

Mimicking Cox-regression

Sven Ove Samuelsen

Department of Mathematics, University of Oslo
P.O. Box 1053 Blindern, 0316 Oslo, Norway

email: osamuels@math.uio.no

Abstract

A class of objective functions related to the Cox partial likelihood is proposed. Examples of the objective functions are applied to binary data with a log-link. It is pointed out that the Peto-Breslow approximation to the partial likelihood with discrete failure times fits a conditional model with a log-link.

1 Introduction

The immensely popular proportional hazard model, Cox (1972), is given by

$$\lambda_i(t) = \exp(\beta'Z_i)\lambda_0(t)$$

where $\lambda_i(t)$ is the hazard for individual i , $\lambda_0(t)$ a baseline hazard, β and Z_i respectively vectors of regression parameters and covariates for individual i . Letting X_i and D_i be respectively the possibly right-censored failure times and indicators of failure for individual i and \mathcal{R}_t the number of individuals at risk right before t Cox' method consists in maximizing the partial likelihood, (Cox, 1975),

$$L_P = \prod_{i=1}^n \left\{ \frac{\exp(\beta'Z_i)}{\sum_{j \in \mathcal{R}_{X_i}} \exp(\beta'Z_j)} \right\}^{D_i}. \quad (1.1)$$

The partial likelihood property requires that no failure times are equal. With truly discrete survival data Cox (1972) instead suggested fitting the conditional logistic model

$$\frac{\lambda_{it}}{1 - \lambda_{it}} = \exp(\beta'Z_i) \frac{\lambda_{0t}}{1 - \lambda_{0t}}$$

⁰*Key words:* Discrete survival data, Log-linear models, Matched sets, Nuisance parameters, Objective functions, Proportional hazard model, Unbiased estimating functions

where λ_{it} is the conditional probability that an individual with covariate Z_i will fail at t given survival up to t . The conditional likelihood given the total number of failures at each time t does not depend on the baseline λ_{0t} and has the same form as a conditional likelihood for logistic regression. It coincides with the usual Cox-likelihood when there is only one failure at each failure time and suggests an approach to dealing with tied survival data.

However this model does not correspond to a proportional hazard model when the ties arise only because the continuous time scale has been measured to crudely. Also direct evaluation of the conditional likelihood can be numerically exhaustive. (An clever algorithm for efficient evaluation has been developed, see Gail, Lubin & Rubinstein (1981) and Howard (1972), but is sometimes not implemented in software packages.)

Other approximations have been suggested by Peto (1972), Efron (1977) and Oakes (1981). The most commonly applied, but also most criticized of these approximations is that of Peto. The objective function (1.1) actually covers this approach. The approaches of Efron and Oakes on the other hand takes into consideration removal of failures from the risk set. The approximation of Peto was also considered by Breslow (1974) and is for this reason often referred to as the Breslow approximation. Here we will use the term Peto-Breslow approximation.

The Peto-Breslow approximation is the inspiration for this paper. It will be shown that under a truly discrete model it actually fits a conditional model with log-link, that is, it fits the model

$$\lambda_{it} = \exp(\beta' Z_i) \lambda_{0t}, \tag{1.2}$$

contrary to the logit link suggested by Cox (1972).

Example

Consider the extreme situation that survival is only inspected at one time point. Also assume that there is only one covariate indicating presence or absence of exposure. Denote the number of exposed and nonexposed failures as respectively A and B and the corresponding number of nonfailures or censored observations as C and D . Applying the Peto-Breslow approximation amounts to maximizing the function

$$\exp(\beta A) / \{(B + D) + (A + C) \exp(\beta)\}^{A+B}.$$

This leads to the estimate $\hat{\beta}$ given by

$$\exp(\hat{\beta}) = \frac{A/(A + C)}{B/(B + D)},$$

that is $\hat{\beta}$ is log of the ratio of frequencies of failure among exposed and unexposed.

When extending the example to (1.2) it turned out that the argument could be presented most easily within a much larger framework and that it suggests a wide class

of objective function with unbiased estimating equations. The objective functions are presented in the next section. In Section 3 we again specialize to binary outcomes, first in a basic case, secondly as discrete censored survival data and third in a matched sets framework. In the final section other applications of the objective function are discussed.

