Marginal Likelihoods Based on Cox's Regression and Life Model

Author(s): J. D. Kalbfleisch and R. L. Prentice
Source: *Biometrika*, Vol. 60, No. 2 (Aug., 1973), pp. 267-278
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: https://www.jstor.org/stable/2334538
Accessed: 17-06-2019 09:14 UTC

*Biometrika* (1973), **60**, 2, *p.* 267
*With* 1 *text-figure*
*Printed in Great Britain*



# Marginal likelihoods based on Cox's regression and life model

BY J. D. KALBFLEISCH AND R. L. PRENTICE*
*State University of New York at Buffalo*

SUMMARY

Marginal likelihoods are obtained for the regression parameters in the model presented by Cox (1972). If no ties occur in the recording of failure time data the results of Cox are given a straightforward justification. If ties occur in the data, results different from those suggested by Cox are obtained. Some Monte-Carlo comparisons of these competing results are made. A discrete model is developed for grouped data from Cox's model and estimates of the survivor function are given for both continuous and grouped data.

*Some key words:* Marginal sufficiency; Marginal likelihood; Survivals with covariates; Ranks; Group invariance; Regression; Censored data; Tied data.

## 1. INTRODUCTION

Cox (1972) suggested a regression model for the failure time $t$ of an individual when the covariate $\mathbf{z}' = (z_1, \dots, z_p)$ is recorded. For $t$ continuous the model was specified by a hazard function

$$\lambda(t, \mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta}\mathbf{z}), \tag{1}$$

where $\boldsymbol{\beta}$ is a row vector of $p$ unknown parameters and $\lambda_0(t)$ is arbitrary. This model, though largely nonparametric, permits the estimation of $\boldsymbol{\beta}$ and leads to estimates of the survivor function of the Kaplan & Meier (1958) type when covariates are present in the data.

Cox (1972) discussed this model in some detail and obtained likelihood results for the estimation of $\boldsymbol{\beta}$. It does not appear possible, however, to interpret his result as a proper conditional likelihood (Kalbfleisch & Prentice, 1972), so that the derivation of these results require some additional justification. In this paper the class of models (1) is restricted to those that possess a strictly monotone survivor function or, equivalently, to those for which the hazard function $\lambda_0(t)$ is not identically zero over an open interval. The invariance of this restricted class, $C$ say, under the group of monotone increasing transformations on $t$ is exploited to derive a marginal likelihood for $\boldsymbol{\beta}$. In the case where no ties occur the marginal likelihood is identical to, and therefore a justification of, the result that Cox obtains. If, however, failure times from $C$ are grouped before being recorded and some ties result, marginal likelihoods are obtained that differ from Cox's corresponding result.

The above derivations are carried out in §§2 and 3. A discrete model is developed for grouped data from model (1) to replace the logistic model introduced by Cox as an approximation. Estimates of the survivor function, alternative to those based on the logistic model, are derived for both continuous and grouped data in §4. A numerical example and some comparison of competing results are given in §5.

Notational difficulties arise in the derivations which follow. In expressing the final result

* On leave from the University of Waterloo, Canada.

the risk set notation has been used to facilitate comparisons with Cox's work. This notation, though very convenient for summarizing results, is not suitable for some derivations. Additional notation has therefore been introduced at the beginning of most sections.

## 2. THE MARGINAL LIKELIHOOD

### 2.1. *General*

The class $C$ of models defined above is invariant under and, in fact, is generated by the group $G$ of differentiable, strictly monotone increasing transformations on $t$. For if $\lambda_0(t)\exp(\boldsymbol{\beta}\mathbf{z})\in C$, then for any $g\in G$ the hazard of $t' = g(t)$ is

$$\lambda_0\{g^{-1}(t')\}\frac{\partial g^{-1}(t')}{\partial t'}\exp(\boldsymbol{\beta}\mathbf{z})\in C.$$

Conversely, suppose $t$ has hazard $\exp(\boldsymbol{\beta}\mathbf{z})$. Then an arbitrary $\lambda_0(t)\exp(\boldsymbol{\beta}\mathbf{z})$ in $C$ can be generated as the hazard of

$$t' = g(t) = \int_0^t \lambda_0(u)\,du,$$

and, since $\lambda_0(u)$ is not identically zero over an interval, $g\in G$.

### 2·2. *Uncensored data*

Let $t_1,\ldots,t_n$ represent survival times arising from $C$ and $\mathbf{z}_1,\ldots,\mathbf{z}_n$ be the corresponding covariate values. The group $G$ acts transitively on the order statistic $\{t_{(1)},\ldots,t_{(n)}\}$, while leaving the rank statistic $\mathbf{r}(\mathbf{t}) = \{(1),\ldots,(n)\}$ invariant. That is, two samples can be mapped one onto the other by an element of $G$ if and only if they yield the same rank statistic. The homomorphic group $H$ acting on the parameter space is transitive on the function $\lambda_0(.)$ and leaves the regression parameter $\boldsymbol{\beta}$ invariant. It follows from Barnard's (1963) definition that $\mathbf{r}(\mathbf{t})$ is marginally sufficient for the estimation of $\boldsymbol{\beta}$, that is 'sufficient for $\boldsymbol{\beta}$ in the absence of knowledge of $\lambda_0(.)$'. Therefore, by the definition of Kalbfleisch & Sprott (1970), or of Fraser (1968, pp. 188–9), the marginal likelihood of $\boldsymbol{\beta}, L_1(\boldsymbol{\beta})$, is obtained from the marginal distribution of the ranks. That is,

