The Annals of Statistics
1981, Vol. 9, No. 5, 1061–1065

A NOTE ON A PAPER BY FERGUSON AND PHADIA

BY C. J. WILD AND J. D. KALBFLEISCH

University of Auckland and University of Waterloo

Ferguson and Phadia have recently discussed the nonparametric Bayesian estimation of a distribution function from a right-censored random sample using process priors that are neutral to the right. The more general problem of estimating the baseline distribution function with right censored data from the proportional hazards model has been studied by Kalbfleisch who uses the more restrictive class of gamma process priors. This note shows that, with a simple modification, the analysis of Ferguson and Phadia can be extended to deal with the proportional hazards situation with constant covariates. Extensions to time dependent covariates and other regression models are also considered.

1. Introduction. Suppose that a right-censored sample of survival times is available from a probability model for survival time x with distribution function $F_0(t)$. Nonparametric Bayesian estimation of $F_0(t)$ has been considered by Susarla and Van Ryzin (1976) using a Dirichlet process, and more recently by Ferguson and Phadia (1979) for the general class of neutral to the right processes. Doksum (1974) has defined a random distribution function as being neutral to the right if the normalized increments

$$F(t_1),\ \{F(t_2) - F(t_1)\}/\{1 - F(t_1)\},\ \cdots,\ \{F(t_k) - F(t_{k-1})\}/\{1 - F(t_{k-1})\}$$

are independent for all $t_1 < t_2 < \cdots < t_k$ and Kalbfleisch (1978) has given an interpretation of this definition in the survival context. The results of Ferguson and Phadia (1979) follow by noting that if the prior distribution of F_0 is neutral to the right, then given either a censored or observed survival time x, the posterior distribution is again neutral to the right. It then follows that the remainder of the data can be brought in sequentially without difficulty.

Let the failure time of individual i be X_i and suppose that $\mathbf{W}_i^T = (W_{i1}, \cdots, W_{ip})$ is a vector of p measured time independent covariates. The proportional hazards model of Cox (1972) specifies that the distribution function of X_i is

$$F_i(x) = 1 - \{1 - F_0(x)\}^{\exp(\boldsymbol{\beta}^T \mathbf{W}_i)}, \tag{1}$$

where $\boldsymbol{\beta}$ is a vector of regression coefficients and F_0 is a baseline distribution function and is left unspecified. It should be noted that in Cox's formulation, the covariates were allowed to be functions of time, $\mathbf{W}_i(x)$ and the model was specified in terms of the hazard function relationship

$$\lambda_i(x) = \lambda_0(x) \exp(\boldsymbol{\beta}^T \mathbf{W}_i),$$

which is valid for both the fixed and time varying cases. The expression (1) is appropriate for fixed covariates only. In what follows, only the case $\mathbf{W}_i$ fixed is considered. Kalbfleisch (1978) discusses estimation of F_0 for arbitrary $\boldsymbol{\beta}$ but for the case where the prior distribution of F_0 is a particular process neutral to the right, the gamma process. In this note, we show that only minor adjustments to the derivations of Ferguson and Phadia enable the

Received November 22, 1979; revised February 10, 1981.

AMS 1970 *subject classifications.* Primary 62C10; Secondary 62605, 62P10, 60G57.

Key words and phrases. Bayesian nonparametric methods, censored data, Dirichlet process, gamma process, neutral-to-the-right processes, proportional hazards model, random probability measures, survival time.

extension of these results to the more general regression model (1) and thus places the analysis of Kalbfleisch in a more general setting.

2. Results for neutral to the right processes. Suppose that X has distribution function $F(t)$ and that $\mathbf{W}$ is the associated covariate. Suppose further that the prior distribution of F_0 is neutral to the right and that F and F_0 are related as in (1). Let $Y_t = -\log\{1 - F(t)\}$ and $Y_{0t} = -\log\{1 - F_0(t)\}$. Since $Y_t = Y_{0t}\exp(\boldsymbol{\beta}^T\mathbf{W})$ it follows from Definition 1 of Ferguson and Phadia that the prior distribution of F is also neutral to the right. As Doksum (1974) has noted, both Y_{0t} and Y_t can be specified as processes with independent non-negative increments.

Theorems 1 and 2 below follow directly by applying Theorems 1, 2, and 3 of Ferguson and Phadia to Y_t and then transforming to corresponding results for Y_{0t}. It should be noted that it is at this point that time dependent $\mathbf{W}$ must be excluded since the relationship (1) between distribution functions is crucial.

