Rank-Based Inference in the Proportional Hazards Model for Interval Censored Data

Author(s): Glen A. Satten

Source: *Biometrika*, Jun., 1996, Vol. 83, No. 2 (Jun., 1996), pp. 355-370

Published by: Oxford University Press on behalf of Biometrika Trust

Stable URL: https://www.jstor.org/stable/2337606

Biometrika (1996), **83**, 2, *pp.* 355–370
Printed in Great Britain

Rank-based inference in the proportional hazards model for interval censored data

BY GLEN A. SATTEN

Division of HIV/AIDS Prevention, National Center for HIV, STD and TB Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia 30333, U.S.A.

SUMMARY

A marginal likelihood approach to fitting the proportional hazards model to interval censored or grouped data is proposed; this approach maximises a likelihood that is the sum over all rankings of the data that are consistent with the observed censoring intervals. As in the usual proportional hazards model, the method does not require specification of the baseline hazard function. The score equations determining the maximum marginal likelihood estimator can be written as the expected value of the score of the usual proportional hazards model, with respect to a certain distribution of rankings. A Gibbs sampling scheme is given to generate rankings from this distribution, and stochastic approximation is used to solve the score equations. Simulation results under various censoring schemes give point estimates that are close to estimates obtained using actual failure times.

Some key words: Cox model; Current status data; Gibbs sampling; Marginal likelihood; Regression; Stochastic approximation; Survival analysis.

1. INTRODUCTION

In studies involving time-to-failure data, whether or not an individual has failed is often determined only at specific monitoring times. If monitoring occurs sufficiently frequently, we may be able to use the time when the failure is first detected in place of the time when the failure occurred. In general, if the time between successive determinations of failure status is ‘small’ compared with the time to failure, such an approximation is reasonable. Ties among the failure times occur as the inter-monitoring interval increases, resulting in grouped data. Finally, if the inter-monitoring interval is highly variable between individuals, we may only be able to specify that a failure has occurred since the last time an individual was observed: in this case, we say that the data are interval censored. Interest in interval-censored data has increased recently because applications involving HIV/AIDS often involve interval censoring. Since the time between infection with HIV and development of clinical disease is long, with a median of approximately 10 years, patients are often monitored approximately every 6 months; see e.g. Brookmeyer & Gail (1994).

When data are not interval censored, the proportional hazards or Cox model is a useful way to assess the effect of covariates on survival (Cox, 1972). Its principal appeal is that it is semiparametric; although a specific functional form for the effects of covariates is assumed, the hazard in the baseline group need not be specified because only the rank order of failure and censoring times is required. When the data are grouped two approaches are commonly used, each of which preserves the semiparametric nature of the proportional hazards model. The partial likelihood approach assumes that failures can only occur at discrete times so that ties may occur; this model is not considered further here. The marginal likelihood approach (Peto, 1972; Kalbfleisch & Prentice, 1972, 1973, 1980; Pettitt 1983) assumes that a true vector of failure times would not have any ties. The marginal likelihood, obtained by summing over all rankings that are consistent with the observed pattern of ties, is maximised to determine parameters. Some recent progress on fitting the marginal likelihood has been made for the case of grouped data (DeLong, Guirguis & So, 1994; Sinha, Tanner & Hall, 1994).

When data are interval censored, we may also want to fit a proportional hazards model. In the general case, however, current approaches to fitting the proportional hazards model require estimation of the baseline hazard (Finkelstein, 1986; Diamond, McDonald & Shah, 1986), thus diminishing the appeal of using the proportional hazards model. In addition, these models do not reduce to the 'usual' proportional hazards model as the censoring intervals shrink. In this paper, we show how the marginal likelihood approach can be extended to interval censored data, and develop numerical methods that allow parameters in the proportional hazards model to be calculated without estimating the baseline hazard. Although this approach is computationally intense, we have been able to carry out many hundreds of simulations in data sets with 200 observations, with good accuracy.

The approach taken here is to construct a sequence of approximants that converges to the maximum marginal likelihood estimator of the parameters in the proportional hazards model. We accomplish this by stochastic approximation using the Robbins–Monro process, in a manner similar to that of Ruppert et al. (1984). The Robbins–Monro process is a stochastic process that converges to the root of a function $f(x)$, when values of $f(x)$ can only be observed with error. Ruppert (1991) provides a good general introduction to stochastic approximation; a discussion of the use of stochastic approximation as a method of numerical optimisation can be found in Ruppert et al. (1984). In § 2, the marginal likelihood for interval-censored data is defined, and the score equations and information matrix are calculated. In § 3·1 we establish the connection between the Robbins–Monro process and the score equations of the marginal likelihood by expressing the score as the expected value of the score for the proportional hazards model for known failure times, with respect to a certain distribution. Section 3·2 shows how random samples from this distribution can be generated using a Gibbs sampling scheme; see e.g. Tanner (1993, pp. 102–46). Section 3·3 considers the details of implementing the stochastic approximation scheme, with special emphasis on making efficient use of the Gibbs sampler when using stochastic approximation. Section 3·4 considers the total variance of the stochastic approximant to the true value of the parameters in the proportional hazards model. Section 4 presents simulation results assessing the performance of the maximum marginal likelihood estimator under both 'light' and 'heavy' censoring and for moderate and large hazards ratios. Section 5·1 summarises our recommendations for how to conduct an analysis using our methodology, and § 5·2 contains a general conclusion.

2. MODEL

For each individual i, suppose that we measure $d_i := (l_i, u_i, x_i)$, where we know that an event occurred in the interval $(l_i, u_i]$, and where x_i is a vector of explanatory variables. As special cases, traditional right censoring corresponds to $u_i = \infty$, while 'exact' knowledge of the time of event corresponds to $l_i = u_i - \varepsilon$, for arbitrarily small ε. If we had observed the exact times of events, we would be able to rank the observations d_i in order of their occurrence. Instead, we may consider the set of all possible such rankings of the ordered observations d_i, which we will denote by $\mathscr{R}$. That is, every element $R := (r_1, r_2, \ldots, r_N)$ of $\mathscr{R}$ is a permutation of the integers from 1 to N, such that it is consistent with the observed intervals $(l_i, u_i]$ to assume that observation d_1 occurred r_1th, observation d_2 occurred r_2th, etc. For example, when none of the intervals $(l_i, u_i]$ overlap, then $\mathscr{R}$ consists of a single element. In a second example, if $u_1 < l_i$, $i \geqslant 2$, then $r_1 = 1$ for every element in $\mathscr{R}$.