$$\begin{aligned} L_1(\boldsymbol{\beta}) &\propto \int_0^\infty\int_{t_{(1)}}^\infty\cdots\int_{t_{(n-1)}}^\infty \prod_{i=1}^n\left[\lambda_0(t_{(i)})\exp(\boldsymbol{\beta}\mathbf{z}_{(i)})\exp\left\{-\exp(\boldsymbol{\beta}\mathbf{z}_{(i)})\int_0^{t_{(i)}}\lambda_0(u)\,du\right\}\right]dt_{(n)}\ldots dt_{(1)} \\ &= \exp\left(\boldsymbol{\beta}\sum_{i=1}^n \mathbf{z}_i\right)\Big/\prod_{i=1}^n\sum_{l\in R(t_{(i)})}\exp(\boldsymbol{\beta}\mathbf{z}_l), \end{aligned} \tag{2}$$

where $R(t_{(i)})$ is the set of individuals at risk at $t_{(i)}-0$, that is $R(t_{(i)}) = \{(i),\ldots,(n)\}$.

The result (2) is identical to the result that Cox obtains in the uncensored case and in fact many of the arguments given in his derivation are similar to the concepts introduced above. Cox notes, for example, that 'No information can be contributed about $\boldsymbol{\beta}$ in time intervals in which no failures occur because the component $\lambda_0(t)$ might conceivably be identically zero in such intervals', and uses this to justify a conditional argument given the times at which failures occur. The group invariance argument given above is a formalization of this intuitive argument and does imply that the lengths of time between successive failures are irrelevant to the inference. It should be stressed, however, that the group invariance leads to consideration of the marginal distribution of the ranks. It would seem that the result (2) cannot be derived by a formal conditional argument as might be inferred from Cox's work.

### 2.3. *Censored data*

If censoring is included, the structure of the model is more complicated. If the entire sample were observed, we have seen by the above arguments that the rank statistic would be marginally sufficient and the relevant inference arises out of the distribution of the ranks. If a censored sample is obtained, however, the entire rank statistic is not observed; we have available only partial information on the ranks. Suppose, for example, that items 1, 2, 3, 4 are observed to have survival times 40, 10, 20*, 30, respectively, where the asterisk indicates a censored survival time. The rank statistic, on the basis of this data, is one of

$$(2,3,4,1), \quad (2,4,3,1) \quad \text{or} \quad (2,4,1,3).$$

In order to make an inference about $\boldsymbol{\beta}$, the marginal probability that the rank statistic should be one of these possible is used. In the above example, this is the sum of three terms of the type (2). This argument for the censored case, although intuitively appealing, cannot be justified directly by the group invariance arguments since the censored model does not possess the group invariant structure. The information being ignored, however, is the exact time of censoring and, the fact that the invariance of the uncensored model makes the lengths of the intervals between successive failures irrelevant, would suggest that little information is lost in this restriction.

Suppose that $k$ items give rise to observed ordered failure times $\{t_{(1)}, \ldots, t_{(k)}\}$ with corresponding covariates $\{\mathbf{z}_{(1)}, \ldots, \mathbf{z}_{(k)}\}$. Suppose further that $q_i$ items with covariates $(\mathbf{z}_{i1}, \ldots, \mathbf{z}_{iq_i})$ are censored in the $i$th interval $[t_{(i)}, t_{(i+1)})$ $(i = 0, 1, \ldots, k)$, where $t_{(k+1)} = \infty$ and $t_{(0)} = 0$. The marginal likelihood of $\boldsymbol{\beta}$ is proportional to the probability that the rank statistic should be one of those possible on this sample. A direct calculation of this would involve the summation of a very large number of terms like that in (2). The derivation, however, is considerably simplified by noting that the event whose probability we are calculating can be written

$$\begin{gathered} t_{(1)} < \ldots < t_{(k)}, \\ t_{(i)} \leqslant t_{i1}, \ldots, t_{(i)} \leqslant t_{iq_i} \quad (i = 0, 1, \ldots, k), \end{gathered} \tag{3}$$

where $t_{i1}, \ldots, t_{iq_i}$ are the unobserved failure times associated with individuals censored in $[t_{(i)}, t_{(i+1)})$. It is easily seen that, given $t_{(i)}$, the event $t_{(i)} \leqslant t_{i1}, \ldots, t_{(i)} \leqslant t_{iq_i}$ has conditional probability

$$\exp\left\{-\sum_{j=1}^{q_i} \exp(\boldsymbol{\beta}\mathbf{z}_{ij}) \int_0^{t_{(i)}} \lambda_0(u)\, du\right\} \quad (i = 0, 1, \ldots, k).$$

Thus, the probability of the event (3) is easily calculated and is proportional to the marginal likelihood of $\boldsymbol{\beta}$. That is

$$\begin{aligned} L_2(\boldsymbol{\beta}) \propto \int_0^\infty \int_{t_{(1)}}^\infty \cdots \int_{t_{(k-1)}}^\infty \prod_{i=1}^k \Bigg(\lambda_0(t_{(i)}) \exp(\boldsymbol{\beta}\mathbf{z}_{(i)}) \\ \times \exp\left[-\left\{\exp(\boldsymbol{\beta}\mathbf{z}_{(i)}) + \sum_{j=1}^{q_i} \exp(\boldsymbol{\beta}\mathbf{z}_{ij})\right\} \int_0^{t_{(i)}} \lambda_0(u)\, du\right]\Bigg)\, dt_{(k)} \ldots dt_{(1)} \\ = \exp\left(\boldsymbol{\beta} \sum_{i=1}^k \mathbf{z}_{(i)}\right) \Bigg/ \prod_{i=1}^k \sum_{l \in R(t_{(i)})} \exp(\boldsymbol{\beta}\mathbf{z}_l), \end{aligned} \tag{4}$$

where $R(t_{(i)}) = \{(i), i1, \ldots, iq_i; \ldots; (k), k1, \ldots, kq_k\}$ is the risk set at $t_{(i)}$. Once again Cox's result is derived as a marginal likelihood.