2 The objective functions

The conditional likelihood for independent Poisson-distributed data Y_1, \dots, Y_n with expectations $EY_i = \lambda_i = \lambda_0 \exp(\beta'Z_i)$ given $Y_\bullet = \sum_{i=1}^n Y_i$, is proportional to

$$L_C = \prod_{i=1}^n \left\{ \frac{\exp(\beta'Z_i)}{\sum_{j=1}^n \exp(\beta'Z_j)} \right\}^{Y_i} = \prod_{i=1}^n \left(\frac{EY_i}{\sum_{j=1}^n EY_j} \right)^{Y_i}. \quad (2.1)$$

The partial likelihood L_P is essentially a product of a such conditional likelihoods. The objective function studied in this paper is in its simplest form identical to the right-hand side of (2.1), namely

$$M_0 = \prod_{i=1}^n \left[\frac{EY_i}{\sum_{j=1}^n EY_j} \right]^{Y_i}, \quad (2.2)$$

however the restriction that Y_i is Poisson distributed is removed. It is only assumed that $EY_i > 0$ and depend on some parameters of interest. This generality is sufficient for unbiased estimating equations as shown below. In the examples we will however only consider log-linear structures $EY_i = \exp(\alpha + \beta'Z_i)$.

Taking the derivative of $\log M_0$ with respect to these parameters we obtain

$$U_0 = \partial \log M_0 = \sum_{i=1}^n Y_i \left(\frac{\partial EY_i}{EY_i} - \frac{\sum_{j=1}^n \partial EY_j}{\sum_{j=1}^n EY_j} \right).$$

Then taking the expectation at the true value of the parameter gives

$$EU_0 = \sum_{i=1}^n EY_i \left(\frac{\partial EY_i}{EY_i} - \frac{\sum_{j=1}^n \partial EY_j}{\sum_{j=1}^n EY_j} \right) = 0, \quad (2.3)$$

thus the estimating equations $U_0 = 0$ are unbiased. Under regularity assumption they yield consistent and asymptotically normal estimators. In general these score equations will not have the property $\text{var}(U_0) = -E\partial U_0$ and they will lead to estimators with variance $\Sigma = \Gamma^{-1}\Delta\Gamma^{-1}$ where $\Gamma = -E\partial U_0$ and $\Delta = \text{var}(U_0)$.

Efficiency may be improved on by weighting, replacing Y_i in (1.3) by $Y_i' = w_i Y_i$ which amounts to solving

$$U_{0w} = \sum_{i=1}^n w_i Y_i \left(\frac{w_i \partial EY_i}{w_i EY_i} - \frac{\sum_{j=1}^n w_j \partial EY_j}{\sum_{j=1}^n w_j EY_j} \right). \quad (2.4)$$

This leads to the question of how the weights should be chosen. A somewhat heuristic approach will be adopted. Since (2.2) is a conditional likelihood for independent Poisson data it is possible that choosing w_i such that $EY_i' = \text{var}(Y_i')$ may give improved estimators. This leads to $w_i = EY_i/\text{var}(Y_i)$. In the appendix it is shown that with $w_i = EY_i/\text{var}(Y_i)$ evaluated at the true parameters of the model the likelihood property $\text{var}(U_{0w}) = -E\partial U_{0w}$ is gained.

In practice these weights are of course unknown. To get around this one may first fit (2.2) without weights. Additional nuisance parameters will also be necessary to estimate like the α under the log-linear model. The w_i are then replaced by \hat{w}_i where the estimated parameters are plugged in and Y_i in (2.2) is replaced by $Y_i' = \hat{w}_i Y_i$.

A generalization of M_0 is what is actually of interest. Let $Y_{it}, i = 1, \dots, n_t$ be random variables, \mathcal{F}_t some set of conditioning variables and E_t the operator taking conditional expectations given \mathcal{F}_t . Again unbiased estimating equations are obtained by solving

$$U = \sum_t \sum_{i=1}^{n_t} Y_{it} \left(\frac{\partial E_t Y_{it}}{E_t Y_{it}} - \frac{\sum_{j=1}^{n_t} \partial E_t Y_{jt}}{\sum_{j=1}^{n_t} E_t Y_{jt}} \right)$$

or, assumed equivalently, maximizing

$$M = \prod_t \prod_{i=1}^{n_t} \left(\frac{E_t Y_{it}}{\sum_{j=1}^{n_t} E_t Y_{jt}} \right)^{Y_{it}}. \quad (2.5)$$

Again improved estimates may be obtained replacing Y_{it} by $\hat{w}_{it} Y_{it}$ for estimated weights $\hat{w}_{it} = E_t \widehat{Y}_{it} / \widehat{\text{var}}_t Y_{it}$. Here var_t denotes conditional variance given \mathcal{F}_t . The main interest again concern log-linear structures

$$E_t Y_{it} = \exp(\alpha_t + \beta' Z_{it})$$

where Z_{it} are covariates and the α_t nuisance parameters or random components. Note that the α_t drops out of objective function and thus β may be estimated without making any assumptions on the nuisance parameters. This is the main reason for considering the objective function M .

