) reduces to

$$
\begin{aligned}
(3.14)\qquad &\sum_{\gamma_{l+1}=0}^{\lambda_{l+1}}\cdots\sum_{\gamma_{l+k}=0}^{\lambda_{l+k}}\prod_{j=l+1}^{l+k}\binom{\lambda_j}{\gamma_j}Q\frac{1}{(\sum\alpha_i+N)}\prod_{j=l+1}^{l+k}[\alpha_{l+k+j}^{(\gamma_j)}\alpha_{l+2k+j}^{(\lambda_j-\gamma_j)}]\\
&\times\Big(\sum_{i=2l+2k+1}^{2n+k+1}+\prod_{r=l+1}^{l+k}(\lambda_r-\gamma_r)+\sum_{i=l+k+1}^{n}(\lambda_i+\mu_i)\Big)\,,
\end{aligned}
$$

where

$$(3.15)\qquad Q=\frac{\Gamma(\sum \alpha_i)}{\Gamma(\sum \alpha_i+N)}\prod_{i=1}^{l+k}\alpha_i^{(\mu_i)}\prod_{i=1}^{l}\alpha_{l+k+i}^{(\lambda_i)}\prod_{i=l+k+1}^{n}\alpha_{l+2k+i}^{(\lambda_i)}\alpha_{n+k+i}^{(\mu_i)},$$

$$\sum \alpha_i=\sum_{i=0}^{2n+k+1}\alpha_i=\alpha(R_2^+),\quad \sum(\mu_i+\lambda_i)=\sum_{i=1}^{n}(\mu_i+\lambda_i)=N.$$

By applying Lemma 3.1 repeatedly to (3.14), we obtain

$$(3.16)\qquad \sum_{\gamma_{l+1}=0}^{l+1}\cdots\sum_{\gamma_{l+k}=0}^{\lambda_{l+k}}\prod_{j=l+1}^{l+k}\binom{\lambda_j}{\gamma_j}\alpha_{l+k+j}^{(\gamma_j)}\alpha_{l+2k+j}^{(\lambda_j-\gamma_j)}=\prod_{j=l+1}^{l+k}(\alpha_{l+k+j}+\alpha_{l+2k+j})^{(\lambda_j)},$$

and

$$(3.17)\qquad \sum_{\gamma_{l+1}=0}^{\lambda_{l+1}}\cdots\sum_{\gamma_{l+k}=0}^{\lambda_{l+k}}\sum_{r=l+1}^{l+k}(\lambda_r-\gamma_r)\prod_{j=l+1}^{l+k}\binom{\lambda_j}{\gamma_j}\alpha_{l+k+j}^{(\gamma_j)}\alpha_{l+2k+j}^{(\lambda_j-\gamma_j)}$$
$$=\sum_{r=l+1}^{l+k}\sum_{\gamma_r=0}^{\lambda_r}(\lambda_r-\gamma_r)\binom{\lambda_r}{\gamma_r}\alpha_{l+k+r}^{(\gamma_r)}\alpha_{l+2k+r}^{(\lambda_r-\gamma_r)}\prod_{\substack{j\neq r\\ j=l+1}}^{l+k}(\alpha_{l+k+j}+\alpha_{l+2k+j})^{(\lambda_j)}.$$

Substitute (3.16) and (3.17) in (3.14) to get the numerator of (3.8).

A similar evaluation shows that the denominator of (3.8) is equal to

$$(3.18)\qquad Q\cdot\prod_{j=l+1}^{l+k}(\alpha_{l+k+j}+\alpha_{l+2k+j})^{(\lambda_j)}.$$

Substituting these two terms in (3.8) and cancelling the common terms, the r.h.s. of (3.8) reduces to

$$(3.19)\qquad \lim_{\epsilon\downarrow 0}\frac{1}{\alpha(R_2^+)+N}\left\{\alpha(D)+\sum_{i=l+k+1}^{n}(\lambda_i+\mu_i)\right.$$
$$\left.+\sum_{r=l+1}^{l+k}\frac{\sum_{\gamma_r=0}^{\lambda_r}(\lambda_r-\gamma_r)\binom{\lambda_r}{\gamma_r}\alpha_{l+k+r}^{(\gamma_r)}\alpha_{l+2k+r}^{(\lambda_r-\gamma_r)}}{(\alpha_{l+k+r}+\alpha_{l+2k+r})^{(\lambda_r)}}\right\}$$
$$=\frac{1}{\alpha(R_2^+)+N}\left\{\alpha(D)+\sum_{i=l+k+1}^{n}(\lambda_i+\mu_i)\right.$$
$$\left.+\sum_{r=l+1}^{l+k}\lambda_r\lim_{\epsilon\downarrow 0}\frac{\alpha(A_{l+2k+r})}{\alpha(A_{l+k+r}\cup A_{l+2k+r})}\right\}$$