## 3. Tied data

In most practical applications, survival times can reasonably be thought of as continuous random variables and so the continuous model (1) is very general in its scope. Nevertheless, in the recording of survival data, measurement errors will always be introduced and some ties may result as, for example, if survivals arising from the model (1) in the class $C$ are subject to grouping into disjoint groups before being recorded. In this section, the marginal likelihood (2) is generalized to include the case where ties are present in the recorded data.

Suppose that, of $n$ individuals under test, individuals $i1, \ldots, im_i$ are recorded as failing at the ordered failure time $t_{(i)}$ $(i = 1, \ldots, k)$, where $\Sigma m_i = n$. This situation can be handled in a manner analogous to the way in which censored data points were handled in § 2. Here again, only partial information is available on the rank vector; the $m_i$ individuals failing at $t_{(i)}$ are known to have ranks less than those recorded as failing at $t_{(j)}$ for $j > i$, but the arrangement of the ranks of the $m_i$ individuals failing at $t_{(i)}$ is otherwise unknown. Viewed in this way, the marginal likelihood of $\boldsymbol{\beta}$ is proportional to the probability that the rank vector should be one of those possible given the sample. This probability is the sum of the probabilities of the $m_1! \ldots m_{k_i}!$ possible rank vectors; it is necessary, therefore, to evaluate a $k$-dimensional sum of terms of the type (2).

The calculation of the marginal likelihood is considerably simplified by noting that the ranks assigned to the $m_i$ individuals who fail at $t_{(i)}$ is unaffected by the ranks assigned the $m_j$ individuals who fail at $t_{(j)}$. The $k$-dimensional sum, therefore, reduces to the product of $k$ one-dimensional sums. Let $Q_i$ be the set of permutations of the symbols $i1, \ldots, im_i$ and let $(p_1, \ldots, p_{m_i})$ be one element in $Q_i$. As before, $R(t_{(i)})$ is the risk set at $t_{(i)}$ and let $R(t_{(i)}, p_r)$ be the set difference $R(t_{(i)}) - (p_1, \ldots, p_{r-1})$. The marginal likelihood for $\boldsymbol{\beta}$ can then be written

$$L_3(\boldsymbol{\beta}) \propto \prod_{i=1}^{k} \exp(\boldsymbol{\beta}\mathbf{s}_{(i)}) \sum_{(p_1, \ldots, p_{m_i}) \in Q_i} \prod_{r=1}^{m_i} \left\{ \sum_{l \in R(t_{(i)}, p_r)} \exp(\boldsymbol{\beta}\mathbf{z}_l) \right\}^{-1}, \tag{5}$$

where

$$\mathbf{s}_{(i)} = \sum_{j=1}^{m_i} \mathbf{z}_{ij}$$

is the sum of the covariates of the individuals observed to fail at $t_{(i)}$. The notation used in writing (5) is sufficiently general that the case where censored data are present is also covered. The likelihoods $L_1(\boldsymbol{\beta})$ and $L_2(\boldsymbol{\beta})$ are special cases of (5).

The likelihood $L_3(\boldsymbol{\beta})$ differs from the result obtained by Cox to cover this case. In this latter result, summations are taken which may involve a very large number of terms and typically (5) would be easier to compute. Some simplification occurs in the calculation of $L_3(\boldsymbol{\beta})$ in that if $p_i$ of the $m_i$ individuals failing at $t_{(i)}$ have the same covariate value, only $m_i!/p_i!$ terms in the sum over $Q_i$ need be computed. More discussion of the computational aspects of these results is contained in § 5.

Cox's likelihood was derived by consideration of a logistic model for the discrete failure time classes. As such, the likelihood he obtains would appear to be a statement of inference about the regression parameter in the logistic model rather than in the model (1). It should be noted, however, that the same difficulties arise in formalizing Cox's result as a conditional likelihood as were mentioned in § 2. The logistic model was shown to provide a first-order approximation to the model (1) so that the parameter meaning would be largely unaffected if the grouping intervals were narrow. If, however, the grouping intervals were broader, this could be a serious drawback. For example, consider the two-sample case where the survival times are exponentially distributed with failure rates 0·2 (sample 0) and $0{\cdot}2e^{0\cdot5}$ (sample 1). Viewing this as a member of the set of models (1), $\beta = 0{\cdot}5$. If survivals are grouped in intervals $[0, 1)$, $[1, 2), \ldots$, however, the corresponding logistic parameter is approximately 0·57. Cox's likelihood would appear to be estimating this parameter. On the other hand, the likelihood $L_3(\boldsymbol{\beta})$ is suitable for estimating the original parameter $\boldsymbol{\beta}$ from the continuous model whatever the size of the grouping interval.