3 Example: Binary data with log-link

In this section we will look at the general objective functions of the previous section under the assumption that outcomes are binary with log-linear probabilities of success. We shall look at three different cases, first a standard case, then discrete time survival data and third, matched sets with a common random intercept in each set.

3.1 The standard case

Let Y_i be independent binary data with $p_i = EY_i = \exp(\alpha + \beta'Z_i)$ where the Z_i are covariates. The objective function M_0 becomes

$$M_0 = \prod_{i=1}^n \left(\frac{EY_i}{\sum_{j=1}^n EY_j} \right)^{Y_i} = \prod_{i=1}^n \left\{ \frac{\exp(\beta'Z_i)}{\sum_{j=1}^n \exp(\beta'Z_j)} \right\}^{Y_i}.$$

The derivatives of $\log(M_0)$ becomes

$$U_0 = \frac{\partial \log(M_0)}{\partial \beta} = \sum_{i=1}^n Y_i \left\{ Z_i - \frac{\sum_{j=1}^n Z_j \exp(\beta'Z_j)}{\sum_{j=1}^n \exp(\beta'Z_j)} \right\}$$

and

$$I_0 = -\frac{\partial^2 \log(M_0)}{\partial \beta^2} = \sum_{i=1}^n Y_i \left[\frac{\sum_{j=1}^n Z_j^{\otimes 2} \exp(\beta'Z_j)}{\sum_{j=1}^n \exp(\beta'Z_j)} - \left\{ \frac{\sum_{j=1}^n Z_j \exp(\beta'Z_j)}{\sum_{j=1}^n \exp(\beta'Z_j)} \right\}^{\otimes 2} \right]$$

where $v^{\otimes 2} = vv'$ for any vector v . From this we obtain

$$\Gamma = EI_0 = \sum_{i=1}^n p_i \left[\frac{\sum_{j=1}^n Z_j^{\otimes 2} \exp(\beta'Z_j)}{\sum_{j=1}^n \exp(\beta'Z_j)} - \left\{ \frac{\sum_{j=1}^n Z_j \exp(\beta'Z_j)}{\sum_{j=1}^n \exp(\beta'Z_j)} \right\}^{\otimes 2} \right]$$

and

$$\Delta = \text{var}(U_0) = \sum_{i=1}^n p_i(1 - p_i) \left[Z_i - \frac{\sum_{j=1}^n Z_j \exp(\beta'Z_j)}{\sum_{j=1}^n \exp(\beta'Z_j)} \right]^{\otimes 2}.$$

Estimates $\hat{\Gamma}$ and $\hat{\Delta}$ are easily obtained inserting the solution $\hat{\beta}$ of $U_0 = 0$ for β and

$$\hat{\alpha} = \log \left\{ \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n \exp(\hat{\beta}'Z_i)} \right\}$$

for α . This estimator of α is an analogy to the Breslow estimator of the base-line hazard in Cox-regression. Note that $\hat{\Gamma}$ is identical to the observed information matrix inserted $\hat{\beta}$.

The estimate $\hat{\beta}$ can generally be obtained from a program for Cox-regression. To obtain the variance estimators $\hat{\Gamma}^{-1} \hat{\Delta} \hat{\Gamma}^{-1}$ one will have to write a separate routine. Note however that a lazy approximation of these variances, strictly valid only for $\beta = 0$, is obtained by multiplying the variances from the Cox-regression program by the proportion of individuals with $Y_i = 0$. Note also that not adjusting the variances, that is using those calculated by the Cox-regression program will always be conservative because

$$\Delta < \sum_{i=1}^n p_i \left\{ Z_i - \frac{\sum_{j=1}^n Z_j \exp(\beta'Z_j)}{\sum_{j=1}^n \exp(\beta'Z_j)} \right\}^{\otimes 2} = \Gamma.$$

An interesting feature of this procedure is that fitted values of p_i greater than one does not create a numerical problem. This would point to a problem with the model that might have been overlooked from the maximum likelihood estimates (MLE).

The preceding section suggests that weighting may improve efficiency and that the weights should be chosen close to the true value of $w_i = EY_i/\text{var}(Y_i) = 1/(1 - p_i)$. It can be shown that with the true p_i this leads to the same variance as the MLE. The obvious choice of weights is $\hat{w}_i = 1/(1 - \hat{p}_i)$. Denote the resulting estimator β^* . Since $\hat{\beta}$ and $\hat{\alpha}$ are consistent under weak assumption it also holds that the \hat{w}_i are similarly consistent for the w_i . Thus we may use the inverse of the observed information matrix as the variance estimator of β^* .