When event times are known with certainty, the proportional hazards model utilises either the marginal or partial likelihoods of the observed rank order of the data rather than the full likelihood of the observed failure times. The reduction of data to ranks is useful because it allows for a semiparametric model for failure-time distributions in the proportional hazards family. Similarly, in the interval-censored case, we must choose what we wish to consider as 'data' when constructing a likelihood. We consider the model in which the set of possible rankings, $\mathscr{R}$, comprises the data. In this case, the likelihood $\mathscr{L}$ is the marginal likelihood, given by

$$\mathscr{L} = \sum_{R \in \mathscr{R}} \operatorname{pr}(R \mid \beta, \{x_i\}), \tag{1}$$

where $\operatorname{pr}(R \mid \beta, \{x_i\})$ is the probability of the ranking R in the standard proportional hazards model, given the vector of regression parameters β and the set $\{x_i\}$ of explanatory variables for all of the observations.

Taking the gradient of $\ln(\mathscr{L})$ with respect to β, we obtain the score function $S_\beta(\mathscr{R})$, given by

$$S_\beta(\mathscr{R}) := \sum_{R \in \mathscr{R}} w_\beta(R) S_\beta(R \mid \{x_i\}), \tag{2}$$

where $S_\beta(R \mid \{x_i\})$ is the score function for the observed data calculated using the standard proportional hazards model assuming ranking R, and where weights $w_\beta(R)$ are given by

$$w_\beta(R) = \frac{\operatorname{pr}(R \mid \beta, \{x_i\})}{\sum_{R \in \mathscr{R}} \operatorname{pr}(R \mid \beta, \{x_i\})}. \tag{3}$$

For simplicity of notation, we have suppressed the dependence of w_β on explanatory variables $\{x_i\}$.

Maximum marginal likelihood estimates of β, denoted by $\hat{\beta}$, may be obtained as usual by solving $S_\beta(\mathscr{R}) = 0$; although our results do not depend on this, we assume without proof that these estimates are asymptotically normal and that the variance–covariance matrix is given by the inverse of the information matrix F, where

$$F_\beta(\mathscr{R}) = \sum_{R \in \mathscr{R}} w_\beta(R)[F_\beta(R \mid \beta, \{x_i\}) - \{S_\beta(R \mid \beta, \{x_i\}) - S_\beta(\mathscr{R})\}\{S_\beta(R \mid \beta, \{x_i\}) - S_\beta(\mathscr{R})\}^{\mathrm{T}}], \tag{4}$$

and where $F_\beta(R \mid \beta, \{x_i\})$ is the information matrix for the observed data calculated using the standard proportional hazards model assuming ranking R. Equations similar to (2)–(4) were derived by Self & Grossman (1986) when considering linear rank tests for interval censored data in the accelerated failure time model, and are implicit in Sinha, Tanner & Hall (1994) for the special case of grouped data.

3. Solving the score equations using stochastic approximation

3·1. *General set-up*

Without a way to calculate maximum likelihood values $\hat{\beta}$, the model of § 2 has limited value. In this section, we show how stochastic approximation may be used to construct a stochastic process that will converge to the maximum likelihood values $\hat{\beta}$. Stochastic approximation is a method for finding the root of a function, whose values can only be observed with error, by constructing a stochastic process that converges to the root of the function almost surely. The simplest such stochastic process is the Robbins–Monro process; see e.g. Ruppert (1991). We will use a simple variant of the Robbins–Monro process used by Ruppert et al. (1984). In our case, we seek the root of the score given in equation (3), or equivalently in equation (5) below.

Note that the weights $w_\beta(R)$ defined in equation (3) form a probability distribution on the set $\mathscr{R}$, so that the score S may be considered an expected value of the score $S_\beta(R \mid \beta, \{x_i\})$ for the proportional hazards model with known ranking R, taken with respect to the distribution w_β; that is

$$S_\beta(\mathscr{R}) = E\{S_\beta(R \mid \{x_i\}) \mid w_\beta(R)\}. \tag{5}$$

Similarly,

$$F_\beta(\mathscr{R}) = \mathscr{F}_\beta(\mathscr{R}) - \Sigma_\beta(\mathscr{R}), \tag{6}$$

where

$$\mathscr{F}_\beta(\mathscr{R}) = E\{F_\beta(R \mid \{x_i\}) \mid w_\beta(R)\},$$

$$\Sigma_\beta(\mathscr{R}) = E\{S_\beta(R \mid \{x_i\}) S_\beta(R \mid \{x_i\})^{\mathrm{T}} \mid w_\beta(R)\}.$$

As a consequence of (5), if we select a random element R of $\mathscr{R}$, with probability $w_\beta(R)$, then $S_\beta(R \mid \beta, \{x_i\})$ is an unbiased estimator of $S_\beta(\mathscr{R})$. As a result, if we can generate rankings R with distribution $w_\beta(R)$, we may use the Robbins–Monro process to estimate the root of $S_\beta(\mathscr{R})$, which is the maximum marginal likelihood estimator $\hat{\beta}$, by constructing a stochastic process which converges almost surely to $\hat{\beta}$.

3·2. *Generating rankings with distribution $w_\beta(R)$*

Although characterisation of $\mathscr{R}$ is itself a difficult problem, by using special properties of the proportional hazards model we have developed simple, easily implemented Gibbs sampling scheme to obtain random samples from $\mathscr{R}$ with probabilities $w_\beta(R)$. A very useful characteristic of the proportional hazards family of distributions is that the probability of obtaining a certain ranking R of the data is independent of the baseline failure-time distribution. Hence, a convenient way to generate rankings is to choose a distribution for the failure times which is in the proportional hazards family, generate failure times using this distribution, and then rank the failures. The exponential distribution with unit hazard for the baseline distribution is the easiest choice. We stress here that this choice does not imply that we are assuming that the underlying distribution is exponential, nor are we using the exponential distribution to impute the missing failure times within the observed intervals; instead, the failure times generated are simply a mathematical construction to allow rankings to be generated.

