J. R. Statist. Soc. B (1978),
40, *No.* 2, *pp.* 214–221

Non-parametric Bayesian Analysis of Survival Time Data

By John D. Kalbfleisch

University of Waterloo

[Received April 1977. Final revision March 1978]

Summary

A Bayesian analysis of the semi-parametric regression and life model of Cox (1972) is given. The cumulative hazard function is modelled as a gamma process. Both estimation of the regression parameters and of the underlying survival distribution are considered. The results are compared to the results obtained by other approaches.

Keywords: SURVIVAL DISTRIBUTIONS; BAYESIAN STATISTICS; NON-PARAMETRICS; RANDOM PROBABILITY MEASURE; REGRESSION

1. Introduction

Let $T>0$ represent the failure time of an individual for whom $\mathbf{z}=(z_1,\ldots,z_p)$ are p measured covariables. Cox (1972) proposed that the distribution of T might be modelled by taking the hazard function of T given $\mathbf{z}$ to be

$$\lambda(t\,|\,\mathbf{z}) = \lim_{\Delta t\to 0} P\{T\in[t,t+\Delta t)\,|\,T\geqslant t,\mathbf{z}\}/\Delta t = \lambda(t)\exp(\mathbf{z}\boldsymbol{\beta}), \tag{1}$$

where $\boldsymbol{\beta}$ is a column vector of p regression parameters and $\lambda(t)$, an arbitrary baseline hazard function, is left unspecified. Alternatively, the model (1) can be specified in terms of the survivor function of T as

$$P\{T\geqslant t\,|\,\mathbf{z}\} = \exp\{-\Lambda(t)\exp(\mathbf{z}\boldsymbol{\beta})\}, \tag{2}$$

where $\exp\{-\Lambda(t)\} = P\{T\geqslant t\,|\,\mathbf{z}=\mathbf{0}\}$ is a base line survivor function and is left unspecified. The expression (1) is appropriate for T a continuous random variable and leads to the name "proportional hazards model". Expression (2) has the advantage of incorporating both discrete and continuous survival distributions. That is, (2) is a valid model whether $\exp\{-\Lambda(t)\}$ is discrete continuous or a mixture of these.

In the continuous case, $\Lambda(t)=\int_0^t \lambda(u)\,du$ and this has led to the name "cumulative hazard" for $\Lambda(t)$. When discrete and continuous cases are being combined, however, this name is misleading since $\Lambda(t)$ is not the sum of discrete hazards.

The analysis of the continuous model specified by (1) and (2) was considered by Cox (1972) who obtained a partial likelihood (Cox, 1975) for $\boldsymbol{\beta}$ by arguing conditionally at each observed failure point. If $t_{(1)},\ldots,t_{(k)}$ are the ordered failure times with corresponding covariates $\mathbf{z}_{(1)},\ldots,\mathbf{z}_{(k)}$, there being no ties, the partial likelihood

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{k}\{\exp(\mathbf{z}_{(i)}\boldsymbol{\beta})/\sum_{l\in R(t_{(i)})}\exp(\mathbf{z}_l\boldsymbol{\beta})\}, \tag{3}$$

where $R(t_{(i)})$ is the set of labels attached to the individuals at risk of failing just prior to $t_{(i)}$. Kalbfleisch and Prentice (1973) also obtained the result (3) as a marginal likelihood for $\boldsymbol{\beta}$ arising out of the distribution of the rank vector associated with the failure times.

In this paper, we consider a Bayesian method of analysing (2). First, $\Lambda(t)$ is treated as a nuisance parameter with a certain prior. Estimation of $\boldsymbol{\beta}$ is carried out by determining the marginal probability distribution of data as a function of $\boldsymbol{\beta}$, $\Lambda(t)$ having been eliminated.

The results of this procedure are compared with the result (3). The final section considers computation of the posterior distribution of $\Lambda(t)$ given the data and the associated problem of estimating $\Lambda(t)$.

This work is closely related to that of Cornfield and Detre (1977) who considered a similar problem. The approach is mathematically more satisfying, however, and the results give corrections to those reported by them. Their methods and results are commented on in a separate communication (Kalbfleisch and MacKay, 1978) and the connection with the present approach is also considered there.