where the equality follows by Lemma 3.2 (ii), and $\alpha(D)=\sum_{i=2l+2k+1}^{2n+k+1}\alpha_i$. Thus for $0<t<s<\infty$,

$$(3.20)\qquad \mathrm{E}[\bar{F}(s,t)|(\boldsymbol{\delta},\boldsymbol{Z})]$$
$$=\frac{1}{\alpha(R_2^+)+N}\left\{\alpha((s,\infty)\times(t,\infty))+N^+(s)+\sum_{\substack{r\ni:\\ t<z_r<s}}\lambda_r\alpha_s'(z_r)\right\}.$$

Similar expressions can be obtained for $0<s\leqq t<\infty$, and combining the two, we get (2.1) completing the proof the theorem.

4. Numerical example

To illustrate the method we rework the example given in Kaplan and Meier [6]. They take eight observations, four of which are censored. We assume here that the observations arise from a bivariate distribution function $F(x, y)$, which represents the joint distribution of lifetime (x-coordinate) and censored (y-coordinate) variables. Further, we observe only δ's and z's. Thus, their data is $\delta_1=\delta_4=\delta_5=\delta_7=1$, $\delta_2=\delta_3=\delta_6=\delta_8=0$, $z_1=0.8$, $z_2=1.0$, $z_3=2.7$, $z_4=3.1$, $z_5=5.4$, $z_6=7.0$, $z_7=9.2$ and $z_8=12.1$ (in months). Based on this data we calculate the Bayes estimator of the survival function $\bar{F}_1(s)=\bar{F}(s, o)$ and compare it with the estimators reported earlier in the literature under the assumption that the lifetime and censoring variables are independent.

We take the Dirichlet process parameter α to be the bivariate exponential distribution discussed in the example in Section 2. Then the Bayes estimator of survival function $\bar{F}(s, 0)$ becomes,

$$\hat{\bar{F}}_s(s, 0)=\frac{1}{N+1}\left\{e^{-(\beta+\gamma)s}(1+\gamma s)+N^+(s)+\sum_{z_r\leq s}\lambda_r e^{(\beta+\gamma)(z_r-s)}\right\}. \tag{4.1}$$

To use this estimator, we need to know β and γ. Here we observe that $\mathrm{E}\,[\delta_i I[Z_i\geqq 1]]=\beta(\beta+\gamma)^{-1}\exp\left(-(\beta+\gamma)\right)$, and $P(\delta_i=1)=\beta(\beta+\gamma)^{-1}$. Hence reasonable estimators of β and γ are given by

$$\begin{aligned}\hat{\beta}&=-\bar{\delta}\ln\left(\frac{\sum\delta_i I[Z_i\geqq 1]}{\sum\delta_i}\right)\\ \hat{\gamma}&=-(1-\bar{\delta})\ln\left(\frac{\sum\delta_i I[Z_i\geqq 1]}{\sum\delta_i}\right)\end{aligned} \tag{4.2}$$

with $\bar{\delta}=\frac{1}{n}\sum\delta_i$. For this data $\bar{\delta}=1/2$, $\hat{\alpha}=\hat{\beta}=.14384$. Plugging these values in (4.1), we evaluate $\hat{\bar{F}}_B$ at data points and it is given in Table 1 along with the Product-Limit estimator $\hat{\bar{F}}_{PL}$ (Kaplan and Meier [6]), and the Bayes estimator with Dirichlet prior assuming independence model $\hat{\bar{F}}_D$ of Susarla and Van Ryzin [10]. Other estimators are given in Ferguson and Phadia [3] but are not reproduced here. In Table 1, where there is only one value of an estimate at a data point, the estimated survival function is continuous. When there are two values, the upper and lower numbers represent the left and right limits of the function at that point.