Let t be a vector with components t_i that are exponentially distributed random variables with exponential distribution parameter $\lambda_i = \exp(\beta x_i)$. In the absence of the order restrictions imposed by $\mathscr{R}$, we could simply rank the components of t and be assured that the probability of observing a ranking R was given by $\mathrm{pr}(R \mid \beta, \{x_i\})$. However, to generate rankings from $\mathscr{R}$ with probabilities $w_\beta(R)$, we must restrict our samplings to rankings in $\mathscr{R}$. Although $\mathscr{R}$ is difficult to characterise in terms of rankings, sets of times t that lead to rankings in $\mathscr{R}$ are easily characterised. Specifically, let b_i denote the set containing all indices of observations that must occur before the ith observation. Then

$$b_i = \{j \mid u_j \leqslant l_i, j = 1, \ldots, N, j \neq i\}; \tag{7}$$

note that, even if $u_j = l_i$, observation j must occur before observation i because we are considering intervals of the form $(l_i, u_i]$. Similarly, let a_i denote the set containing all indices of observations that must occur after the ith observation, so that

$$a_i = \{j \mid u_i \leqslant l_j, j = 1, \ldots, N, j \neq i\}. \tag{8}$$

The times t that are consistent with $\mathscr{R}$ are those satisfying

$$t_i > t_j \quad (i = 1, \ldots, N, j \in b_i), \tag{9}$$

$$t_i < t_j \quad (i = 1, \ldots, N, j \in a_i). \tag{10}$$

A random sample of times t, in which each component is exponentially distributed with parameter $\lambda_i = \exp(\beta x_i)$, satisfying constraints (9) and (10), when ranked, will be a random sample from $\mathscr{R}$ with distribution $w_\beta(R)$.

Random samples of the type described above may be obtained from a Gibbs sampling scheme. Let $T_i^+(t)$ be defined by

$$T_i^+(t) = \begin{cases} \min(t_j, j \in a_i) & \text{if } a_i \neq \varnothing, \\ \infty & \text{otherwise.} \end{cases}$$

Similarly, let

$$T_i^-(t) = \begin{cases} \max(t_j, j \in b_i) & \text{if } b_i \neq \varnothing, \\ 0 & \text{otherwise.} \end{cases}$$

Then $T_i^\pm$ form upper and lower bounds for t_i in the sense that (9) and (10) are equivalent to

$$T_i^-(t) < t_i < T_i^+(t).$$

Hence, the distribution of t_i, conditional on all other t_j $(j \neq i)$ fixed, and subject to constraints (9) and (10), is the truncated exponential distribution

$$\operatorname{pr}(t_i = s \mid \{t_j, j \neq i\}) = \frac{\lambda_i e^{-\lambda_i s}}{e^{-\lambda_i T_i^-(t)} - e^{-\lambda_i T_i^+(t)}} \quad (T_i^-(t) < s < T_i^+(t)).$$

These conditional distributions can be used in a Gibbs sampling scheme to generate random variates t, which can in turn be ranked to obtain samples from $w_\beta(R)$. Finally, updating the values sequentially, either in ascending or descending order, with probability 0·5, ensures the time-reversal symmetry of the Gibbs sampler (Besag, 1986; Geyer, 1992).

Starting values for t can be obtained from the data by defining $y_i := l_i + \alpha_i(u_i - l_i)$, where α_i is a number between 0 and 1. Although the y_i's satisfy the constraints (9)–(10), they may be an unlikely outcome from the set of exponential distributions considered above. However, this is easily remedied by scaling the y_i's to obtain starting values t_i' given by

$$t_i' := \frac{y_i}{\sum_{i'=1}^{N} y_{i'} e^{\beta x_{i'}}}. \tag{11}$$

3·3. *Stochastic approximation scheme*

Once we have available rankings R independently selected from $\mathscr{R}$ with probability $w_\beta(R)$, we can implement a stochastic approximation scheme. A slight modification of the scheme described by Ruppert et al. (1984) is appropriate for our case. Briefly, if we denote by β_k the kth approximant to $\hat{\beta}$, and if R_k is a ranking in $\mathscr{R}$ selected using distribution $w_\beta(R)$, then the $(k+1)$th approximant to $\hat{\beta}$ is given by

$$\beta_{k+1} = \beta_k + \frac{1}{k}(\mathscr{F}_{k+1} - \Sigma_{k+1})^{-1} S(R_k | \beta_k, \{x_i\}) \quad (k \geqslant 1); \tag{12}$$

similarly, $\mathscr{F}$ and Σ are updated using

$$\mathscr{F}_{k+1} = \mathscr{F}_k + \frac{1}{k}\{F(R_k | \beta_k, \{x_i\}) - \mathscr{F}_k\} \quad (k \geqslant 1), \tag{13}$$

$$\Sigma_{k+1} = \Sigma_k + \frac{1}{k}\{S(R_k | \beta_k, \{x_i\}) S^{\mathrm{T}}(R_k | \beta_k, \{x_i\}) - \Sigma_k\} \quad (k \geqslant 1). \tag{14}$$

Those not familiar with stochastic approximation may consider (12) to be a variant of a Newton–Raphson iteration, except that convergence must be forced by taking ever smaller steps.

Although the starting value β_1 in (12) is arbitrary in theory, in practice a good starting value β_1 is important, and is discussed below. Note that $\mathscr{F}_1$ and Σ_1 do not need to be specified to begin the iterations. In defining Σ, we have used the fact that the expected value of S at $\hat{\beta}$ is zero, so that the covariance of S is the expected value of SS^{T}. The stochastic approximation scheme is run until termination at $k = k^*$; we will discuss the choice of k^* later. After termination, the most current values of β, $\mathscr{F}$ and Σ, denoted β^*, $\mathscr{F}^*$ and Σ^*, are used as estimates of $\hat{\beta}$, $\mathscr{F}_{\hat{\beta}}(\mathscr{R}) =: \hat{\mathscr{F}}$ and $\Sigma_{\hat{\beta}}(\mathscr{R}) =: \hat{\Sigma}$, respectively. We depart from Ruppert et al. (1984) only by using the updated values of $\mathscr{F}$ and Σ when determining the next value of β. Note that, although we have written recursive expressions for $\mathscr{F}_k$ and Σ_k, $\mathscr{F}_{k+1}$ can also be expressed as is the arithmetic average of $F(R_{k'} | \beta_{k'}, \{x_i\})$ for $k' = 1, \ldots, k$; a similar result holds for Σ_{k+1}.

Standard asymptotic results for stochastic approximation cited in Ruppert et al. (1984) show that, conditional on the observed data,

$$(k^*)^{\frac{1}{2}}(\beta^* - \hat{\beta}) \to N(0, \hat{F}^{-1}\hat{\Sigma}\hat{F}^{-1}) \tag{15}$$

in distribution as $k^* \to \infty$, where $\hat{F} := \hat{\mathscr{F}} - \hat{\Sigma}$ is the information matrix at $\hat{\beta}$.