2. The Prior Distributions

In this section, the class of tail-free or neutral to the right random probabilities considered by Doksum (1974) is outlined; a particular example, a gamma process, is chosen from this class and its properties are examined. For the moment, we consider only specification of the distribution of $\Lambda(t)$. The model (2) is returned to and analysed in Section 3.

Let the survivor function of T be $\exp\{-\Lambda(t)\} = P\{T \geqslant t \mid \Lambda\}$. The probability statement is shown conditional on Λ since Λ, the parameter in the model, is the realization of a stochastic process to be defined. We consider a partition of $[0,\infty)$ into a finite number k of disjoint intervals $[a_0 = 0, a_1), [a_1, a_2), \ldots, [a_{k-1}, a_k = \infty)$ and define the hazard contribution of the ith interval as

$$q_i = P\{T \in [a_{i-1}, a_i) \mid T \geqslant a_{i-1}, \Lambda\} \tag{4}$$

if $P\{T \geqslant a_{i-1} \mid \Lambda\} > 0$ and otherwise $q_i = 1$ $(i = 1, \ldots, k)$. Clearly then we have $\Lambda(a_0) = 0$ and

$$\Lambda(a_i) = \sum_{j=1}^{i} -\log(1-q_j) = \sum_{j=1}^{i} r_j \quad (i = 1, \ldots, k). \tag{5}$$

Doksum (1974) has considered this situation and has shown that a probability distribution can be specified on the space $\{\Lambda(t)\}$ by specifying the finite dimensional distributions of $q_1, \ldots, q_k$ for each partition $[a_{i-1}, a_i)$ $(i = 1, \ldots, k)$. Accordingly, independent prior probability densities can be specified for $q_1, \ldots, q_k$ subject to some consistency conditions and the resulting processes are called tailfree or neutral to the right by Doksum. It should be stressed that the distributional assumptions hold for all possible partitions and that there must be consistency between partitions. Examination of (5) makes it clear that $\Lambda(t)$ is by this construction a non-decreasing process with independent increments. This process in $\Lambda(t)$ is called a subordinator in Kingman (1975).

The problem then reduces to the specification of a non-decreasing independent increments process for $\Lambda(t)$. To do this, one need only specify independent priors for the r_i's (or for the q_i's) subject to the condition that the distribution of $r_i + r_{i+1}$ must be the same as would be obtained by direct application of the rules to the combined interval $[a_{i-1}, a_i)$.

The Dirichlet process (Ferguson, 1973, 1974) was the first random probability measure used in a Bayesian context. The process can be defined in very general probability spaces but, if the observable T has a natural ordering as above, it is convenient to think of the process as unfolding sequentially. In this formulation, the q_i's have independent beta prior distributions and full definition of the process proceeds as for the gamma process discussed below. Recently, Susarla and Van Ryzin (1976) have considered applications of the Dirichlet process to the estimation of the survivor function with censored data. Some of the results in their paper are more easily obtained by viewing the process as independent increments in $\Lambda(t)$.

A second process, similar to the Dirichlet process, is obtained if $\Lambda(t)$ is a gamma process similar to that used by Moran (1959) in modelling dam inputs. More specifically, let c be a positive real number and $\exp\{-\Lambda^*(t)\}$ be a completely specified survivor function. Let $r_i = -\log(1-q_i)$ have the independent gamma distributions

$$r_i \sim G(\alpha_i - \alpha_{i-1}, c) \quad (i = 1, \ldots, k), \tag{6}$$

where $\alpha_i = c\Lambda^*(a_i)$. The notation $X \sim G(a, b)$ indicates that X has the gamma distribution with parameters a (shape) and b (scale). The conventions adopted here are that $G(0, c)$ is the distribution with unit mass at 0 and $G(\infty, c)$ or $G(\infty-\infty, c)$ is the distribution with unit mass at ∞. It is clear that $\Lambda(t)$ is by this construction a gamma process and we shall write $\Lambda \sim \mathbf{G}(c\Lambda^*, c)$.

The parameters c and Λ^* can be seen to have simple interpretations similar to the interpretations of the analogous parameters of the Dirichlet process (Ferguson, 1973). Consideration of the partition $(0, t]$, $[t, \infty)$ shows immediately from (6) that $\Lambda(t) \sim G\{c\Lambda^*(t), c\}$ and $E\{\Lambda(t)\} = \Lambda^*(t)$, $\operatorname{var}\{\Lambda(t)\} = \Lambda^*(t)/c$. The specified Λ^* can be viewed as an initial guess at Λ and c is a specification of the weight attached to that guess.