The hazard function for the discrete model obtained by grouping the model (1) into disjoint time intervals can be obtained directly. Suppose that survival times in the interval $[a_{i-1}, a_i)$ are recorded as $t_i\,(i = 1, 2, \ldots)$ $(a_0 = 0)$. The probability of an observed failure time $t_i$ when the covariate is $\mathbf{z}$ is given by

$$P(t_i|\mathbf{z}) = \{1-\alpha_i^{\exp\,(\boldsymbol{\beta}\mathbf{z})}\} \prod_{j=1}^{i-1} \alpha_j^{\exp(\boldsymbol{\beta}\mathbf{z})},$$

where

$$\alpha_j = \exp\left\{-\int_{a_{j-1}}^{a_j} \lambda_0(u)\,du\right\} \quad (j = 1, \ldots, k).$$

The contribution of the hazard function at $t_i$ when $\mathbf{z} = \mathbf{0}$ is, therefore, $1-\alpha_j$ and for an arbitrary $\mathbf{z}$ is $1-\alpha_j^{\exp\,(\boldsymbol{\beta}\mathbf{z})}$. To rewrite this result in a more general way, let $\lambda(t_i;\,\mathbf{z})\,dt_i$ be the hazard, or conditional probability of failure, at $t_i$. Then

$$\lambda(t_i;\,\mathbf{z})\,dt_i = 1-\{1-\lambda_0(t_i)\,dt_i\}^{\exp\,(\boldsymbol{\beta}\mathbf{z})} \tag{6}$$

specifies both the discrete and the continuous model. The estimation of the hazard function using the discrete model from (6) is considered in § 4.

We (Kalbfleisch & Prentice, 1972) discussed an example in which two items with scalar covariates $z_1$ and $z_2$ are recorded as tied and pointed out that Cox's likelihood, in this case, is constant indicating that no one value of $\beta$ is to be preferred to any other. It was argued that if $z_1 \neq z_2$, $\beta = 0$ should have been a preferred value. Since in the derivation of $L_3(\beta)$, no assumptions are introduced about the width of the grouping intervals, it also gives the uniform result in this example. At the time of writing those comments, we were considering the case of multiplicities with a different approach from that discussed above; the general group structure in § 2 was used to produce an invariant partition of the sample space which allowed for ties. This approach does give a unique maximum of the marginal likelihood at $\beta = 0$ in the above example although the general result stated in our comments is incorrect. This alternative approach to the multiplicities case is outlined in the appendix. Since the computations required, however, in this approach are generally very difficult and the approach would seem relevant only if the grouping intervals were small, we are of the opinion that the computationally simpler and more generally applicable result (5) is the relevant likelihood in the multiplicities case.

## 4. The estimation of the survivor function

Cox (1972) suggested an iterative procedure for estimating the survivor function at $\mathbf{z} = \tilde{\mathbf{z}}$ given a set of data from the model (1) and an estimate of $\boldsymbol{\beta}$. In obtaining the estimate, it was assumed that $\lambda_0(t)$ was identically zero aside from mass points at the observed failure times. The data were taken as having arisen from his assumed logistic discrete analogue of the model (1) and a separate maximum likelihood estimation of the hazard at each failure point was then proposed. One difficulty with this approach is that the logistic model cannot be obtained by grouping the continuous model (1), as was pointed out in § 3. As a result of this, the estimates of the survivor function obtained for different values of $\mathbf{z}$ relate to one another through the logistic model only; continuous survivor functions cannot be constructed which are arbitrarily close to these estimates and relate to one another through the model (1). The result, therefore, is not a legitimate supremum of the likelihood defined on the model (1).

A procedure more compatible with the continuous model is obtained by considering the discrete model (6) instead of the logistic case. Following Cox's approach, the estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ from the marginal likelihood is used. If the model (6) is adopted, the maximum likelihood estimate of the contribution $\hat{\lambda}_0(t_{(i)})\,dt_{(i)} = 1-\hat{\alpha}_i$ to the hazard at $t = t_{(i)}$ is given by

$$\sum_{k\in F_i} \frac{\exp(\hat{\boldsymbol{\beta}}\mathbf{z}_k)}{1-\hat{\alpha}_i^{\exp(\hat{\boldsymbol{\beta}}\mathbf{z}_k)}} = \sum_{l\in R(t_{(i)})} \exp(\hat{\boldsymbol{\beta}}\mathbf{z}_l), \tag{7}$$

where $F_i$ is the set of individuals failing at $t_{(i)}$. If only a single failure occurs at $t_{(i)}$ $(m_i = 1)$, or all individuals in $F_i$ have the same covariate value, this equation can be solved analytically for $\hat{\alpha}_i$. Otherwise, an iterative solution is required. A suitable starting value for the iteration is

$$\hat{\alpha}_{i0} = \exp\{-m_i/\sum_{l\in R(t_{(i)})} \exp(\hat{\boldsymbol{\beta}}\mathbf{z}_l)\}. \tag{8}$$

In fact, since expression (8) is obtained by substituting

$$1+\exp(\hat{\boldsymbol{\beta}}\mathbf{z}_k)\log\hat{\alpha}_i \doteq \exp\{\exp(\hat{\boldsymbol{\beta}}\mathbf{z}_k)\log\hat{\alpha}_i\}$$

in (7), we expect $\hat{\alpha}_{i0}$ will approximate $\hat{\alpha}_i$ very closely if there are many distinct failure times, that is the $\hat{\alpha}_i$'s are near 1. It now easily follows from (6) that the estimated survivor function for a covariate value $\mathbf{z} = \tilde{\mathbf{z}}$ is

$$\hat{\mathscr{F}}_{\tilde{\mathbf{z}}}(t) = \prod_{\{i|t_{(i)}<t\}} \hat{\alpha}_i^{\exp(\hat{\boldsymbol{\beta}}\tilde{\mathbf{z}})}. \tag{9}$$

Expression (9) is a legitimate supremum of the likelihood function for the continuous case while, as noted above, the logistic result is not. A sequence of continuous survivor functions can be constructed statisfying the model (1) which converge to (9) for all values of $\tilde{\mathbf{z}}$. As could be expected, when the covariate $\mathbf{z}$ is the same for all individuals sampled, (9) reduces to the Kaplan–Meier product limit estimate.