The stochastic approximation scheme described above is inefficient if independent rankings R_k must be obtained from a Gibbs sampling scheme. Assuming that we plan to use a single long run from our Gibbs sampler, the inefficiency is a result of the requirement that we must run the Gibbs sampler long enough to ensure that R_k is independent of R_{k-1} and that the equilibrium distribution characterised by parameter value β_k has set in; hence many unused sets of times and rankings are generated. However, a simple modification of the stochastic approximation scheme allows us to use a single Gibbs stream and in many cases to use each iterate generated. Instead of obtaining a single ranking R_k for each distinct value of β_k, we propose taking $M_0 + M$ rankings for each β_k; the significance of M_0 and M is explained below. We denote these rankings by R_{km} $(m = 1, \ldots, M_0 + M)$. Associated with each ranking is a vector of times, which we denote by $t^{(k,m)}$. Hence, starting with $\beta = \beta_1$, the Gibbs sampler is used to generate times $t^{(1,1)}, \ldots, t^{(1,M_0+M)}$; after β is updated to β_2 the Gibbs stream is continued by restarting it with initial value $t^{(1,M_0+M)}$ to generate $t^{(2,1)}, \ldots, t^{(2,M_0+M)}$; β is updated to β_3 and so on until k^*. Define

$$\bar{S}_k := \frac{1}{M} \sum_{m=M_0+1}^{M_0+M} S(R_{km} \mid \beta_k, \{x_i\}). \tag{16}$$

Similarly, let

$$\bar{\mathscr{F}}_k := \frac{1}{M} \sum_{m=M_0+1}^{M_0+M} F(R_{km} \mid \beta_k, \{x_i\}), \tag{17}$$

$$\bar{\Sigma}_k := \frac{1}{M} \sum_{m=M_0+1}^{M_0+M} S(R_{km} \mid \beta_k, \{x_i\}) S(R_{km} \mid \beta_k, \{x_i\})^{\mathrm{T}}. \tag{18}$$

Because only M of the $M_0 + M$ are used to construct $\bar{S}_k$, $\bar{\mathscr{F}}_k$ and $\bar{\Sigma}_k$, we will refer to M_0 as the number of blanks, while we will call M the block size. If we let the block size M increase, the $\bar{S}_k$'s are asymptotically independent, regardless of the choice of M_0; additionally, the functional central limit theorem for reversible Markov chains, see Kipnis & Varadhan (1986) and also Geyer (1992), ensures that they are asymptotically unbiased estimators of $S_{\beta_k}(\mathscr{R})$. Hence, the stochastic approximation scheme (12)–(14) can be used, except that $\bar{S}_k$ and $\bar{\Sigma}_k$ are used instead of S_k and Σ_k. Because M is finite in actual use, a nonzero value of M_0 will help achieve practical independence of the $\bar{S}_k$'s for smaller values of M in cases of strong correlation.

We consider first the case where $M_0 = 0$, where $T := Mk^*$ is the total number of rankings generated. If the rankings R_{km} were independent, the variance of $\bar{S}_k$ would be Σ/M; hence, equation (15) shows that, under independence, the asymptotic variance of β^* depends on M and k^* only through their product T. In particular, no asymptotic loss of efficiency occurs by taking multiple rankings at each β_k but fewer stochastic approximation steps. Note also that $\bar{S}_k$, $\bar{\Sigma}_k$ and $\bar{\mathscr{F}}_k$ can each be recursively calculated, so that the rankings R_{km} do not need to be stored; thus, the recursive character of stochastic approximation is not lost. However, because the rankings R_{km} are dependent, the variance of $\bar{S}_k$ will typically be larger than if the R_{km}'s were independent. In the examples we consider in § 4, this effect ranges from negligible to fairly prominent.

If M and M_0 are chosen such that the $\bar{S}_k$'s are nearly independent, the variance of $\bar{S}_k$, denoted by V, may be estimated by V_k, where

$$V_{k+1} = V_k + \frac{1}{k}(\bar{S}_k \bar{S}_k^{\mathrm{T}} - V_k) \quad (k \geqslant 1). \tag{19}$$

In writing (19) we have used the fact that the expected value of $\bar{S}_k$ is zero; as for $\mathscr{F}$ and Σ, the value of V_1 does not need to be specified. Then, the asymptotic variance of β_k is given by (15), with Σ replaced by V. An additional saving in time required for computation when using (16)–(18) is obtained in higher dimensions, because the matrix $(\bar{\mathscr{F}}_k - \bar{\Sigma}_k)$ need only be inverted every M steps. Finally, for cases with strong correlation, we may use $M_0 > 0$ to assist in achieving independence of the $\bar{S}_k$'s. Note that, even if we choose $M_0 = M$, we are still using half of the Gibbs iterates generated, a great gain in efficiency over using (12)–(13) while trying to ensure that each $S(R_k \mid \beta_k, \{x_i\})$ is independent.

As $\mathscr{F} - \Sigma$ may fail to be positive definite if β_1 is far from $\hat{\beta}$, we use a short 'burn-in' period to obtain a starting value. We begin with an arbitrary value of β, usually $\beta = 0$, and initial values t' as given in (11), and run the Gibbs sampler 50 times to obtain an initial sample $t^{(0)}$ from the constrained distribution. Then we run the stochastic approximation scheme (12)–(13) with $\Sigma_k \equiv 0$ for n_b steps, taking only one ranking at each value of β_k, corresponding to $M = 1$, and with no concern for the independence of the R_k's. The value of β_{n_b} is then taken as the starting value β_1. The Gibbs sampler is begun using the values of t that were obtained at the end of the burn-up period. In the simulations in § 4, we used $n_b = 100$ for simulations with a hazard ratio of $\lambda = 2$, and $n_b = 500$ for simulations with $\lambda = 10$. If convergence of (12) is still problematic due to a poor initial step, Ruppert et al. (1984) have suggested replacing $1/k$ by $1/(k + k_0)$ for $k_0 > 0$ to avoid spoiling a good starting value.

3·4. *Unconditional confidence intervals*

If the stochastic approximation scheme is terminated at a finite value of k^*, then the final value β^* differs from $\hat{\beta}$ by an error that is approximately normally distributed, with asymptotic variance given by $(1/k^*)\hat{F}^{-1}\hat{V}\hat{F}^{-1}$, where by $\hat{V}$ we mean the variance of $\bar{S}_k$ when $\beta = \hat{\beta}$. As a result, if we use β^* as an estimator of β, the 'true' parameter, we can expect that the confidence interval for β should be increased to reflect this additional variability. A standard decomposition of the variance of β^* is

$$\text{var}(\beta^*) = E_{\{d_i\}}\{\text{var}(\beta^* \mid \{d_i\})\} + \text{var}_{\{d_i\}}\{E(\beta^* \mid \{d_i\})\}, \tag{20}$$

where $\{d_i\}$ denotes the observed data. Since the asymptotic expected value of β^* is $\hat{\beta}$, the second term in (20) is the usual large-sample variance estimator of $\hat{\beta}$. An asymptotically unbiased estimator of the first term is $(1/k^*)(F^*)^{-1}V^*(F^*)^{-1}$, since $\beta_{k^*} \rightarrow \hat{\beta}$ as $k^* \rightarrow \infty$ for fixed N, and $\hat{\beta} \rightarrow \beta$ as $N \rightarrow \infty$. Hence, the variance of β is inflated linearly by the additional variability introduced by estimating $\hat{\beta}$ by β^*.