One deficiency of the gamma process and the Dirichlet process is that with probability one, a realization $\Lambda(t)$ is discrete. This property has several ramifications in the next two sections where the gamma process is used as a prior distribution for $\Lambda(t)$ in the model (2). The difficulties with discreteness are discussed further in Sections 5 and 6.

3. Application to the Regression Model

Let T_i be a random variable with conditional survivor function

$$P(T_i \geq t_i \mid \mathbf{z}_i, \Lambda) = \exp\{-\Lambda(t_i)\exp(\mathbf{z}_i \boldsymbol{\beta})\} \quad (i = 1, \ldots, n) \tag{7}$$

independently. We suppose that $\Lambda \sim \mathbf{G}(c\Lambda^*, c)$ and consider the problem of estimating $\boldsymbol{\beta}$ on the basis of the data $(t_1, \mathbf{z}_1), \ldots, (t_n, \mathbf{z}_n)$. One way to proceed is to calculate the probability density of $t_1, \ldots, t_n$ conditional on the $\mathbf{z}_i$'s, Λ having been eliminated, and to interpret this as a likelihood function for $\boldsymbol{\beta}$ (Cox, 1976). Conditional on Λ,

$$P(T_1 \geq t_1, \ldots, T_n \geq t_n \mid \boldsymbol{\beta}, \mathbf{Z}, \Lambda) = \exp\{-\textstyle\sum \Lambda(t_i)\exp(\mathbf{z}_i \boldsymbol{\beta})\}, \tag{8}$$

where $\mathbf{Z}$ is the design matrix with ith column $\mathbf{z}_i$. Without loss of generality, we consider $t_1 \leqslant t_2 \leqslant \ldots \leqslant t_n$ and define $r_i = \Lambda(t_i) - \Lambda(t_{i-1})$ $(i = 1, \ldots, n+1)$ where $t_0 = 0$ and $t_{n+1} = \infty$. The t_i's are now playing the role of the a_i's in the last section. We have further that

$$r_i \sim G\{c\Lambda^*(t_i) - c\Lambda^*(t_{i-1}), c\} \quad (i = 1, \ldots, n+1) \tag{9}$$

independently, and since $\Lambda(t_i) = \sum_1^i r_j$ $(i = 1, \ldots, n+1)$, (8) implies

$$P(T_1 \geq t_1, \ldots, T_n \geq t_n \mid \boldsymbol{\beta}, \mathbf{Z}, r_1, \ldots, r_{n+1}) = \exp\left\{-\sum_1^n r_j A_j\right\}, \tag{10}$$

where

$$A_j = \sum_{l \in R(t_j)} \exp(\mathbf{z}_l \boldsymbol{\beta}) \quad (j = 1, \ldots, n) \tag{11}$$

and $R(t_j)$ is the set of individuals at risk at time $t_j - 0$. Integrating (10) with respect to the distribution (9) of $r_1, \ldots, r_n$ gives

$$P(T_1 \geq t_1, \ldots, T_n \geq t_n \mid \boldsymbol{\beta}, \mathbf{Z}) = \exp\{-\textstyle\sum cB_j \Lambda^*(t_j)\}, \tag{12}$$

where $B_j = -\log\{1 - \exp(\mathbf{z}_j \boldsymbol{\beta})/(c + A_j)\}$.

The expression (12) is valid for any cumulative hazard $\Lambda^*(t)$. In order to avoid problems with fixed discontinuities, however, we shall assume that $\Lambda^*(t)$ is absolutely continuous. The multiple decrement function (11) is then absolutely continuous except along any hyperplane with $t_i = t_j$ for some $i \neq j$. Thus if there are no ties in the data, $(t_1 < t_2 < \ldots < t_n)$ the p.d.f. of $T_1, \ldots, T_n$ is computed by differentiation and yields

$$L(\beta) = c^n \exp\{-\textstyle\sum cB_j \Lambda^*(t_j)\} \displaystyle\prod_1^n \{\lambda^*(t_i) B_i\}, \tag{13}$$

