INTERVAL CENSORED SURVIVAL DATA: A GENERALIZED LINEAR MODELLING APPROACH

C. P. FARRINGTON

PHLS Statistics Unit, 61 Colindale Avenue, London NW9 5EQ, U.K.

SUMMARY

A method is described for weak parametric modelling of arbitrarily interval censored survival data using generalized linear models. The method makes use of an associated Bernoulli model, with standard errors based on the observed information matrix. Three types of models are discussed: additive and multiplicative hazard models with piecewise constant baseline hazard, and a proportional hazards model with discrete baseline survivor function. These models may be fitted in the statistical package GLIM.

1. INTRODUCTION

Observations are said to be interval censored when they are known only to lie in some interval between $a$ and $b$. Left and right censoring may be regarded as particular cases of interval censoring, in which $a = 0$ or $-\infty$ (left censoring) or $b = +\infty$ (right censoring). Interval censoring of survival data arises in many situations. To give but one important example, repeat testing for occult events (such as tumour genesis or asymptomatic infection) gives rise to censored data, the censoring intervals being determined by the times at which testing is undertaken.

Estimation from interval censored data has been considered by many authors. In the context of survival data, Peto and Turnbull describe non-parametric estimation from a single sample. A variety of procedures have been proposed for estimating and comparing survival curves, including the proportional hazards model of Finkelstein and other methods involving, for instance, the EM algorithm.

Although interval censoring arises naturally in many contexts, especially in biology and medicine, commonly available statistical packages do not appear to allow statistical modelling of such data other than by fully parametric methods. This contrasts with the variety of approaches described for fitting right censored survival data. A relatively straightforward procedure has been described for interval censored data when different individuals have the same observation times. However, this method does not apply when observation times differ between individuals. The purpose of this paper is to show how arbitrarily interval censored survival data may be cast in a weakly parametric generalized linear modelling framework. Fitting the models can then be undertaken in standard statistical packages such as GLIM. The method makes use of an associated Bernoulli model, an approach which has been used for fitting logistic regression models to right censored data.

The next sections describe the associated Bernoulli model and appropriate parametrizations, leading up to the proportional hazards model for interval censored data. These are followed by a section on implementation in GLIM.
2. THE LIKELIHOOD AND ASSOCIATED BERNOULLI MODEL

Suppose that time to a defined event for patients \( i = 1, \ldots, n \) with fixed (that is, time independent) covariates \( x_i \) is censored into intervals \( A_i = (a_i, b_i] \), where \( b_i \) may equal \(+\infty\) (in which case the observation is right censored). It is assumed that time to the event \( T \) is measured from \( T = 0 \). The data thus consist of triplets \( (a_i, b_i, x_i) \). It is assumed throughout that the event of interest occurs with probability 1, and that the censoring mechanism is non-informative. We let \( S(t|x) \) and \( \lambda(t|x) \) denote the survival and hazard functions at time \( t \), given covariates \( x \). The likelihood function is then:

\[
L = \prod_{i=1}^{n} \{ S(a_i|x_i) - S(b_i|x_i) \}.
\]

We suppose without loss of generality that the data are ordered in such a way that the first \( l \) observations are left censored \( (a_i = 0) \), the next \( r \) are right censored \( (b_i = +\infty) \), and the final \( c \) observations are confined \( (0 < a_i < b_i < +\infty) \), with \( n = l + r + c \). The likelihood may be written:

\[
\prod_{i=1}^{l} \left\{ 1 - S(b_i) \right\} \prod_{i=l+1}^{l+r} S(a_i) \prod_{i=l+r+1}^{n} S(a_i) \left\{ 1 - S(b_i)/S(a_i) \right\}
\]