4. SIMULATION RESULTS

4·1. *Performance of the method*

To assess the performance of the stochastic approximation scheme, as well as to develop confidence in the marginal likelihood approach to interval censored data, we conducted simulation studies on artificially generated data. Results were obtained for univariate analyses with a single binary covariate x. We considered six cases, corresponding to a moderate hazard ratio ($\beta = \log 2$) and a large hazard ratio ($\beta = \log 10$), in the presence of both 'light' censoring, 'heavy' censoring, or grouping. Data for all simulations had 200 observations, with 100 observations in each group.

One hundred data sets were created for each value of β. In each case, the failure-time distribution in the 'baseline' group was assumed to follow a Weibull distribution with scale parameter $a = 10^{-4}$ and shape parameter $b = 2$ so that the mean and variance were $50\pi^{\frac{1}{2}}$ and 7500π, respectively. The simulated failure times were then interval censored using two mechanisms. In the 'heavy' censoring case, a renewal process was begun at time 0 with log-normally distributed increments that had mean 20 and coefficient of variation $e^4 - 1$, and the endpoints of the interval containing the failure were recorded. In the 'light' censoring case, intervals were constructed around the 'true' failure time so that the interval containing the ith of the ordered failure times included only the $(i - 1)$th and $(i + 1)$th failure times. Grouped data were obtained by observing failure status for all individuals every 10 time units, resulting in an average of 23 failure time groups.

To assess the degree of censoring, we computed the average number of censoring intervals each censoring interval overlaps with. Specifically, if a data set has N observations, then the proportion of observed intervals overlapping with the ith observation, denoted by $\mathscr{P}_i$, is given by

$$\mathscr{P}_i := 1 - \frac{\#(a_i) + \#(b_i)}{N},$$

where $\#(.)$ denotes the number of elements in a set, and a_i and b_i are defined in (7) and (8). The mean $\bar{\mathscr{P}}$ of the $\mathscr{P}_i$'s is then the proportion of observations which a typical observation overlaps. The data sets for the 'heavy' censoring case had an average value of $\bar{\mathscr{P}}$ of approximately 0·66 for $\beta = \log 2$ and 0·62 for $\beta = \log 10$. In the 'light' censoring case, $\bar{\mathscr{P}} \simeq 3/200$. For the grouped data, $N\bar{\mathscr{P}}$ is the average number of failures in each group; for the data sets with $\beta = \log 2$ and $\beta = \log 10$, the average values of $N\bar{\mathscr{P}}$ were 8·7 and 9·0 respectively, although the maximum numbers of ties in a single group were substantially higher, 33 for $\beta = \log 2$ and 45 for $\beta = \log 10$.

For both 'light' censoring analyses, we used a block size of $M = 50$, with no blanks ($M_0 = 0$). For the 'heavy' censoring analyses, we used a block size of $M = 150$, with no blanks for $\beta = \log 2$, and $M_0 = 150$ for $\beta = \log 10$. For grouped data analyses, we used a block size of $M = 50$, with no blanks for $\beta = \log 2$, and $M_0 = 250$ for $\beta = \log 10$. In all analyses, the stochastic approximation scheme was run for $k^* = 400$ steps. How we chose the parameters of the stochastic approximation is discussed in § 4·2.

For each of the simulated data sets, we computed estimates of the hazard ratio using the 'true' failure time and using only the censoring interval. For the 'heavy' censoring case we also computed hazard ratio estimates using the midpoint of the censoring interval; in the 'light' censoring case, the midpoint of the interval is the 'true' failure time by construction. The results of these simulations are shown in Tables 1 and 2. The first lines compare estimates of the log hazard ratio β for the four analyses considered. The estimators of β using the true failure times are very close to those obtained under light censoring and grouping and only slightly different from those obtained under heavy censoring. The estimator obtained using the midpoints of the heavy censoring intervals is highly biased. The second lines of Tables 1 and 2 give a direct comparison of $\hat{\beta}_e$, the maximum likelihood estimator of the standard proportional hazards model using the exact failure times, with β^* for the interval censored and grouped cases, and with $\hat{\beta}_m$, the maximum likelihood estimator for the standard proportional hazards model using the midpoint data.

To determine the amount of variability in β^* that is accounted for by variability in $\hat{\beta}_e$, we computed the correlation between $\hat{\beta}_e$ and the values of β^* obtained for each censoring scheme. For $\beta = \log 2$, we found the correlation between $\hat{\beta}_e$ and β^* for 'light' censoring, grouped data, 'heavy censoring' and midpoint analyses to be 0·99998, 0·997, 0·83 and 0·53, respectively. For $\beta = \log 10$, the equivalent correlations were 0·999997, 0·986, 0·74 and 0·24.

The third lines of Tables 1 and 2 compare the observed information for the standard proportional hazards model with its expected value under the distribution $w_\beta(R)$. Even in the heavy censoring case, the value of $\mathscr{F}^*$ differs only slightly from the value of F_β obtained using the true failure times. In the $\beta = \log 10$ case, the value of F_β obtained using midpoint data is also severely biased.

The remainder of Tables 1 and 2 compare the information available in the three censoring cases. The small values of Σ^* relative to $\mathscr{F}^*$ in the light censoring and grouped cases indicate that very little information is lost by censoring given equation (6), while the relatively large value of Σ^* compared with $\mathscr{F}^*$ in the heavy censoring case results in an appreciable loss of information. This loss of information also manifests itself in the values of $(V^*/k^*)^{\frac{1}{2}}/(\mathscr{F}^* - \Sigma^*)$, which is the standard error of β^* around $\hat{\beta}$: although the block size in the heavy censoring case was three times as great as that in the light censoring

Table 1. *Summary of results from* 100 *simulated data sets with a single binary explanatory variable* x, 100 *observations at each level of* x, *and* $\beta = \log 2$