where $\lambda^*(t) = (d/dt)\Lambda^*(t)$. The expression (13) can be interpreted as a likelihood function for $\boldsymbol{\beta}$ on the data $T_1 = t_1, \ldots, T_n = t_n$. Right censoring is easily accommodated since the appropriate likelihood is obtained by differentiating (12) only with respect to the observed failure times. This gives

$$\exp\{-\sum cB_j \Lambda^*(t_j)\} \prod_1^n \{c\lambda^*(t_i) B_i\}^{d_i}, \tag{14}$$

where $d_j = 0$ or 1 for censored or failure times t_j respectively. The standard convention is adopted here that censored times tied with failure times are adjusted an infinitesimal amount to the right.

Two cases are of particular interest. If c is near 0, then to a first-order approximation

$$L(\boldsymbol{\beta}) \cong K \prod_1^n -\log\{1-\exp(\mathbf{z}_j \boldsymbol{\beta})/(c+A_i)\} \cong K \prod \exp(\mathbf{z}_i \boldsymbol{\beta})/A_i. \tag{15}$$

The last term in (15) is proportional to the partial likelihood or the marginal likelihood of $\boldsymbol{\beta}$. Small values of c correspond to placing little faith in the prior estimate $\Lambda^*(t)$ of $\Lambda(t)$. On the other hand,

$$\lim_{c\to\infty} L(\boldsymbol{\beta}) = \exp\{-\sum \Lambda^*(t_j)\exp(\mathbf{z}_j \boldsymbol{\beta})\} \prod \lambda^*(t_i)\exp(\mathbf{z}_i \boldsymbol{\beta}),$$

which is the appropriate likelihood if it is assumed that $\Lambda(t) = \Lambda^*(t)$ at the outset.

In effect, (13) gives a spectrum of likelihoods ranging from truly non-parametric situations (c near 0) to situations where $\Lambda(t)$ is assumed completely known. By allowing $\Lambda^*(t)$ to depend on one or more unknown parameters, for example $\Lambda^*(t) = \lambda t$, the likelihood (14) corresponds to a generalization of the usual parametric analysis ($c \to \infty$) with exponential survivals. An examination of this likelihood for varying c can lead to an evaluation of how assumption dependent is the analysis.

The leukaemia data of Gehan (1965) reproduced in Table 1 can be used to illustrate this

TABLE 1.

Times of remission in weeks of leukaemia patients

Group 1 (Z = 0·5)	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22+ε†, 23+ε†
Group 0 (Z = −0·5)	6, 6, 6, 6‡, 7, 9‡, 10, 10‡, 11‡, 13, 16, 17‡, 19‡, 20‡, 22, 23, 25‡, 32‡, 32‡, 34‡, 35‡

† These data were 22, 23 but were adjusted slightly in the positive direction to break the ties with items in group 0.
‡ Censored.

point. The covariable is treatment group specified by the indicator variable $z = -0{\cdot}5, 0{\cdot}5$. We select $\Lambda^*(t) = \lambda t$ where λ is left unspecified. The censored results (14) then gives a joint likelihood for λ and β for each specified c

$$L(\beta, \lambda) \exp(-c\lambda \sum t_j B_j) \prod (c\lambda B_i)^{d_i}.$$

Maximum likelihood equations are easily formed and the maximization is carried out using a Newton–Raphson iterative technique. For the data in Table 1, there are ties present and these were broken at random in order to apply the results above.

Table 2 summarizes the estimation of β for various c values and gives also the results for the exponential regression model ($c \to \infty$) and for the proportional hazards model. The

TABLE 2.

Estimation of β

	Proportional hazards	$c=1$	$c=5$	$c=25$	$c=125$	$c=\infty$
$\hat{\beta}$	1·652	1·606	1·512	1·461	1·490	1·58
var $(\hat{\beta})$	0·148	0·134	0·124	0·130	0·146	0·158
$\hat{\beta}/\surd\{\text{var}(\hat{\beta})\}$	4·29	4·39	4·29	4·05	3·90	3·97

entries var$(\hat{\beta})$ give the asymptotic variance of $\hat{\beta}$ based on the entry in the inverse of the information matrix for β, λ. The estimation of β is very stable over the range $0<c<\infty$.