where \( S(\cdot) = S(\cdot|x_i) \). This may be written as a particular realization of an associated model involving \( n + c \) independent Bernoulli trials with probability \( p_i \), response \( r_i \), and covariate vector \( z_i \), where \( i = 1, \ldots, n + c \). The associated model is defined as follows. Consider the \( i \)th observation; if it is left censored, it contributes one Bernoulli trial, with \( p_i = 1 - S(b_i) \), \( r_i = 1 \), \( z_i = x_i \). If, on the other hand, the observation is right censored, it also contributes one Bernoulli trial, with \( p_i = 1 - S(a_i) \), \( r_i = 0 \), \( z_i = x_i \), and \( B_i = (0, a_i] \). Finally, if the \( i \)th observation is confined (with \( i \) therefore lying between \( n - c + 1 \) and \( n \)), it contributes two Bernoulli trials, indexed \( i \) and \( c+i \). The first has \( p_i = 1 - S(a_i) \), \( r_i = 0 \), \( z_i = x_i \), \( B_i = (0, a_i] \), while the second has \( p_{c+i} = 1 - S(b_i)/S(a_i) \), \( r_{c+i} = 1 \), \( z_{c+i} = x_i \), and \( B_{c+i} = (a_i, b_i] \). It is easily seen that:

\[
L = \prod_{i=1}^{n+c} p_i^r (1 - p_i)^{1-r_i}.
\]

Thus maximizing the log-likelihood is equivalent to maximizing the log-likelihood of the associated model. The probabilities \( p_i \) may be written as follows:

\[
p_i = 1 - \exp \left\{ - \int_{B_i} \lambda(t|x_i) \, dt \right\}.
\]

In the next sections, various parameterizations are considered. For simplicity, the distinction between \( z_i \) and \( x_i \) will be dropped.

3. MODELS FOR THE HAZARD FUNCTION

The baseline hazard \( \lambda_0(t) = \lambda(t|x_i = 0) \) is assumed piecewise constant on time intervals \((t_{j-1}, t_j]\), \( j = 1, \ldots, k \), with \( t_0 = 0 \) and \( t_k = +\infty \). Let \( \lambda_j > 0 \) be the value of the baseline hazard in the \( j \)th time interval. For each \( i = 1, \ldots, n + c \) and \( j = 1, \ldots, k \) define the time segments \( c_{ij} = |B_i \cap (t_{j-1}, t_j]| \), and \( u_i = |B_i| \), the length of interval \( B_i \).

3.1. Additive model

Consider first an additive model. The hazard function is:

\[
\lambda(t|x_i) = x_i^T \alpha + \lambda_j \quad \text{if } t_{j-1} < t \leq t_j
\]
and the probabilities $p_i$ may be related to a linear predictor via the link function:

$$\eta_i = -\ln(1 - p_i) = x_i^T \alpha u_i + \sum_{j=1}^k \lambda_j c_{ij}.$$ 

This defines a generalized linear model with covariates $x_i u_i$ and $c_i$. Some care in fitting this model is required to ensure that all linear predictors remain positive, even if the maximum likelihood estimates are strictly positive. A method which works well using GLIM is described later.

Since the link function is not canonical, the observed and expected information matrices differ. The expected information matrix is calculated with respect to the associated Bernoulli model, and is therefore inappropriate. The expected information with respect to the original model depends on the censoring process, which in general is not known. However, standard errors may be estimated consistently using the observed information matrix $X'VX$ where $V = \text{diag}(r_i(1 - p_i)/p_i^2)$ and $X$ is the full covariate matrix, including the $x_i u_i$ and $c_i$. Note that $V_i = 0$ when $r_i = 0$, so that right censored observations contribute no information in the additive model, though they do affect the maximum likelihood estimates. The observed information matrix may be inverted in GLIM by means of a supplementary regression.

### 3.2. Multiplicative model

Consider now the multiplicative model. This is:

$$\lambda(t|x_i) = \exp(x_i^T \alpha) \lambda_j \text{ if } t_{j-1} < t \leq t_j.$$ 

This model can only partly be linearized using the complementary log-log link:

$$\eta_i = \ln \left\{ -\ln(1 - p_i) \right\} = x_i^T \alpha + \ln \left( \sum_{j=1}^k \lambda_j c_{ij} \right).$$

The model, however, is quite easy to fit using generalized linear modeling software by exploiting the fact that it may be written:

$$-\ln(1 - p_i) = \sum_{j=1}^k \lambda_j c_{ij} \exp(x_i^T \alpha)$$

and hence that, conditional on $\alpha$, it reduces to an additive model with no fixed covariates and time segments $c_i^* = c_i \exp(x_i^T \alpha)$.