	Exact	Light censoring	Heavy censoring	Grouped	Midpoint
$\hat\beta$ or β^*	0·669 (0·283, 0·991)	0·669 (0·283, 0·991)	0·685 (0·191, 1·028)	0·672 (0·276, 0·999)	0·352 (0·055, 0·750)
$\hat\beta_e - \beta^*$ or $\hat\beta_e - \hat\beta_m$	—	$0{\cdot}12 \times 10^{-3}$ ($-9{\cdot}58 \times 10^{-3}$, $0{\cdot}63 \times 10^{-3}$)	−0·011 (−0·242, 0·248)	$1{\cdot}9 \times 10^{-3}$ (−0·0394, 0·0313)	0·315 (−0·072, 0·750)
F or $\mathscr{F}^*$	44·66 (39·65, 48·68)	44·65 (39·65, 48·66)	44·29 (38·47, 48·63)	44·63 (39·99, 48·66)	48·43 (43·98, 49·62)
Σ^*	—	0·075 (0·015, 0·498)	14·62 (10·93, 18·55)	0·33 (0·19, 0·84)	—
V^*	—	$2{\cdot}2 \times 10^{-3}$ ($0{\cdot}4 \times 10^{-3}$, $14{\cdot}0 \times 10^{-3}$)	0·141 (0·087, 0·262)	$7{\cdot}6 \times 10^{-3}$ ($4{\cdot}5 \times 10^{-3}$, $17{\cdot}8 \times 10^{-3}$)	—
$(V^*/k^*)^{\frac{1}{2}}(\mathscr{F}^* - \Sigma^*)^{-1}$	—	$4{\cdot}8 \times 10^{-5}$ ($2{\cdot}3 \times 10^{-5}$, $14{\cdot}0 \times 10^{-5}$)	$6{\cdot}3 \times 10^{-4}$ ($4{\cdot}4 \times 10^{-4}$, $10{\cdot}5 \times 10^{-4}$)	$9{\cdot}8 \times 10^{-5}$ ($7{\cdot}3 \times 10^{-5}$, $14{\cdot}8 \times 10^{-5}$)	—
MV^*/Σ^*	—	1·47 (1·19, 1·92)	1·44 (1·00, 2·45)	1·17 (1·01, 1·36)	—

Values shown are means of 100 simulations, unless otherwise indicated; ranges of 100 simulations shown in parentheses.

For $\hat\beta$ or β^*, results shown for exact and midpoint analyses are $\hat\beta_e$ and $\hat\beta_m$, respectively, the estimators from the standard proportional hazards model; for light and heavy censoring, results shown are values of β^*.

For $\hat\beta_e - \beta^*$ or $\hat\beta_e - \hat\beta_m$, results shown are the median of 100 simulations.

For F or $\mathscr{F}^*$, results shown for exact and midpoint analyses are F, the observed information matrix from the standard proportional hazards model; for light and heavy censoring, results shown are values of $\mathscr{F}^*$.

Table 2. *Summary of results from* 100 *simulated data sets with a single binary explanatory variable x,* 100 *observations at each level of x, and* $\beta = \log 10$

	Exact	Light censoring	Heavy censoring	Grouped	Midpoint
$\hat\beta$ or β^*	2·285 (1·797, 2·837)	2·286 (1·797, 2·838)	2·379 (1·961, 3·537)	2·290 (1·760, 2·856)	0·997 (0·597, 1·382)
$\hat\beta_e - \beta^*$ or $\hat\beta_e - \hat\beta_m$	—	$6{\cdot}3 \times 10^{-4}$ $(-21{\cdot}3 \times 10^{-4}, 11{\cdot}1 \times 10^{-4})$	−0·072 (−0·839, 0·298)	$5{\cdot}4 \times 10^{-3}$ (−0·0887, 0·0768)	1·303 (0·643, 1·964)
F or $\mathscr{F}^*$	21·80 (13·91, 29·16)	21·79 (13·89, 29·15)	20·78 (9·74, 26·33)	21·73 (13·94, 29·55)	45·44 (41·13, 48·18)
Σ^*	—	0·025 (0·009, 0·041)	8·68 (5·76, 12·47)	0·57 (0·17, 1·36)	—
V^*	—	$7{\cdot}6 \times 10^{-4}$ $(2{\cdot}8 \times 10^{-4}, 16{\cdot}8 \times 10^{-4})$	0·173 (0·068, 0·408)	0·021 (0·004, 0·081)	—
$(V^*/k^*)^{\frac{1}{2}}(\mathscr{F}^* - \Sigma^*)^{-1}$	—	$6{\cdot}4 \times 10^{-5}$ $(3{\cdot}7 \times 10^{-5}, 11{\cdot}2 \times 10^{-5})$	$1{\cdot}8 \times 10^{-3}$ $(0{\cdot}9 \times 10^{-3}, 6{\cdot}6 \times 10^{-3})$	$3{\cdot}4 \times 10^{-4}$ $(1{\cdot}5 \times 10^{-4}, 10{\cdot}7 \times 10^{-4})$	—
MV^*/Σ^*	—	1·54 (1·19, 2·06)	2·98 (1·27, 6·52)	1·73 (1·30, 3·16)	—

Values shown are means of 100 simulations, unless otherwise indicated; ranges of 100 simulations shown in parentheses.

For $\hat\beta$ or β^*, results shown for exact and midpoint analyses are $\hat\beta_e$ and $\hat\beta_m$, respectively, the estimators from the standard proportional hazards model; for light and heavy censoring, results shown are values of β^*.

For $\hat\beta_e - \beta^*$ or $\hat\beta_e - \hat\beta_m$, results shown are the median of 100 simulations.

For F or $\mathscr{F}^*$, results shown for exact and midpoint analyses are F, the observed information matrix from the standard proportional hazards model; for light and heavy censoring, results shown are values of $\mathscr{F}^*$.

case, the standard errors in the light censoring case are still at least an order of magnitude smaller. However, in all cases we were able to achieve fairly accurate estimates of $\hat{\beta}$ even while working with 100 data sets. In the worst case, corresponding to the $\beta = \log 10$ 'heavy' censoring simulations, there is some uncertainty in the second decimal place.

The final lines of Tables 1 and 2 relate to the loss of efficiency caused by using correlated values of $S(R_{km} | \beta, \{x_i\})$ to calculate the $\bar{S}_k$'s. Under independence, we could use Σ^*/M in place of V^* in calculating the variance of β^* around $\hat{\beta}$. Thus, the ratio MV^*/Σ^* measures the loss of efficiency in this variance caused by inflation of V^* by the correlation induced by the Gibbs sampler: MV^*/Σ^* is also a convenient summary of the magnitude of this correlation and is used in this way in § 4·2.

The time required to carry out the analysis of a single data set ranged from about 12 minutes for the grouped ($\beta = \log 2$) and light censoring cases ($M = 50$, $M_0 = 0$, $k^* = 400$) to about 100 minutes for the $\beta = \log 10$ heavy censoring case ($M = 150$, $M_0 = 150$, $k^* = 400$). All computations were performed on a 486/60 IBM compatible PC equipped with an NDP i860 floating point processor.