4. POSTERIOR DISTRIBUTION OF Λ

In this section, the posterior distribution of the underlying process Λ is obtained when a sample $(t_1, \mathbf{z}_1), \ldots, (t_n, \mathbf{z}_n)$ is obtained from the model (7). The prior distribution of Λ is the gamma process with parameters c and absolutely continuous Λ^* as before.

Consider again a partition of $[0,\infty)$ into disjoint intervals $[a_0=0, a_1), \ldots, [a_{k-1}, a_k=\infty)$ and suppose that the data are $(t_1, \mathbf{z}_1)$ so that $n=1$. The extension to other n values follows easily from the results for $n=1$. As before $r_j=-\log(1-q_j)$ where q_j is the hazard contribution of the jth interval ($j=1,\ldots,k$). Assume that $a_{i-1} \leqslant t_1 < a_i$ and $r_{i1}=-\log(1-q_{i1})$ and $r_{i2}=-\log(1-q_{i2})$, where q_{i1} and q_{i2} are the hazard contributions of the intervals $[a_{i-1}, t_1)$ and $[t_1, a_i)$ respectively. Then clearly, from (9),

$$P\{T_1 \geqslant t_1 \mid \mathbf{r}, r_{i1}, \mathbf{z}_1\} = \exp\{-(r_1+\ldots+r_{i-1}+r_{i1})\exp(\mathbf{z}_1\boldsymbol{\beta})\}, \tag{16}$$

and

$$P\{T_1 \geqslant t_1, r_j \in (r_{0j}, r_{0j}+dr_{0j}), j=1,\ldots,k \mid \mathbf{z}_1\} = h(t_1, r_{01}, \ldots, r_{0k} \mid \mathbf{z}_1) \prod dr_{0j}$$

is obtained by taking the product of (16) with the independent gamma probability elements for the prior $r_1, \ldots, r_{i-1}, r_{i1}, r_{i2}, r_{i+1}, \ldots, r_k$ and integrating over $r_{i1}+r_{i2}=r_{i0}$. The posterior density of $\mathbf{r}$ given $T_1=t_1$ is then

$$f(\mathbf{r} \mid T_1=t_1, \mathbf{z}_1) = \frac{(\partial/\partial t_1)\, h(t_1, r_1, \ldots, r_k \mid \mathbf{z}_1)}{c\lambda^*(t_1)\log(c/c_1)(c/c_1)^{c\Lambda^*(t_1)}},$$

where $c_1=c+\exp(\mathbf{z}_1\boldsymbol{\beta})$ and the denominator is the marginal density of T_1 from (13). There does not appear to be a simple closed form expression for this, but the probability laws can be simply characterized using moment generating functions. Evidently $h(t_1, r_1, \ldots, r_k, \mathbf{z}_1)$ is a product of k factors, the jth of which involves r_j only while t_1 is in only the ith such factor. It follows that $r_1, \ldots, r_k$ are *a posteriori* independent and that the density for r_j ($j \neq 1$) is gamma. Thus

$$M_{\mathbf{r}}(\theta \mid T_1=t_1) = E\{\exp(\theta_1 r_1+\ldots+\theta_k r_k) \mid T_1=t_1\} = \prod M_{r_j}(\theta_j \mid T_1=t_1),$$

where

$$M_{r_j}(\theta_j \mid T_1=t_1) = \{c_1/(c_1-\theta_j)\}^{\alpha_j-\alpha_{j-1}}, \quad j=1,\ldots,i-1$$

$$M_{r_j}(\theta_j \mid T_1=t_1) = \{c/(c-\theta_j)\}^{\alpha_j-\alpha_{j-1}} \quad j=i+1,\ldots,k.$$

For the ith interval, we evaluate

$$\frac{\partial}{\partial t_1}\int\int \exp\{\theta_i(r_{i1}+r_{i2})\}\exp\{-r_{i1}\exp(\mathbf{z}_1\boldsymbol{\beta})\}\, g_1(r_{i1})\, g_2(r_{i2})\, dr_{i1}\, dr_{i2},$$

where g_1 and g_2 are the independent prior gamma densities of r_{i1} and r_{i2}. The ratio of this to the marginal density of t_1 is proportional to the moment generating function of r_i,

$$\left(\frac{c_1}{c_1-\theta_i}\right)^{\alpha(t_1)-\alpha_{i-1}}\left(\frac{c}{c-\theta_i}\right)^{\alpha_i-\alpha(t_1)}\log\left(\frac{c_1-\theta_i}{c-\theta_i}\right)\Big/\log\left(\frac{c_1}{c}\right). \tag{17}$$