The estimate (9) of the survivor function will typically require an iterative solution when ties are present in the data, although the computations are fairly simple. Further, (9) is a step function estimate of the survivor function and in many instances a continuous estimate would be preferable, especially for suggesting a parametric form for $\lambda_0(t)$ or for communicating information to non-statisticians. In the remainder of this section, a continuous estimate of the survivor function is derived which has the further advantage of being computationally simpler than (9).

We begin by approximating the true log survivor function by a connected series of straight lines or, equivalently, by approximating the hazard function by a step function

$$\lambda_0(t) = \begin{cases} \lambda_1 & t\in[b_0, b_1) = I_1, \\ \lambda_2 & t\in[b_1, b_1+b_2) = I_2, \\ \vdots & \\ \lambda_r & t\in[b_1+\ldots+b_{r-1}, b_1+\ldots+b_r) = I_r, \end{cases} \tag{10}$$

where $b_0, b_1, \ldots, b_r$ are prespecified with $b_0 = 0$ and $b_r = \infty$. If (10) is assumed as part of the model (1) with $\lambda_1, \ldots, \lambda_r$ unknown, then for any specified value of $\boldsymbol{\beta}$ the quantities $\lambda_1, \ldots, \lambda_r$

are easily estimated and the corresponding survivor function obtained. The regression parameter is therefore assumed specified as $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the estimate from the marginal likelihood of $\boldsymbol{\beta}$.

Suppose that $t_{i1}, \ldots, t_{im_i}$ are observed survival times in the $i$th interval and that the corresponding covariate values are $\mathbf{z}_{i1}, \ldots, \mathbf{z}_{im_i}$. Suppose further that $d_i$ of these times correspond to observed failures and let

$$C_i = \sum_{l=i+1}^{r} \sum_{j=1}^{m_l} \exp(\hat{\boldsymbol{\beta}} \mathbf{z}_{lj}),$$

$$D_i = \sum_{j=1}^{m_i} (t_{ij} - b_1 - \ldots - b_{i-1}) \exp(\hat{\boldsymbol{\beta}} \mathbf{z}_{ij}).$$

The likelihood function of $\lambda_1, \ldots, \lambda_r$ given that $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ is easily found to be proportional to

$$\prod_{i=1}^{r} \lambda_i^{d_i} \exp\{-\lambda_i(D_i + b_i C_i)\},$$

so that the maximum likelihood estimate of $\lambda_i$ is

$$\hat{\lambda}_i = d_i/(D_i + b_i C_i).$$

The corresponding estimate of the survivor function is

$$\log \mathscr{F}_0(t) = -t\hat{\lambda}_{k(t)} - \sum_{i=1}^{k(t)} \{\hat{\lambda}_i - \hat{\lambda}_{k(t)}\} b_i \quad (t > 0), \tag{11}$$

where $k(t) = j$ if $t \in I_j$. The plot of this function is a connected sequence of straight lines. It should be noted that although (11) is the estimated survivor function for $\mathbf{z} = \mathbf{0}$, the estimated survivor function for an arbitrary $\mathbf{z}$ can be obtained merely by replacing $\hat{\lambda}_i$ by

$$\hat{\lambda}_i \exp(\hat{\boldsymbol{\beta}} \tilde{\mathbf{z}}) \quad (i = 1, \ldots, r).$$

This approach to estimating the survivor function is similar to the approach suggested by Oakes (1972) and Breslow (1972). Instead of selecting the intervals $I_1, \ldots, I_r$ independently of the data, however, they suggest defining the intervals as beginning and ending at the observed failure times; the calculations then proceed as outlined above. In order to write down the likelihood of the parameters $\lambda_1, \ldots, \lambda_r$ we must assume that the model is of the form (10) and consequently that the intervals are chosen independently of the data. The actual survival curves obtained will be similar, however, in the two approaches.

## 5. Numerical illustrations

Monte-Carlo comparisons of the marginal and conditional results are made difficult by the fact that both likelihoods can be very complex to compute when many ties are reported in the data. The choice of examples is also somewhat restricted by the fact that the discrete logistic model used in obtaining the conditional result is not compatible with the grouped continuous model. Accordingly, the comparisons that have been made have all been within a two sample problem with exponential survival distributions and with constant grouping intervals. The discrete logistic model fits exactly in this situation. The examples presented in this section are for exponentials with failure rates $0{\cdot}2$ and $0{\cdot}2e^{0{\cdot}5}$, respectively. In the following discussion, the marginal estimate of $\beta$, $\hat{\beta}_M$, will mean the estimate of $\beta$ arising out of (5) while the conditional estimate, $\hat{\beta}_C$, will mean the estimate arising out of the conditional likelihood of Cox.