4·2. *Performance of the Gibbs sampler*

To determine parameters such as the block size M, the number of blanks M_0, and the number of stochastic approximation steps k^*, we must assess the correlation induced by the Gibbs sampler between nearby values of $S(R_n | \{x_i\})$. To accomplish this, for each value of β ($= \log 2$, $\log 10$), and for each type of censoring, we studied the data set that showed evidence of the strongest correlation between observations, as determined by having the highest value of MV^*/Σ^*; see § 4·1. Keeping β set at β^*, we generated a single Gibbs stream of length $T = 100\,000$ of rankings R_t and computed the autocorrelation function of $S(R | \{x_i\})$. Two examples out of the six are shown in Fig. 1 and are discussed below.

For our simulations with 'light' censoring, we found a rapidly decaying correlation in our worst-case data sets for both $\beta = \log 2$ and $\beta = \log 10$, with a correlation length of about 4. Figure 1(a) shows the autocorrelation function for the worst-case data set from the $\beta = \log 10$ simulation, along with 95% confidence bands for the autocorrelation function based on the null hypothesis of no correlation. Results for $\beta = \log 2$ were similar and are not shown. Given this rapidly decaying correlation, we chose a block size of $M = 50$ with $M_0 = 0$.

For our simulations under 'heavy' censoring, the decay of the autocorrelation function in the worst case was much slower than in the 'light' censoring case. Figure 1(b) shows the autocorrelation function for the worst-case data set for $\beta = \log 10$, which has a correlation length of approximately 150. As a result, larger blocks were used ($M = 150$), and blanks were inserted ($M_0 = 150$) to ensure that the $\bar{S}_k$'s were independent. The autocorrelation function for the worst-case data set for $\beta = \log 2$, not shown, had a correlation length of about 40, but the magnitudes of the correlations were smaller; based on these results we chose a block size of $M = 150$ with $M_0 = 0$.

For our simulations using grouped data, the autocorrelation function in the worst-case data set for $\beta = \log 2$ decayed more rapidly than in the worst-case 'light' censoring data set, while for $\beta = \log 10$ the autocorrelation function of the worst-case data set had a slightly slower decay than in the 'heavy' censoring simulations. Based on these results, we chose $M = 50$ for both simulations but used $M_0 = 0$ for $\beta = \log 2$, and $M_0 = 250$ for $\beta = \log 10$.

We also tested the independence of the $\bar{S}_k$'s in the six worst-case data sets directly, by

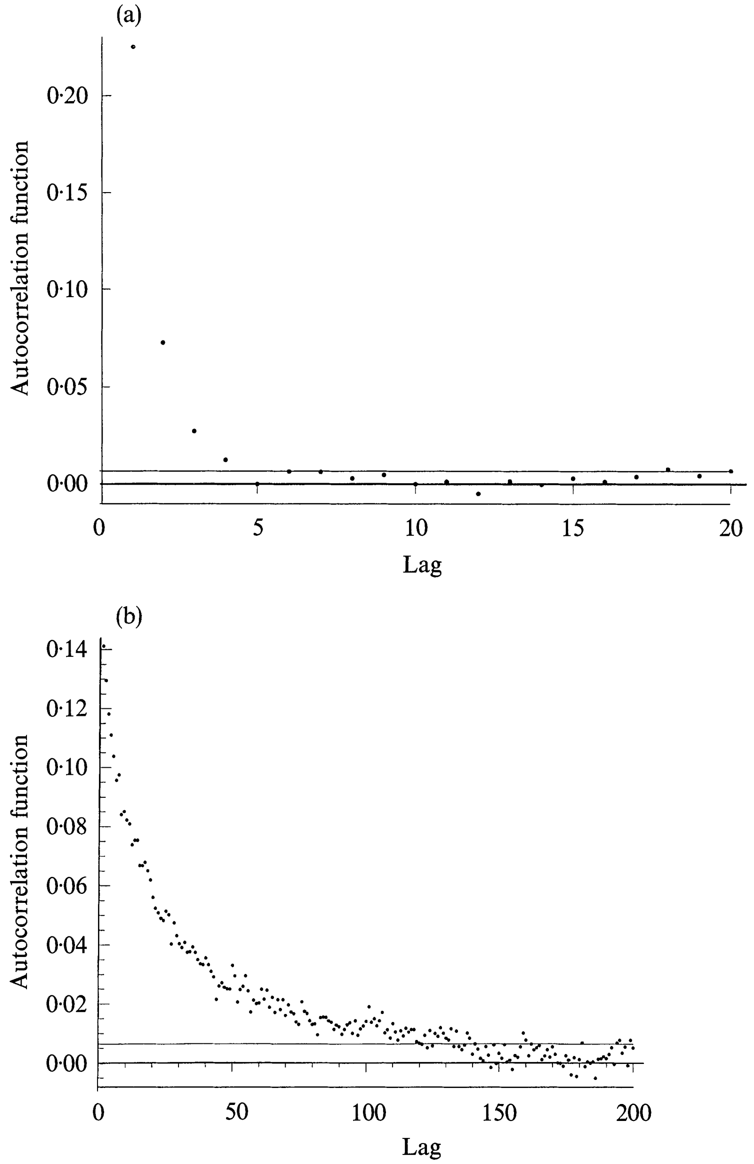

Fig. 1. Autocorrelation function for the worst-case data set (a) from the 'light censoring' simulations, and (b) from the 'heavy censoring' simulations; $\beta = \log 10$. Also shown are the 95% confidence limits for the autocorrelation function under the hypothesis of no correlation.

simulating a long Gibbs stream with β held fixed at β^*, calculating values of $\bar{S}_k$ using (16), and testing their independence using the rank von Neumann ratio (Madansky, 1988, p. 116–8). Block sizes M and blanks M_0 were chosen to match the values used in the simulations. We tested for independence using 10 $k^* = 4000$ blocks; even with the larger number of blocks little evidence of correlation was found, with all p-values $\geqslant 0{\cdot}08$. It should be noted that we have been very conservative in choosing M and M_0 for all simulations by examination of the worst-case data set; for most data sets smaller values could have been used.

5. DISCUSSION

5·1. *Computational strategies*

Our experience with finding the maximum marginal likelihood estimator $\hat{\beta}$ for the proportional hazards model using stochastic approximation coupled with a Gibbs sampler suggests the following modelling strategy. First, some estimates of the magnitude of $\hat{\beta}$ can be made without worrying about the correlation between the $\bar{S}_k$'s, with block size M and with $M_0 = 0$ and k^* chosen more with computation time in mind than overall accuracy. Then, the correlation structure of the $S(R \mid \beta, \{x_i\})$'s induced by the Gibbs sampler can be studied via the autocorrelation function. Once this has been done, values of M and M_0 can be chosen, with $M_0 = 0$ unless there is fairly long-range correlation. A fairly short run will then give an estimate of the magnitudes of Σ and V; these values can be used to decide on the value of k^*.