It is clear that the present estimator is also continuous at the censored observations and has jumps only at the uncensored observations. However, it should be noted that the size of jump at each

Table 1. Numerical comparison of 3 estimators

t_δ	0.8	1.0	2.7	3.1	5.4	7.0	9.2	12.1
	1	0	0	1	1	0	1	0
$\hat{\bar{F}}_{PL}$	1.000			.875	.700		.525	
	.875	.875	.875	.700	.525	.525	.263	.263
$\hat{\bar{F}}_{D}$	.992	.878	.861	.856	.684	.522	.502	.271
	.880			.707	.535		.292	
$\hat{\bar{F}}_{B}$	.987	.873	.806	.781	.554	.415	.327	.158
	.876			.670	.443		.216	

real observation is the same $1/(N+1)$, which was not the case with either the PL estimator or the Bayes estimator. In these cases the amounts of jump was random and dependent upon the number of censored observations between two consecutive lifetime observations. Furthermore, it is clear from the table that this dependent model estimator of the survival function is below the other two estimators. This is significant and reasonable in the context of competing risk model where, if more than one related causes are competing for the life of a component (patient) the probability of surviving after time s would be smaller.

5. Concluding remarks

We make some comments concerning the large sample behaviour of $\hat{F}_\alpha(\cdot)$ in the classical setup, that is, under the assumption that $(X_1, Y_1), \cdots, (X_n, Y_n)$ are i.i.d. F_0. In this case, the Bayes estimator given in Section 2 need not be consistent because F_0 need not be identifiable if only one gets to observe the identified minimum (For special cases of F_0, Basu and Ghosh [1], and Gilliland and Hannan [4] show that F_0 is identifiable by observing only the identified minimum). However, one can get the large sample behaviour of $\hat{F}_\alpha(s, s)$, the Bayes estimator of $F_0(s, s)=P(Z_1>s)$ under squared error loss. In this case, $\hat{F}_\alpha(s, s)$ is uniformly (in s) a.s. consistent and converges weakly to a Gaussian process with mean function $=F_0(s, s)$ and covariance function $F_0(s, s)\times(1-F_0(t, t))$ for $0<t<s<\infty$. In a similar fashion, if one considers the Bayes estimator $\hat{F}_1(s)$ of $F_1(s)=P(\delta=1, Z_1>s)$, $0<s<\infty$, statements similar to those given above also hold for $\hat{F}_1(s)$. Note that both $F_0(s, s)$ and $F_1(s)$ are important characteristics from the practical point of view. For example, in the competing risk model, $F_0(s, s)$ is the probability of survival of more than s units of time when two causes of death or failure are present in the system. A similar interpretation can be made for $F_1(s)$. The results presented here can be generalized to more than two causes of failure.

Acknowledgements

The authors would like to thank the referee for carefully reading the manuscript and suggesting some notational improvements.

WILLIAM PATERSON COLLEGE, WAYNE, N.J., U.S.A.
STATE UNIVERSITY OF NEW YORK-BINGHAMTON, U.S.A.

REFERENCES

[1] Basu, A. P. and Ghosh, J. K. (1978). Identifiability of the multinormal and other distributions under competing risks model, *J. Multivariate Anal.* 8, 413-429.

[2] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems, *Ann. Statist.*, **1**, 209-230.

[3] Ferguson, T. S. and Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data, *Ann. Statist.*, **7**, 163-186.

[4] Gilliland, D. and Hannan, J. (1980). Identification of the ordered bivariate normal distribution by minimum variate, *J. Amer. Statist. Ass.*, **75**, 651-654.

[5] Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics-Continuous Multivariate Distributions*, J. Wiley & Sons, New York.

[6] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *J. Amer. Statist. Ass.*, **53**, 457-481.

[7] Lagokos, S. W. and Williams, J. S. (1978). Models for censored survival analysis: A cone class of variable-sum models, *Biometrika*, **65**, 181-189.

[8] Langberg, N., Proschan, F. and Quinzi, A. J. (1981). Estimating dependent life lengths, with applications to the theory of competing risks, *Ann. Statist.*, **9**, 157-167.

[9] Miller, R. (1980). Survival analysis, *Tech. Report* No. 58, Divison of Biostatistics, Stanford University.

[10] Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations, *J. Amer. Statist. Ass.*, **71**, 897-902.