As for the additive model, the standard errors based on the expected information matrix relative to the associated model are incorrect. The asymptotically correct standard errors may be obtained from the observed information matrix. As it turns out, this matrix may be expressed in a way that enables the required matrix manipulations to be performed within GLIM. Thus, let $D = [A, B]$ denote the matrix of derivatives with $A_{ij} = \delta \eta_i / \delta x_j$, $j = 1, \ldots, s$ and $B_{ij} = \delta \eta_i / \delta \lambda_j$, $j = 1, \ldots, k$, and $U$ and $V$ the diagonal matrices:

$$U_i = r_i(1 - p_i) \{ \ln(1 - p_i)/p_i \}^2 + (r_i - p_i) \{ \ln(1 - p_i)/p_i \}$$

$$V_i = \{ \ln(1 - p_i)/p_i \}^2 r_i(1 - p_i).$$

After a little algebra it can be seen that the observed information matrix $I$, namely minus the second derivative of the log-likelihood of the associated model, may be obtained from $D^T UD$ by replacing its lower right $k \times k$ submatrix by $B^T VB$. In view of this fact, $I$ may be inverted within GLIM by means of supplementary regressions as described later.
4. A MODEL FOR THE SURVIVOR FUNCTION

In practice, it is often possible to regard the process as occurring in discrete time. The baseline survivor function $S_0(t)$ may then be modelled as a discrete function with mass at a subset of the ordered censoring times $t_j$. This leads to the proportional hazards model. Define:

$$\theta_j = \ln \{S_0(t_{j-1})/S_0(t_j)\}.$$  

The $\theta_j$ induce a parameterization of the survivor function $S_0(t_j)$, and $\theta_j \geq 0$. Under the proportional hazards model,

$$S(t_j|x) = \{S_0(t_j)\}^{\exp(x^T \alpha)}$$

and hence the associated model may be written:

$$\ln \{ - \ln (1 - p_i) \} = x_i^T \alpha + \ln \left( \sum_{j=1}^k \theta_j d_{ij} \right)$$

where $d_{ij}$ equals 1 if $t_j \in B_i$ and 0 otherwise. This model is fitted in exactly the same way as the multiplicative model in the previous section.

Fitting a model with a parameter for each censoring time is usually unrealistic, and some reduction is necessary. The following procedure may be used. First, find a minimal subset $M$ of the censoring times, such that $B_i \cap M$ is non-empty for all $i$, and fit the model using covariates $d_i$ derived from $M$. Then expand $M$ until no further improvement of the model deviance is possible, subject to the $\theta_j$ remaining positive. Note that the maximum likelihood estimators $\theta_j$ under $M$ must be non-negative, since for each $j$ there is an $i$ with $B_i \cap M = \{t_j\}$ and hence $\eta_i = x_i^T \alpha + \ln(\theta_j)$.

Note finally that the fitting procedure described for the survivor function model is valid only when the set of possible censoring times is finite. Otherwise the number of parameters $\theta_j$ may increase indefinitely with the sample size thus calling into question the asymptotic validity of the estimation method. In practice censoring times are often rounded or grouped, and hence may be regarded as discrete, so that the method described remains valid.

5. FITTING IN GLIM

The statistical package GLIM provides a versatile environment in which to fit generalized linear models and their extensions. The present discussion refers to the fourth release. The manipulations required to define and fit the associated model are readily performed using GLIM's extensive macro facility. An example of some of the procedures described below is provided in the Appendix. General interactive macros which handle variable numbers of covariates and time intervals are available from the author. These macros set up the environment for the associated Bernoulli model and fit the three models described in the paper.

5.1 Additive model

As mentioned previously, one difficulty in fitting the additive model is that some of the linear predictors may become negative during the iteration. Two steps are required to circumvent this problem. First, negative values of the vector of linear predictors should be replaced by small positive quantities. This allows iteration to continue, although a check should be kept of the number of corrected values, a final non-zero count indicating that the model is inappropriate. In addition, for iteration to converge it may be necessary to control the step size. GLIM uses Fisher
scoring to update the parameter estimates, in which the step vector is $$(X^T WX)^{-1} X^T W \xi$$, where $X$ is the full matrix of covariates including $x_i u_i$ and the $c_i$, $W$ is the diagonal matrix of iterative weights corresponding to the associated model, and $\xi_i = (r_i - p_i) d\eta_i / dp_i$. The step size may therefore be reduced by a factor $f$ by scaling the covariates by $f$. The linear predictors must then be corrected by a factor $1/f$, which is possible in GLIM since the vectors $\eta$ and $X^T \beta$, where $\beta = (\alpha, \lambda)$, are held in different system structures, %eta and %lp, respectively. The convergence tolerance %cc must also be reduced by the factor $f$ to retain the same degree of accuracy of estimation. Note, however, that the number of cycles required for convergence will increase.