THEOREM 1. *If $X = x$ is a random sample of size one from F and the prior distribution of F_0 is neutral to the right, then*

(i) *the posterior distribution of F_0 is neutral to the right;*

(ii) *the posterior distribution of an increment in Y_{0t} to the right of x is the same as its prior distribution;*

(iii) *an increment $Y_{0t} - Y_{0s}$ with prior density $dG(y)$ for $s < t < x$, has posterior density proportional to*

$$\exp(-ye^{\boldsymbol{\beta}^T\mathbf{W}})\, dG(y);$$

(iv) *there is a jump discontinuity at x in the posterior process Y_{0t}, $Y_{0x} - Y_{0x}^-$, whether or not there was one in the prior;*

(v) *if $S = Y_{0x} - Y_{0x}^-$ has prior density $dG_x(s)$, then it has posterior density*

$$dH_x(s) \propto \{1 - \exp(-se^{\boldsymbol{\beta}^T\mathbf{W}})\}\, dG_x(s)$$

If there is no prior jump in Y_{0t} at x, the posterior distribution of $Y_{0x} - Y_{0x}^-$ is specified in terms of Lévy measures by similarly generalizing Theorem 5 of Ferguson and Phadia (1979) or Case 2, Section 4 of Ferguson (1974). In examples, however, it is frequently simpler to use the method described by Kalbfleisch (1978) involving differentiating normalized moment generating functions.

We now consider the posterior process given a single right censored observation. Exclusive right censoring ($X > u$) is most common in practice and for clarity of presentation this is the only form of censoring discussed.

THEOREM 2. *The posterior distribution of F_0 given $X > x$ is neutral to the right. The posterior distribution of an increment in Y_{0t} to the right of x is the same as the prior, while the posterior distribution of $Y_{0t} - Y_{0s}$, $0 < s < t \le x$, has posterior density as in Theorem* 1 *(iii).*

Using these results and bringing the data points in sequentially, we obtain Theorem 3, which generalizes Theorem 4 of Ferguson and Phadia. Assume the data $\{x_j; \mathbf{w}_j\}$ are of two types, failure times and right censoring times. Let $u_1 < u_2 < \cdots < u_k$ be the ordered distinct x_j's, and $R(t)$ the set of labels of items at risk immediately prior to time t. Let $C(i)$, respectively $D(i)$, represent the set of labels of items censored, respectively failing, at u_i. Designate one member of $D(i)$ as "$i1$" and let $D(i)^- = D(i) - \{i1\}$. The prior distribution of a jump, S, in Y_{0t} at u is denoted by $G_u(s)$ with associated moment generating function $M_S^{(0)}(\theta)$, while $H_{u_i}(s)$ denotes the posterior distribution of the jump S_i in Y_{0t} at u_i given the failure at u_i of the item labelled. "$i1$". Let $j(t)$ be the number of u_i's less than or equal to t, $M_t(\theta)$ denote the moment generating function of Y_{0t} and $M_t^-(\theta) = \lim_{s\uparrow t} M_s(\theta)$. It should be noted that all moment generating functions exist for $\theta \le 0$ and $M_t(-1\,|\,\text{Data})$ is the posterior expectation of $F_0(t)$. Then arguing as in Ferguson and Phadia (1979), we obtain

THEOREM 3. *The posterior moment generating function of Y_{0t} is*

$$(2)\quad M_t(\theta \mid \text{Data}) = \frac{M_t(\theta - h_{j(t)+1})}{M_t(-h_{j(t)+1})} \times \prod_{i=1}^{j(t)} \left[\frac{M_{u_i}^-(\theta - h_i)}{M_{u_i}^-(-h_i)} \cdot \frac{C_{u_i}(\theta - h_{i+1} - g_i, D(i))}{C_{u_i}(-h_{i+1} - g_i, D(i))} \cdot \frac{M_{u_i}(-h_{i+1})}{M_{u_i}(\theta - h_{i+1})} \right],$$

where $h_i = \sum_{l \in R(u_i)} \exp(\boldsymbol{\beta}^T \mathbf{W}_i)$ and $g_i = \sum_{l \in C(i)} \exp(\boldsymbol{\beta}^T \mathbf{W}_i)$.