Here as before $\alpha_j = c\Lambda^*(a_j)$ $(j = 1,\ldots,k)$ and $\alpha(t_1) = c\Lambda^*(t_1)$. Thus r_i is distributed as the sum of three independent random variables X, Y, U where X and Y are gamma variables and $U \sim A(c_1, c)$ with density

$$u^{-1}\{\exp(-cu)-\exp(-c_1 u)\}/\log(c_1/c),$$

whose moment generating function is the last factor in (17). All finite dimensional distributions of the posterior process have now been obtained and characterization of the process is straightforward. This special case ($n = 1$) is covered in the next paragraph.

The generalization of these results to obtain the posterior distribution of Λ given $(t_1, \mathbf{z}_1), \ldots, (t_n, \mathbf{z}_n)$ is straightforward. Given $t_1 < t_2 < \ldots < t_n$, $\Lambda(t)$ is an independent increments process. At t_i the increment is

$$U_i \sim A(c+A_i, c+A_{i+1}), \quad i = 1, \ldots, n,$$

where $A_i = \sum_i^n \exp(z_j \beta)$ as before. Between $[t_{i-1}, t_i)$ increments occur as for the gamma process $\mathbf{G}(c\Lambda^*, c+A_i)$. This result is easily seen by inserting t_n first then $t_{n-1}, \ldots, t_1$. Insertion of t_i affects the process only at points $t \leqslant t_i$.

The estimation of the survivor function can be performed in several ways. For example, $E\{\Lambda(t)\,|\,\mathbf{t}, \mathbf{Z}, \boldsymbol{\beta}\}$ would yield an optimum estimate if losses were squared error loss in $\Lambda(t)$. If $t_1 < t_2 < \ldots < t_n$ and $t_{i-1} \leqslant t < t_i$, then the posterior distribution of $\Lambda(t)$ is that of the sum of independent variables $X_1 + U_1 + \ldots + X_{i-1} + U_{i-1} + \delta_i$ where $X_j \sim G\{c\Lambda^*(t_{j-1}) - c\Lambda^*(t_j), c+A_j\}$, $U_j \sim A(c+A_j, c+A_{j+1})$ $(j = 1, \ldots, i-1)$ and $\delta_i \sim G(c\Lambda^*(t) - c\Lambda^*(t_{i-1}), c+A_i)$. Now easy calculation shows that

$$E(U_j) = \frac{\exp(\mathbf{z}_j \boldsymbol{\beta})}{(c+A_j)(c+A_{j+1})}\Big/\log\left(\frac{c+A_j}{c+A_{j+1}}\right)$$

and

$$E(\Lambda(t)\,|\,\mathbf{t}, \mathbf{Z}, \beta\} = \sum_1^{i-1}\{E(X_j) + E(U_j)\} + E(\delta_i).$$

For c small, $E(U_j) \cong A_j^{-1}$ $(j = 1, \ldots, n-1)$ and $E(X_j) \cong E(\delta_i) \cong 0$ so

$$E(\Lambda(t)\,|\,\mathbf{t}, Z, \beta) \cong \sum_{j|t_j<t} A_j^{-1}. \tag{18}$$

Kalbfleisch and Prentice (1973) obtained the maximum likelihood estimate of $\Lambda(t)$ as

$$\begin{aligned}\hat{\Lambda}(t) &= -\sum_{j|t_j<t} \exp(\mathbf{z}_j \boldsymbol{\beta}) \log\{1 - \exp(\mathbf{z}_j \boldsymbol{\beta})/A_j\} \\ &\cong \sum_{j|t_j<t} A_j^{-1}.\end{aligned} \tag{19}$$

Again, the two results are similar.

For the familiar Kaplan Meier situation, $z_j = 0$ $(j = 1, \ldots, n)$ and $A_j = n-j+1$. The estimate (18) is a first-order approximation to the Kaplan Meier result

$$\sum_1^{i-1} \log\{1-(n-j)^{-1}\}$$

to which (19) reduces in this special case.