In Table 1, the estimation of the parameter $\beta$ is exemplified. For this example, samples of size 40 were generated, one from each of the two uncensored exponential distributions described above, and the estimates $\hat{\beta}_M$ and $\hat{\beta}_C$ were computed for various grouping intervals. It is easily seen that, while the marginal estimate is fairly stable over changes in the grouping interval, the conditional estimate increases as the grouping interval broadens. This increase is in correspondence to an increase in the logistic parameter. This type of behaviour of the estimates has been well substantiated in several trials and the summary in Table 1 is quite typical. From these simulations it would appear that the conditional estimate is of the logistic parameter rather than of the parameter in the original continuous model. This is not immediately clear since the conditional likelihood in the multiplicities case can be derived by applying the same argument to the discrete or to the continuous model.

Table 1. *Marginal* ($\hat{\beta}_M$) *and conditional* ($\hat{\beta}_C$) *estimates of the ratio* (0·5) *of failure rates from simulated grouped exponential samples*

| Grouping interval size | $\hat{\beta}_M$ | $\hat{\beta}_C$ | Logistic parameter value |
|---|---|---|---|
| 0·25 | 0·418 | 0·428 | 0·516 |
| 0·50 | 0·429 | 0·453 | 0·533 |
| 0·75 | 0·438 | 0·476 | 0·550 |
| 1·00 | 0·428 | 0·476 | 0·568 |
| 1·50 | 0·432 | 0·496 | 0·603 |
| 2·00 | 0·404 | 0·504 | 0·641 |
| 2·50 | 0·477 | 0·646 | 0·680 |
| 3·00 | 0·474 | 0·617 | 0·720 |
| 3·50 | 0·417 | 0·622 | 0·762 |

Further simulations were done with a grouping interval of length 1. A simulation involving 100 trials in which two samples of size 200 were generated each time gave an average value of $\hat{\beta}_C$ of 0·570 with an estimated standard deviation for the average of 0·012. This seems very compatible with the logistic parameter which for this case is 0·568. Large scale simulations for the marginal estimate for samples of this size are made difficult by the fact that the calculations are extremely laborious. However, using an approximation outlined below, 50 trials were carried out and the marginal and conditional estimates were computed. The average value of the marginal estimate was 0·485 and of the conditional estimate 0·546. The correlation observed between the two estimates was 0·50. These calculations, together with the larger simulation done for the conditional estimate, would suggest that the distribution of the marginal estimate is nearly centred on the true value $\beta = 0·5$. The standard deviations of the estimates obtained in these trials were 0·100 and 0·113 for the marginal and conditional cases, respectively.

Other simulations for smaller sample sizes and grouping interval showed that the conditional estimate has an average value about 0·07 units larger than the marginal estimate. The standard deviation of the marginal estimate was less than that of the conditional estimate in all of the cases considered. In small samples, both estimators are biased with mean values greater than 0·5.

As hinted above, for the two sample cases the computations involved in the marginal likelihood are much more complex than those for the conditional likelihood. This is by no means true in general. If continuous covariates are present in the data and ties occur, the conditional likelihood quickly becomes uncomputable. Consider, for example, the two-sample data quoted by Cox (1972) and given in Table 2. Suppose that a continuous covariate such as white blood cell count were included for each individual in the sample. The number of terms in the sum of the denominator at the 8th failure point time is $28!/(4!\,24!)$ which is

Table 2. *Times of remission in weeks of leukemia patients*

| | |
|---|---|
| Sample 0 | 6*, 6, 6, 6, 7, 9*, 10*, 10, 11*, 13, 16, 17*, 19*, 20*, 22, 23, 25*, 32*, 32*, 34*, 35* |
| Sample 1 | 1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23 |

* Censored.

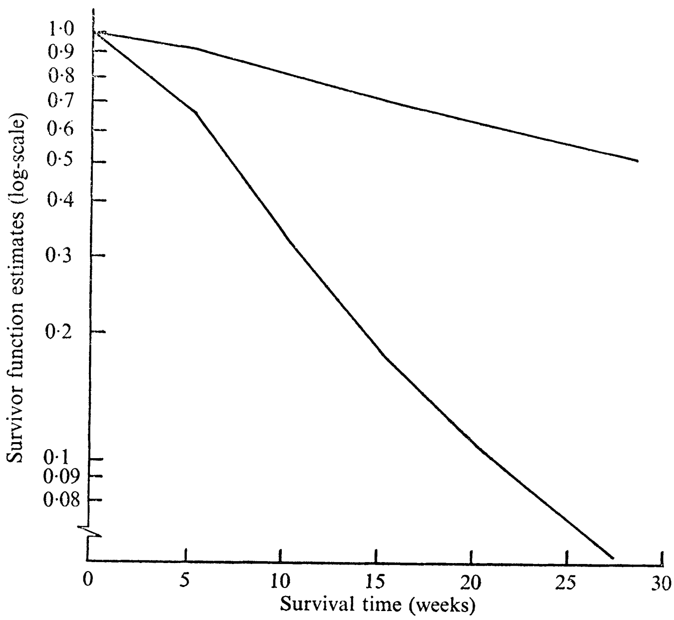


Fig. 1. Continuous survivor function estimates from the data of Table 2. Upper line, sample 0; lower line, sample 1.

already becoming sizeable. The marginal likelihood, on the other hand, will involve the product of 17 sums with each summation being over at most 24 terms. In fact, the marginal likelihood will offer no computational difficulties unless the number of observations tied at a given time point is quite large; 7 or 8 begins to cause difficulties.