If u_i is a prior fixed point of discontinuity of Y_{0t} then

$$(3)\quad C_{u_i}(\alpha, D(i)) = \int_0^\infty e^{\alpha s} \prod_{l \in D(i)} \{1 - \exp(-s e^{\boldsymbol{\beta}^T \mathbf{W}_l})\}\, dG_{u_i}(s),$$

and otherwise

$$(4)\quad \begin{aligned} C_{u_i}(\alpha, D(i)) &= 1, \quad \text{if } D(i) \text{ is empty, and} \\ &= \int_0^\infty e^{\alpha s} \prod_{l \in D(i)} - \{1 - \exp(-s e^{\boldsymbol{\beta}^T \mathbf{W}_l})\}\, dH_{u_i}(s), \quad \text{otherwise.} \end{aligned}$$

Now (3) can be written

$$C(\alpha, D(i)) = \sum_{j=0}^{m_i} (-1)^j \sum_{b_{ij} \in B_{ij}} M_{S_i}^{(0)}\{\alpha - \sum_{l \in b_{ij}} \exp(\boldsymbol{\beta}^T \mathbf{W}_l)\},$$

where m_i is the number of failures at u_i and B_{ij} is the class of all subsets of j labels taken from $D(i)$. An obvious modification gives a similar expression for (4).

3. Sampling from (1) with a Dirichlet process prior. For various reasons, including ease of interpretation, Ferguson and Phadia recommend the Dirichlet process among prior processes. Kalbfleisch (1978) uses the gamma process in sampling from (1) mainly for mathematical convenience, but he also gives interpretations of the parameters similar to those available for the Dirichlet process. In this section, we use the results above to find the posterior process based on a censored sample from (1) with a Dirichlet prior. This serves to illustrate the results in Section 2 and also facilitates comparison with the results with a gamma process prior in Kalbfleisch (1978).

For a Dirichlet process with measure α on $(0, \infty)$, the prior increments of Y_{0t} have a-log beta distribution with moment generating function

$$M_t(\theta) = \frac{B(\alpha_t - \theta, c - \alpha_t)}{B(\alpha_t, c - \alpha_t)},$$

where $c = \alpha(0, \infty)$, $\alpha_t = \alpha[t, \infty)$ and $B(a, b)$ is the beta function. In order to simplify the algebra for purposes of illustration, we restrict attention to the case of no ties among the uncensored observations and of α absolutely continuous with respect to Lebesgue measure. The latter avoids complications with fixed points of discontinuity and the former makes evaluation of $C_{u_i}(\alpha, D(i))$ less cumbersome. The jump S in the posterior process at $t = u$ has moment generating function

$$M_S^{(1)}(\theta) \propto \psi(\alpha_u + e^{\boldsymbol{\beta}^T \mathbf{W}} - \theta) - \psi(\alpha_u - \theta),$$

where $\psi(x)$ is the digamma function $\psi(x) = d \log \Gamma(x)/dx$, from which $H_u(s)$ can be obtained. If $\boldsymbol{\beta} = \mathbf{0}$ then S has a $-\text{logbeta}(\alpha_u, 1)$ distribution but there is no simple form when $\boldsymbol{\beta} = \mathbf{0}$. For $t \in [u_{i-1}, u_i)$ the posterior distribution of Y_{0t} is that of a sum of independent random variables

$$X_1 + S_1 + \cdots + X_{i-1} + S_{i-1} + \delta_i$$

where $X_j \sim -\text{logbeta}(\alpha_{u_j} + h_j, \alpha_{u_{j-1}} - \alpha_{u_j})$, S_j has moment generating function proportional to

$$\psi(\alpha_{u_j} + h_j - \theta) - \psi(\alpha_{u_j} + h_{j+1} - \theta)$$

and $\delta_i \sim -\text{logbeta}(\alpha_t + h_i, \alpha_{u_{i-1}} - \alpha_t)$.

As shown by Ferguson and Phadia, when $\boldsymbol{\beta} = \mathbf{0}$ the posterior expectation of $F_0(t) = M(-1 \mid \text{Data})$ reduces to the estimator of the distribution function derived by Susarla and Van Ryzin (1976). This estimator tends to the maximum likelihood estimator of Kaplan and Meier (1958) as $c = \alpha[0, \infty) \to 0$. Using a gamma process prior, the posterior expectation of $F_0(t)$ does not tend to the Kaplan-Meier estimate as $c \to 0$ but rather a first order approximation to it. It is argued by Ferguson and Phadia that, for this reason, the Dirichlet process prior is preferable.