Censoring causes no additional complications. For a single censored observation the generating function of r_i is (17) except for the last factor. Thus, a censored observation does not insert the random jump U_i into the posterior distribution of Λ.

The estimates for the survivor function, though appearing complicated, are nonetheless simple to compute. They amount to a slight smoothing of the usual step function estimates though for $c<\infty$ jumps occurs at each of the failure points. The tabulations of Cornfield and Detre (1977) are characteristic of these estimates, though the functions they consider are different.

5. THE CASE OF MULTIPLICITIES

If ties are present in the data, the calculations become more difficult. The difficulty arises because with probability one, a realization of the gamma process is discrete and so ties occur with non-zero probability. In this section, appropriate modifications of the previous results are derived to cover the case of ties.

Consider first the posterior of Λ. We begin by examining the simple case $T_1 = T_2 = t_1$ and from this the extensions are easily seen. From Section 4, conditional on $T_1 = t_1$, $\Lambda(t)$ is an independent increments process. At t_1 the increment is $U \sim A\{c+\exp(\mathbf{z}_1, \boldsymbol{\beta}), c\}$ and up to t_1 increments occur as for the gamma process $G\{c\Lambda^*, c+\exp(\mathbf{z}_1 \boldsymbol{\beta})\}$. Let $R = \Lambda(t_1-0)$. It is immediate that

$$P(T_2 = t_1 | R, U, T_1 = t_1) = \exp\{-R\exp(\mathbf{z}_2 \boldsymbol{\beta})\}[1-\exp\{-U\exp(\mathbf{z}_2 \boldsymbol{\beta})\}]. \tag{20}$$

The distribution of R and U (given $T_1 = t_1$) are known and a simple application of Bayes theorem shows that given $T_1 = T_2 = t_1$, the increment in $\Lambda(t)$ at t_1 is U with density

$$f(u|T_1 = T_2 = t_1) \propto (\exp(-cu) - \exp[-\{c+\exp(\mathbf{z}_1 \boldsymbol{\beta})\}u] - \exp[-\{c+\exp(\mathbf{z}_2 \boldsymbol{\beta})\}u] \\ + \exp[-\{c+\exp(\mathbf{z}_1 \boldsymbol{\beta})+\exp(\mathbf{z}_2 \boldsymbol{\beta})\}u])/u.$$

Up to time t_1, increments in $\Lambda(t)$ occur as for the gamma process

$$G\{c\Lambda^*, c+\exp(\mathbf{z}_1 \boldsymbol{\beta})+\exp(\mathbf{z}_2 \boldsymbol{\beta})\}.$$

This argument is informal, but can be formalized by considering partitions as in Section 4. The appropriate formula for ties now becomes apparent. Suppose the data are

$$T_{11} = \ldots = T_{1m1} = t_1 < T_{21} = \ldots = T_{2m2} = t_2 < \ldots < T_{k1} = \ldots T_{kmk} = t_k.$$

In the interval (t_{i-1}, t_i), the increments in $\Lambda(t)$ are those of the gamma process $G(c\Lambda^*, c+A_i)$ $(i = 1, \ldots, k+1)$ where $t_0 = 0$, $t_{k+1} = \infty$. At t_i the increment is U_i with density proportional to

$$u^{-1}\sum_{j=0}^{m_i}(-1)^j \sum_{p_j \in P} \exp[-\{c+A_{i+1}+\sum_{l \in p_j}\exp(\mathbf{z}_{il} \boldsymbol{\beta})\}u], \tag{21}$$

where P_j is the class of all subsets of j items chosen from $1, 2, \ldots, m_i$, $A_{k+1} = 0$ and A_i is defined in (11).

The result (21) would suggest that some inclusion exclusion argument should be available but no such argument has been found.