If the number of ties is large at some survival times a simple approximation to the marginal likelihood is easily obtained. The $i$th term in the product (5) is the sum over all possible rank arrangements of the $m_i$ items tied at the $i$th failure time and so involves $m_i!$ terms. A sample of the possible rank configurations at the $i$th failure time can be drawn at random and the sum computed over these configurations only, when $m_i$ is too large. In our simulations, we have found that a sample of size 50 at each of the problem failure times is sufficient to yield reliable results. This approximation makes possible the computation of the marginal likelihood in even the most complex examples.

The continuous estimate of the survivor function (11) can be illustrated using the data of Table 2. The marginal likelihood estimate is $\hat{\beta}_M = 1{\cdot}59$ for these data. Figure 1 gives the continuous estimates of the survivor functions for sample 0 and for sample 1. The continuous estimates have the definite advantage of being more easily interpreted than the step function estimate from (9). The intervals used in obtaining the continuous estimates are defined by $b_0 = 0$, $b_1 = 5{\cdot}5$, $b_2 = b_3 = b_4 = 5$, $b_5 = \infty$.

## 6. Discussion

The above derivations, when no ties are present in the data, can be regarded as a formalization of the derivations of Cox (1972). The marginal likelihoods arise in the model (1) in a very natural way, following directly from the group invariance properties; the extensions to censored and grouped data gives a unified approach to the model. The results throughout this paper are exact and follow naturally from the continuous model (1). The introduction of the approximate logistic discrete analogue of (1) is unnecessary for the estimation either of $\boldsymbol{\beta}$ or of the survivor function.

Asymptotic results for the estimate arising out of the marginal likelihood (4) are important in cases where the parameter $\boldsymbol{\beta}$ is of high dimension. Cox (1972) has considered this problem for the conditional likelihood; his results are identical to those for the marginal case when no ties are present. When ties are present in the data, however, some modification is needed. As in the conditional case, however, the second derivative of the log likelihood from (5) can be computed in a straightforward manner and little difficulty is encountered in obtaining asymptotic approximations. The likelihoods in this case all arise out of marginal distributions and consequently asymptotic results should be more easily proved than in the conditional case. Andersen (1970) has discussed the asymptotic properties of conditional maximum likelihood estimates and has noted that conditions need to be placed on incidental nuisance parameters eliminated by conditioning. This may have implications for asymptotic results in conditional likelihoods, or approximate ones, arising from the discrete logistic model.

One fundamental difference between the results in this paper and the analogous conditional results is that the latter allow the inclusion of time dependent covariates. In the context of the above derivations, however, such covariates would remove the group invariance of the model (1). Our results can be derived only if the covariates are specified and not functions of time. The fact that the group structure acts transitively on the order statistic removes the possibility of including a time dependent covariate.

The principal use made of time dependent covariates in the conditional argument was to check the assumption of proportional hazards. In many instances this assumption can be checked by an alternative method. For instance, in the two sample problem with covariates, $\lambda_0(t)$ may be allowed to differ between the two samples, and separate estimates of the hazard function may be derived. A plot of these estimates on a log scale should give approximately constant differences along the time axis. A formal test to check for departures from an hypothesis of constant difference would be desirable.

Finally, the restriction of the class of models (1) to those possessing a strictly monotone survivor function is trivial in that any member of (1) can be arbitrarily closely approximated by an element of $C$.

Our thanks go to Professor D. R. Cox for his suggestions in revising this paper and the interest he has shown in this work. We would also like to thank Professors Norman Breslow and Marvin Zelen for interesting discussions on this model and related problems. This work was partially supported by a Veterans Administration Grant and by a PHS Training Grant from the National Cancer Institute.

## APPENDIX

### *Tied data from the class C*

Suppose that failure time data, including ties, are regarded as being exact, continuous and as arising from the class of models $C$. In § 2 the group $G$ acting on the sample space with continuous failure times left invariant the partition of the sample space defined by the rank statistic $\mathbf{r}(\mathbf{t}) = \{(1), \ldots, (n)\}$, where $n$ individuals are under test. If ties are allowed, the maximal invariant partition is the one with sets defined by

$$\begin{gathered} t_{i_1} < t_{i_2} < \ldots < t_{i_n}, \\ t_{i_1} = t_{i_2} < \ldots < t_{i_n}, \\ t_{i_1} < t_{i_2} = t_{i_3} < \ldots < t_{i_n}, \\ \vdots \\ t_{i_1} = t_{i_2} = \ldots = t_{i_n}, \end{gathered}$$

where $(i_1, \ldots, i_n)$ is a permutation of $(1, \ldots, n)$.

This partition is the minimal marginally sufficient statistic for $\boldsymbol{\beta}$ by Barnard's (1963) definition. In observing the survival times, one of the possible sets in the partition is determined and the marginal likelihood of $\boldsymbol{\beta}$ is proportional to the probability, or probability element, of that set. The problem is therefore to obtain these expressions.