If $\boldsymbol{\beta}$ is known, a Bayes estimator of $F_0(t)$ under squared error loss can be obtained in similar manner as the posterior expectation of $F_0(t)$. Alternatively, if losses are proportional to the squared error in the estimate of $\log\{1 - F_0(t)\}$, then the posterior expectation of Y_{0t} is the Bayes estimator. The corresponding nonparametric maximum likelihood estimator ($\boldsymbol{\beta}$ known) has been given by Kalbfleisch and Prentice (1973). With either the gamma or Dirichlet process priors, both of the above Bayes estimators tend to a first order approximation of the maximum likelihood estimator but neither to that estimator itself when $\boldsymbol{\beta} \neq \mathbf{0}$. Only when $\boldsymbol{\beta} = 0$ does the maximum likelihood estimator arise as the limit of Bayes estimators under a Dirichlet process prior. This suggests that the preference of Ferguson and Phadia for the Dirichlet process prior based on its relationship to the maximum likelihood estimator is somewhat artificial.

A fairly general method for the estimation of $\boldsymbol{\beta}$ from the marginal distribution of the data, $F_0(\cdot)$ having been eliminated, has been described by Kalbfleisch (1978) who gives explicit results for the gamma process prior. Explicit results for a Dirichlet process prior have been given by Wild (1979). In general, once an estimate of $\boldsymbol{\beta}$ has been found, empirical Bayes estimators of $F_0(t)$ can be found by replacing $\boldsymbol{\beta}$ in the Bayes estimator by its estimated value.

4. Possible Extensions. The above results depend, for their simplicity, upon the fact that the prior and the posterior distributions of $F_0(\cdot)$ and $F(\cdot)$ are neutral to the right. In this section, we show that the proportional hazards regression model is the only regression model in a broad class of models that has this property. It appears that application to other regression models will therefore be more difficult and require somewhat different arguments.

For the neutral to the right property, we require that

$$F_0(t_1),\ 1 - \{1 - F_0(t_2)\}/\{1 - F_0(t_1)\}$$

be independent for all $0 < t_1 < t_2$. Suppose that $F_0(\cdot)$ is a baseline distribution function for covariate value $\mathbf{W} = \mathbf{0}$, and that the distribution function F, given the covariate value $\mathbf{W}$, is obtained from F_0 by a regression model according to the relationship

$$1 - F(t) = g(1 - F_0(t)),$$

where g is strictly monotone. This formulation includes the proportional hazards class of models with constant covariates, or indeed any model that can be transformed to have additive errors with a strictly increasing distribution function.

For F to be neutral to the right, we must have

$$1 - g(X),\ 1 - g(Y)/g(X)$$

independent where $X = 1 - F_0(t_1)$ and $Y = 1 - F_0(t_2)$. For this to hold, it must be that $g(y)/g(x) = h(y/x)$ for almost all x, y, $0 < y < x < 1$. Consider $0 < y < x < z < 1$ and note that $h(y/x)h(x/z) = h(y/z)$ must hold. Equivalently,

$$h(s)h(t) = h(st)$$

for all $0 < s, t < 1$. This is a standard form of Cauchy's functional equation with an essentially unique solution $h(s) = s^c$ for some constant c. Now since $g(1) = 1$ it follows that

$$g(x) = x^c$$

with $c > 0$ since g is monotone increasing and this is the proportional hazards class with constant covariates.

REFERENCES

COX, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. Ser. B*, **34** 187–220.

DOKSUM, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2** 183–201.

FERGUSON, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2** 615–629.

FERGUSON, T. S. and PHADIA, E. G. (1979). Bayesian nonparametric estimation based on censored data. *Ann. Statist.* **7** 163–186.

KALBFLEISCH, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *J. Roy. Statist. Soc. Ser. B* **40** 214–221.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika* **60** 267–278.

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

SUSARLA, V. and VAN RYZIN, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *J. Amer. Statist. Assoc.* **71** 897–902.

WILD, C. J. (1979). Problems in estimation with matched data. Unpublished Ph.D. thesis, Univ. Waterloo, Canada.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF AUCKLAND
AUCKLAND, NEW ZEALAND

DEPARTMENT OF STATISTICS
UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO
CANADA