The correct standard errors may be obtained as follows. Having obtained the maximum likelihood estimate $\hat{\beta} = (\hat{\alpha}, \hat{\lambda})$ and linear predictor $\eta(\hat{\beta})$ with the associated model, a further iteration step is used to obtain the correct estimated variance–covariance matrix $(X^T V(\hat{\beta}) X)^{-1}$ described in Section 3 by weighted normal errors regression on the full covariate matrix $X$ with scale factor 1 and weight vector $V = \text{diag}(r_i(1 - \hat{p}_i)/\hat{p}_i^2)$. This additional regression step is simply a device to perform the required matrix algebra. The dependent variable used in this step is immaterial, since it does not affect the variance–covariance matrix. The most convenient choice is $\eta(\hat{\beta})$ since the $\hat{\theta}$ are then the parameter estimates returned by GLIM.

5.2. Multiplicative model

For the first step, $\alpha$ is set to zero and an initial estimate of $\lambda$ is obtained using the additive model. The multiplicative model is then fitted with offset $\ln(x_j c_j)$ to obtain a revised value for $\lambda$. The additive model is then refitted with covariates $c_i u_i = c_i \exp(x_i^T \hat{\alpha})$ to update the $\hat{\lambda}_j$, the cycle continuing until convergence. The fitting procedure is similar to the relaxation method used for fitting Weibull and piecewise exponential models in GLIM.

To calculate the standard errors, the matrix $B^T V B$ is obtained as for the additive model by setting up an additional regression with covariates $\delta \xi / \delta \lambda_j$ and weights $V_i$ evaluated at the maximum likelihood estimates. The matrix $B^T V B$ is saved prior to inversion via the GLIM structure %tri called by the $\text{load}$ command. Matrix $D^T U D$ is obtained in a similar manner and the $\text{load}$ command is used to replace the relevant submatrix by $B^T V B$. The standard errors calculated by GLIM are then based on the correct inverse observed information matrix $I^{-1}$. Note that these regressions are purely devices to perform the required matrix operations, and the dependent variables used are immaterial as previously explained. These procedures all require $\text{method}$ to be set to Gauss–Jordan.

6. BREAST CANCER

Finkelstein and Wolfe give data on months to breast retraction in 94 patients with breast cancer, following primary radiotherapy and adjuvant chemotherapy (48 patients) or primary radiotherapy alone (46 patients). Patients were followed up at clinic visits, generating observations as follows: 5 left censored where retraction occurred prior to the first follow-up visit; 38 right censored where no retraction had occurred by the final follow-up visit; and 51 confined, in which retraction occurred between two successive follow-up visits.

To illustrate the additive and multiplicative models, the baseline hazard was modelled as piecewise exponential with time intervals $0-5$, $5-10$, $10-15$, $15-20$, $20-25$, $25-30$, $30-35$, $35-40$, and $40-60$ months. Narrower intervals after 40 months produced negative baseline hazards in this region. The fixed covariate $x_i$ was 0 (radiotherapy alone) or 1 (radiotherapy and chemotherapy), fitted in the additive model as $x_i u_i$ where $u_i = \text{length of interval } B_i$. Starting fitted values were 0.9 if $r_i = 1$, 0.1 if $r_i = 0$. With the step reduction factor set to 10 and the convergence
Table I. Breast cancer data: model deviances and treatment effects