A second question of modelling strategy is whether k^* should be chosen to be sufficiently large so that there is a very high probability that the difference between β^* and $\hat{\beta}$ is smaller than the number of significant digits reported. In this case we can be quite confident that the reported value β^* is 'equal to' $\hat{\beta}$. Alternatively, we may report a value β^* which may differ from $\hat{\beta}$ as a separate estimate of the effect of covariates on survival; in this case, we must also report the larger confidence interval derived in § 3·3. The strategy of using a β^* which has not 'fully converged' to $\hat{\beta}$ is particularly appealing during model selection: if the extra width of the confidence interval appears irrelevant to the decision on whether to include or exclude a given effect, little is gained by carrying out a full analysis with maximal k^* for an intermediate result. Computational resources and data set size will play an important role in these decisions.

5·2. *Summary*

We have presented a stochastic approximation scheme using Gibbs sampling that allows calculation of maximum marginal likelihood estimates of parameters in the proportional hazards model. Like the usual proportional hazards model, these parameters can be estimated without specifying the baseline hazard. In addition, if the censoring intervals shrink so that each individual's censoring interval does not overlap with any other censoring interval so that $\mathscr{R}$ contains only 1 element, the observed ranking, then our method rapidly converges to the usual proportional hazards model estimates.

Although the sums in (1) or (2) could be performed in some cases, the number of such terms can grow very rapidly. For example, in the 'light' censoring case we considered, although each individual's censoring interval only overlaps with that of one individual before and one individual after, the number of distinct rankings in $\mathscr{R}$ grows exponentially with N; for $N = 200$, as in § 4, the number of elements in $\mathscr{R}$ is about $4{\cdot}5 \times 10^{41}$. In the grouped data examples considered, the number of distinct rankings in $\mathscr{R}$ ranged from $1{\cdot}3 \times 10^{144}$ to $2{\cdot}3 \times 10^{172}$; we have no expression for the size of $\mathscr{R}$ in the 'heavy censoring' case, although it is presumably much larger than the grouped or 'light censoring' cases. Although $\mathscr{R}$ is large in the cases we considered, our methods appear to provide accurate estimates of $\hat{\beta}$, as measured by the similarity between β^* and $\hat{\beta}_e$, the maximum likelihood estimator of the usual proportional hazards model using exact failure times; this similarity, shown in Tables 1 and 2, is especially noteworthy in the 'light censoring' and grouped cases, where the information lost by interval censoring is small and we would expect $\hat{\beta}$ to be close to $\hat{\beta}_e$.

A considerable portion of the discussion in §§ 3·3 and 4 was devoted to achieving conditional independence of the observed scores used in the stochastic approximation. We believe that independence is required primarily so that we can estimate the variance of β^* around $\hat{\beta}$, since long runs of the stochastic approximation procedure in several data sets from our simulation study appear to converge to the same value regardless of whether the errors in the stochastic approximation are independent. Hence, another computational approach may be to obtain several shorter stochastic approximation runs to estimate the variability of β^* about $\hat{\beta}$, with the point estimate taken to be the average value. Further results are also needed to establish the asymptotic distribution of $\hat{\beta}$ itself.

Although the simulation results presented here are for a single binary explanatory variable, we have also conducted a smaller number of simulation studies using continuous and bivariate explanatory variables, results not shown. The method appeared to perform equally well in these cases.

In conclusion, we would like to reiterate those points that are key to our ability to analyse large interval censored data sets using the marginal likelihood of the proportional hazards model. First, previous work in this area (Self & Grossman, 1986; Sinha et al., 1994) has been hampered by the lack of an efficient way to generate samples from $w_\beta(R)$; we have provided a simple Gibbs sampling scheme for generating these samples. Secondly, use of stochastic approximation provides an efficient, recursive methodology for maximising the marginal likelihood; finally, by using multiple rankings for each trial value of parameters, we make efficient use of the dependent iterates generated by a Gibbs sampler in the stochastic approximation scheme.

References

Besag, J. (1986). On the statistical analysis of dirty pictures. *J. R. Statist. Soc.* B **48**, 259–302.

Brookmeyer, R. & Gail, M. H. (1994). *AIDS Epidemiology: A Quantitative Approach.* New York: Oxford University Press.

Cox, D. R. (1972). Regression models and life-tables (with Discussion). *J. R. Statist. Soc.* B **26**, 187–220.

DeLong, D. M., Guirguis, G. H. & So, Y. C. (1994). Efficient computation of subset selection probabilities with application to Cox regression. *Biometrika* **81**, 607–11.

Diamond, I. D., McDonald, J. W. & Shah, I. H. (1986). Proportional hazards models for current status data: application to the study of differentials in age at weaning in Pakistan. *Demography* **23**, 607–20.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* **42**, 845–54.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statist. Sci.* **7**, 473–511.

Kalbfleisch, J. D. & Prentice, R. L. (1972). Discussion of paper by D. R. Cox. *J. R. Statist. Soc.* B **26**, 215–6.

Kalbfleisch, J. D. & Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika* **60**, 267–78.

Kalbfleisch, J. D. & Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: John Wiley.

Kipnis, C. & Varadhan, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple excursions. *Comm. Math. Phys.* **104**, 1–19.

Madansky, A. (1988). *Prescriptions for Working Statisticians.* New York: Springer-Verlag.

Peto, R. (1972). Discussion of paper by D. R. Cox. *J. R. Statist. Soc.* B **26**, 205–7.

Pettitt, A. N. (1983). Approximate methods using ranks for regression with censored data. *Biometrika* **70**, 121–32.

Ruppert, D. (1991). Stochastic approximation. In *Handbook of Sequential Analysis*, Ed. B. K. Ghosh and P. K. Sen, pp. 503–29. New York: Marcel Dekker.

Ruppert, D., Reish, R. L., Deriso, R. B. & Carroll, R. J. (1984). Optimization using stochastic approximation and Monte Carlo simulation (with application to harvesting of Atlantic menhaden). *Biometrics* **40**, 535–45.

Self, S. G. & Grossman, E. A. (1986). Linear rank tests for interval-censored data with application to PCB levels in adipose tissue of transformer repair workers. *Biometrics* **42**, 521–30.

SINHA, D., TANNER, M. A. & HALL, W. J. (1994). Maximization of the marginal likelihood of grouped survival data. *Biometrika* **81**, 53–60.

TANNER, M. A. (1993). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions.* New York: Springer-Verlag.

[*Received June* 1994. *Revised November* 1995]