For the estimation of $\boldsymbol{\beta}$; we consider again a simple case where $T_1 = T_2 = \ldots = T_k = t_1$ is observed and then generalize to other situations. Consider first that $T_1 = t_1$ has been observed; integration of (20) with respect to the conditional distributions of U and R given $T_1 = t_1$ gives $P(T_2 = t_1 | T_1 = t_1)$. The (integrated) likelihood of $\boldsymbol{\beta}$ given $T_1 = T_2 = t_1$ is thus obtained as the product of this with the marginal density of T_1 at t_1 (from (13)). This gives

$$\left\{\frac{c}{c+\exp(\mathbf{z}_1 \boldsymbol{\beta})+\exp(\mathbf{z}_2 \boldsymbol{\beta})}\right\}^{c\Lambda^*(t)} [\log(c) - \log\{c+\exp(\mathbf{z}_1 \boldsymbol{\beta})\} - \log\{c+\exp(\mathbf{z}_2 \boldsymbol{\beta})\} \\ + \log\{c+\exp(\mathbf{z}_1 \boldsymbol{\beta})+\exp(\mathbf{z}_2 \boldsymbol{\beta})\}]. \tag{22}$$

By an inductive argument, the contribution of $T_1 = \ldots = T_k = t_1$ is easily obtained. The likelihood of $\boldsymbol{\beta}$ on the data giving rise to (21) is then

$$\prod_{i=1}^{k}\left\{\frac{c+A_{i+1}}{c+A_i}\right\}^{c\Lambda^*(t_i)}\sum_{j=0}^{m_i}(-1)^j\sum_{p_j\in P_j}\exp\left[-\left\{c+A_{i+1}+\sum_{l\in p_j}\exp(\mathbf{z}_{il}\boldsymbol{\beta})\right\}u\right]. \tag{23}$$

Both (21) and (23) are easily adjusted to accommodate censoring as for the case of no ties.

6. Discussion

The case of ties in the data is generally awkward both for the estimation of $\boldsymbol{\beta}$ and for the estimation of the survivor function. If many ties are present, it is probably best to use a discrete model for the data where there would be present only a finite number of scalar parameters. Standard Bayesian techniques could then be applied. The specialization of the above arguments to the discrete case would require specification of independent gamma prior distributions for each of the discrete hazard components. Kalbfleisch and MacKay (1978) consider the discrete case and show that the results obtained here also arise as a limit of the discrete analysis.

One deficiency of the gamma hazard process or the Dirichlet process is that with probability 1 the resulting Λ is a discrete survivor function and further the prior parameter c indexes in both cases the degree of discreteness in that for $c \to 0$, the probability that two random observations chosen from Λ are equal tends to one. As a consequence, the increase in variance of $\Lambda(t)$ for c near zero is to a large extent accounted for by the fact that Λ shall (with high probability) exhibit a single very large jump at some random point.

The above approach can also be criticized on the grounds that the hazard contributions have independent priors, a criticism which applies both to the Dirichlet process and to the gamma process. This would typically not be a reflection of prior opinion about the hazard function. Considering mixtures of gamma processes would introduce correlation between hazard contributions without unduly complicating the analysis. Such mixtures also would allow separate parameters measuring variation and discreteness.

Acknowledgements

I would like to thank Professor R. J. MacKay and Mr. C. J. Wild for many helpful comments and discussions on this work. My thanks also go to the referees whose constructive criticisms and queries led me to improve the exposition over earlier drafts. This work was supported by a grant from the National Research Council of Canada.

References

CORNFIELD, J. and DETRE, K. (1977). Bayesian life table analysis. *J. R. Statist. Soc.* B, **39**, 264–296.
COX, D. R. (1972). Regression models and life tables. *J. R. Statist. Soc.* B, **34**, 187–220.
—— (1975). Partial likelihood. *Biometrika*, **62**, 269–276.
—— (1976). A note on partially Bayes inference and the linear model. *Biometrika*, **63**, 651–654.
DOKSUM, K. A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Prob.*, **2**, 183–201.
FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, **1**, 209–230.
—— (1974). Prior distributions on spaces of probability measures. *Ann. Statist.*, **2**, 615–629.
GEHAN, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily single-censored samples. *Biometrika*, **52**, 203–224.
KALBFLEISCH, J. D. and MACKAY, R. J. (1978). Remarks on a paper by Cornfield and Detre. *J. R. Statist. Soc.* B, **40**, 175–177.
KALBFLEISCH, J. D. and PRENTICE, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika*, **60**, 267–278.
KINGMAN, J. F. C. (1975). Random discrete distributions (with discussion). *J. R. Statist. Soc.* B, **37**, 1–22.
MORAN, P. A. P. (1959). *The Theory of Storage*. London: Methuen.
SUSARLA, V. and VAN RYZIN, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *J. Amer. Stat. Ass.*, **71**, 896–902.