<table>
<thead>
<tr>
<th></th>
<th>Hazard function models</th>
<th>Survivor function model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Additive</td>
<td>Multiplicative</td>
</tr>
<tr>
<td>Null deviance</td>
<td>294.64</td>
<td>294.64</td>
</tr>
<tr>
<td>$\alpha$ (s.e.)</td>
<td>0.01217</td>
<td>0.9409</td>
</tr>
<tr>
<td>Treatment model deviance</td>
<td>291.23</td>
<td>283.76</td>
</tr>
<tr>
<td>Change in deviance</td>
<td>3.39</td>
<td>10.88</td>
</tr>
<tr>
<td>$p$</td>
<td>0.07</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Figure 1. Breast cancer data: hazard function for the additive model
--- radiotherapy only  --- radiotherapy and chemotherapy

criterion set to 0.00001, the fitting procedure for the additive model eventually entered the feasible region (all linear predictors positive) at cycle 55 and converged at cycle 93. The fitting procedure for the multiplicative model converged after 7 complete cycles. The results of the fits are summarized in Table I, and the hazard functions under each model are shown in Figures 1 and 2.
For the survivor function model, a minimal set $M$ of censoring times was chosen by listing the 145 intervals $B_i = (u_i, v_i]$ and choosing $t_i = \inf \{ v_i \}$, $t_r = \inf \{ u_i; u_i \geq t_{r-1} \}$. This is easily done by first sorting the $B_i$ in increasing nested order, respectively, of $v_i$ and $u_i$. For the breast cosmesis data this produced the set of points $M = \{4, 8, 12, 17, 23, 30, 34, 39, 48\}$. Stepwise inclusion of each of the 32 remaining censoring times produced no significant reduction in deviance without at least one $\theta_j$ becoming negative. The results of the fit using the set $M$ are given in Table I, and the survival curves are shown in Figure 3.

The likelihood ratio test statistics for the treatment effects are 3.27 for the additive hazard model, 10.88 for the multiplicative hazard model, and 8.43 for the survivor function model, all on one degree of freedom. The two proportional hazards models produce similar results, and the likelihood ratio statistic may be compared with the score statistic $Z = 2.86$ obtained by Finkelstein. Wald tests based on the estimates and their standard errors lead to similar inferences.

In this example, since the emphasis is on estimating the relative rather than the absolute hazards, the survivor function model is the most relevant. Using this model we may conclude that the hazard of breast retraction is increased if adjuvant chemotherapy is administered with primary radiotherapy, compared with primary radiotherapy alone ($p = 0.004$, relative hazard 2.3, 95 per cent confidence interval 1.3 to 4.1).
7. DISCUSSION

The aim of this paper was to show how interval censored survival data may be modelled without strong parametric assumptions using methods developed for generalized linear models. The advantage of this approach is that the models may be fitted using standard generalized linear modelling software. It is hoped that the flexibility which this entails will lead to greater use of weakly parametric models for this type of data. Although the emphasis throughout has been on implementation using GLIM, a similar approach could work in principle with other software designed to fit generalized linear models, provided it possesses an adequate macro facility. This should be sufficiently flexible to permit the iterative fitting of the multiplicative models, and to make the required adjustments to linear predictors and step size for the additive model. The latter requirement may not always be necessary provided good starting values can be found. Additional software may be needed to perform the matrix algebra involved in obtaining the correct standard errors, although in their absence inferences can be made by comparing deviances in nested models.

Choosing between the three models presented is to some extent guided by the problem in hand. Thus if the hazards are of intrinsic interest the hazard function models should be used, whereas if the emphasis is primarily on estimating the effect of treatment on survival the survivor function model is preferable. One shortcoming of the methods described is the range restriction on the

Figure 3. Breast cancer data: survivor function for the proportional hazards model
--- radiotherapy only ---- radiotherapy and chemotherapy
parameters \( \lambda_j \) and \( \theta_j \), which are required to be non-negative. However, negative estimates sometimes result from an inappropriate parameterization or sparse data, and in such cases may be handled by omitting or combining time intervals. Alternatively a more formal constrained maximum likelihood approach may be used.\(^{14}\) Further work is required on developing appropriate diagnostics to assess model fit, in GLIM or otherwise.