Suppose that failure times from the model $C$ are written

$$t_{11}, \ldots, t_{1m_1}, t_{21}, \ldots, t_{2m_2}, \ldots, t_{k1}, \ldots, t_{km_k}$$

and consider calculation of the probability element corresponding to the event

$$t_{11} = \ldots = t_{1m_1} < t_{21} = \ldots = t_{2m_2} < \ldots < t_{k1} = \ldots = t_{km_k}. \tag{12}$$

The joint probability element of the $t_{ij}$'s is

$$\prod_{i=1}^{k} \prod_{j=1}^{m_i} \left[ \lambda_0(t_{ij}) \exp(\boldsymbol{\beta}\mathbf{z}_{ij}) \exp\left\{ -\exp(\boldsymbol{\beta}\mathbf{z}_{ij}) \int_0^{t_{ij}} \lambda_0(u)\, du \right\} \right] dt_{ij}, \tag{13}$$

where $\mathbf{z}_{ij}$ is the covariate corresponding to $t_{ij}$. Setting

$$y_{ij} = \int_0^{t_{ij}} \lambda_0(u)\, du,$$

(13) can be rewritten as

$$\prod_{i=1}^{k} \left[ \exp(\boldsymbol{\beta}\mathbf{s}_i) \exp\left\{ -\sum_{j=1}^{m_i} \exp(\boldsymbol{\beta}\mathbf{z}_{ij})\, y_{ij} \right\} \prod_j dy_{ij} \right],$$

where

$$\mathbf{s}_i = \sum_{j=1}^{m_i} \mathbf{z}_{ij}.$$

The $y_{ij}$'s can be reparameterized ($m_i \neq 1$) as

$$r_i = y_{i1}, \quad \tan\theta_{i2} = y_{i2}/y_{i1}, \quad \ldots, \quad \tan\theta_{im_i} = y_{im_i}/y_{i1}, \tag{14}$$

and the Jacobian of the transformation is $|J| = r_i^{m_i-1} g(\boldsymbol{\theta}_i)$, where $g(\boldsymbol{\theta}_i)$ is a function of $\theta_{i2}, \ldots, \theta_{im_i}$. This yields the joint probability element of the $r_i$'s and $\theta_{ij}$'s as

$$\prod_{i=1}^{k} \exp(\boldsymbol{\beta}\mathbf{s}_i) \exp\left\{-r_i \sum_{j=1}^{m_i} \exp(\boldsymbol{\beta}\mathbf{z}_{ij}) \tan\theta_{ij}\right\} r_i^{m_i-1} g(\boldsymbol{\theta}_i)\, dr_i\, d\theta_{i2} \ldots d\theta_{im_i},$$

where $\tan\theta_{i1} = 1$. If $m_i = 1$, the transformation (14) reduces to $r_i = y_{i1}$.

The condition (12) is equivalent to the conditions

$$r_1 < \ldots < r_k$$
$$\theta_{ij} = \tfrac{1}{4}\pi \quad (j = 2, \ldots, m_i;\ i = 1, \ldots, k;\ m_i \neq 1).$$

The marginal probability element of the event $r_1 < \ldots < r_k$ and the $\theta_{ij}$'s is

$$\int_0^\infty \int_{r_1}^\infty \cdots \int_{r_{k-1}}^\infty \prod_1^k \left[\exp(\boldsymbol{\beta}\mathbf{s}_i) \exp\left\{-r_i \sum_{j=1}^{m_i} \exp(\boldsymbol{\beta}\mathbf{z}_{ij}) \tan\theta_{ij}\right\} r_i^{m_i-1} g(\boldsymbol{\theta}_i)\right] dr_k \ldots dr_1 \prod d\theta_{ij}.$$

The marginal likelihood for $\boldsymbol{\beta}$ is proportional to the above integral and can be written

$$L_4(\boldsymbol{\beta}) \propto \exp\left(\boldsymbol{\beta} \sum_1^k \mathbf{s}_i\right) \times \sum_{a_k=1}^{M_k} \sum_{a_{k-1}=a_k+1}^{M_{k-1}} \cdots \sum_{a_2=a_3+1}^{M_2} \prod_{i=1}^{k} \left[(M_i - a_{i+1} - 1)^{(a_i - a_{i+1} + 1)} \Big/ \left\{\sum_{l \in R(t_{(i)})} \exp(\boldsymbol{\beta}\mathbf{z}_l)\right\}^{a_i - a_{i+1}}\right], \qquad (15)$$

where $M_i = m_i + \ldots + m_k$ and $a_{k+1} = 0$. Expression (15) is also suitable if censored data are included. The notation $M^{(a)}$ in (15) refers to $M!/(M-a)!$

## REFERENCES


ANDERSEN, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimates. *J.R. Statist. Soc.* B **32**, 283–301.

BARNARD, G. A. (1963). Some aspects of the fiducial argument. *J. R. Statist. Soc.* B **25**, 111–4.

BRESLOW, N. (1972). Contribution to the discussion of paper by D. R. Cox. *J. R. Statist. Soc.* B **34**, 216–7.

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

FRASER, D. A. S. (1968). *The Structure of Inference*. New York: Wiley.

KALBFLEISCH, J. D. & PRENTICE, R. L. (1972). Contribution to discussion of paper by D. R. Cox. *J. R. Statist. Soc.* B **34**, 215–6.

KALBFLEISCH, J. D. & SPROTT, D. A. (1970). Application of likelihood methods to models involving a large number of parameters (with discussion). *J. R. Statist. Soc.* B **32**, 175–208.

KAPLAN, E. L. & MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Statist. Ass.* **53**, 457–81.

OAKES, D. (1972). Contribution to the discussion of paper by D. R. Cox. *J. R. Statist. Soc.* B **34**, 208.


[*Received September* 1972. *Revised December* 1972]