Throughout, only fixed covariates were considered. However, the hazard function models may easily be extended to incorporate piecewise constant time-dependent covariates. The additive model is fitted in much the same way with additional terms in the linear predictor. The multiplicative model requires one or more additional steps in the fitting procedure, and correct standard errors are no longer easily obtainable within GLIM. The methods of this paper may also be extended to incorporate left truncated data. This is achieved by replacing the survivor function \( S(t|x_i) \) by its conditional version \( S(t|x_i)/S(t_{c(i)}|x_i) \) where \( t_i \) is the left truncation time for the \( i \)th observation, and altering the intervals \( B_i \) accordingly, the \( t_i \) being assumed independent of the time to event and the covariates. It is not clear however whether the methods can be adapted to handle right truncation.

**APPENDIX**

General macros to fit all three models described in the paper are available from the author. As an illustration, this appendix contains the GLIM file to fit the multiplicative model with one fixed covariate and two time intervals. It is assumed that the environment for the associated model has been defined with standard length \( n + c \) and the following variables: binary response \( r \); fixed covariate \( x \); time segments \( c_1 \) and \( c_2 \). The step reduction factor is held in \%f.

```glim
$\text{str}$
$\text{scal} \%c = \%cc/\%f$

$\text{ml Xse}$
$\text{ml2 cl + c2e}$

$\text{alk}$
$\text{pred} \%fv = 1 - \exp(-\%beta): \%dr = 1/\exp(-\%beta) \%e}$

$\text{models}$

$\text{cycle}$
$\text{cal z = \exp (pr) *\%f}$

$\text{y}$
$\text{rser b n$\text{il o alk}$me * pred$\text{ini adin}$

$\text{cy}$
$\text{100 50 %c$\text{f (#m2)} .z - 1}$

$\text{ext}$
$\text{pe$\text{cal lamda = %pe: \%y = \%dv}$

$\text{cal}$
$\text{w = log(\%beta/\%exp(pr))}$

$\text{li c$\text{o w$\text{me $f \#m1 - 1}$

$\text{ext}$
$\text{pe$\text{cal alpha = %pe$}

$\text{cal}$
$\text{pr = \%beta - \%cos: \%f = \%fv$o}$

$\text{cal}$
$\text{gb = \%y - \%dv: \%a = \%if ((\%b > 0.005), 1, 0):\%y = \%dv}$

$\text{tril}$
$\text{cal tri = %tril}$

$\text{sub}$
$\text{cal w = r* (1 - \%fv) * (\%log(1 - \%fv)/\%fv)**2$}

$\text{y}$
$\text{n$\text{ser n$\text{li i$w w$\text{me J$\text{sc 1$\text{o a tri 1}$}

$\text{f (#m2). u - 1$as$\text{tmp = 1, 1, 1, 0, 0, O}$

$\text{pi tri tri$\text{mp}$e}$

$\text{tril}$
$\text{as i = 3, 5, 6$cal %tril(i) = tri$\text{e}$}

$\text{mult}$
$\text{v 1 alpha:2 lamda}$
```

This fits the multiplicative model by iterating between fits of the time and fixed covariates macros for altering the triangle prior to inversion initialises
$\alpha = 0; \beta = 0; n = 1; \% \alpha = 1$

$\alpha \; \text{and calls}$

$\beta = (r = 1), \; 0.9, \; 0.1)$

$\text{the fitting}$

$\% \beta$ cycle$\%$

$\text{algorithm}$

$e^{-0.5}(1-r) \times (1-\beta)$

$\text{inverts the}$

$\% \beta$ cycle$\%$

$\text{correct}$

$\% \beta$ cycle$\%$

$\text{information}$

$u = \alpha, \lambda$.

$\text{matrix}$

$w = \text{exp}(u) \times \beta$

$\text{prints correct}$

$\% \beta$ cycle$\%$

$\text{deviance and}$

$u = \alpha, \lambda$

$\text{estimates}$

For the breast cancer data, $n + c = 145$. Using the intervals 0–20 and 20–60 months and a step reduction factor $\% \beta = 10$, the following output is routed to the transcript file:

In final additive fit 0. values
of %lp were out of range
model deviance is 293.3

<table>
<thead>
<tr>
<th>PAR</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.89743</td>
</tr>
<tr>
<td>2</td>
<td>0.01076</td>
</tr>
<tr>
<td>3</td>
<td>0.02391</td>
</tr>
</tbody>
</table>

REFERENCES