Example

A small simulation was set up with population size $n = 1000$, $\alpha = \log(0.01) = -4.605$, $\beta = \log(50) = 3.912$ and the covariates Z_i uniformly distributed. Thus we have $0.01 < p_i < 0.5$. The simulation was repeated 1000 times. The average unweighted estimate of β was 3.930 with empirical variance of 0.161, average of unadjusted variances (i.e. the inverse of the observed information $\hat{\Gamma}$) of 0.183 and average of adjusted variances of 0.155. For the weighted estimator of β the average was 3.924, the empirical variance 0.158 and the average variance estimate 0.150. The MLE was also calculated and the average of those β estimates were 3.927, the empirical variance 0.158 and the average variance estimate 0.146. For the estimates of α the averages were -4.627, -4.623 and -4.625 for the unweighted, the weighted and the MLE respectively. The corresponding empirical variances were 0.108, 0.107 and 0.107.

3.2 Discrete survival data

Assume that survival times T_i follow discrete distributions and let λ_{it} equal the conditional probability that an individual alive up to t will fail at that time. As mentioned in the introduction Cox (1972) suggested to model such data by a conditional logit model. Peto (1972) proposed using the partial likelihood (1.1) as an approximation also in the presence of moderately tied data. We shall here show that for discrete data maximizing the partial likelihood (1.1) amounts to fitting the log-linear model (1.2) for λ_{it} . If the λ_{it} are all small then logit-link and log-link models are also approximately equal. The link for λ_{it} that corresponds to a proportional hazard model is the complementary log-log link which is an intermediate case between the log-link and the logit-link (McCullagh & Nelder, 1989).

Let Y_{it} denote a failure of individual i at time t and assume that the failures only can take place at times $\tau_1, \tau_2, \dots, \tau_m$. Let \mathcal{R}_t denote the set of individuals under observation right before time t . Furthermore let \mathcal{F}_t be the available information right before time t , that is \mathcal{F}_t consist of \mathcal{R}_t , the outcomes $Y_{is}, s < t$ and the covariates

$Z_{is}, s \leq t$. We assume the model

$$P(Y_{it} = 1 | \mathcal{F}_t) = I(i \in \mathcal{R}_t) \lambda_{it} = I(i \in \mathcal{R}_t) \exp(\alpha_t + \beta' Z_{it}),$$

that is the censoring is independent with respect to the filtration \mathcal{F}_t (Andersen et al., 1993). Then applying the objective function (2.5) we obtain

$$M = \prod_t \prod_{i=1}^n \left(\frac{E Y_{it} | \mathcal{F}_t}{\sum_{j=1}^n E Y_{jt} | \mathcal{F}_t} \right)^{Y_{it}} = \prod_t \prod_{i \in \mathcal{R}_t} \left\{ \frac{\exp(\beta' Z_{it})}{\sum_{j \in \mathcal{R}_t} \exp(\beta' Z_{jt})} \right\}^{Y_{it}}$$

which is the Peto-Breslow objective function and coincides with (1.1). The product over t is of course a discrete product over the τ_j . The arguments of Section 2 shows that under model (1.2) this leads to unbiased estimating equations and under regularity assumptions on the censoring and the covariates to consistency and asymptotic normality of the maximizer $\hat{\beta}$. The baseline conditional probabilities may be estimated by the Breslow type estimator

$$\hat{\lambda}_{0t} = \frac{\sum_{i \in \mathcal{R}_t} Y_{it}}{\sum_{i \in \mathcal{R}_t} \exp(\hat{\beta}' Z_{it})} \quad (3.1)$$

With $\hat{\lambda}_{it} = \exp(\hat{\beta}' Z_{it}) \hat{\lambda}_{0t}$ the covariance matrix of $\hat{\beta}$ is estimated by $\hat{\Gamma}^{-1} \hat{\Delta} \hat{\Gamma}^{-1}$ where

$$\hat{\Delta} = \sum_t \sum_{i \in \mathcal{R}_t} \hat{\lambda}_{it} (1 - \hat{\lambda}_{it}) \left\{ Z_{it} - \frac{\sum_{j \in \mathcal{R}_t} Z_{jt} \exp(\hat{\beta}' Z_{jt})}{\sum_{j \in \mathcal{R}_t} \exp(\hat{\beta}' Z_{jt})} \right\}^{\otimes 2}$$

and

$$\hat{\Gamma} = \sum_t \sum_{i \in \mathcal{R}_t} \hat{\lambda}_{it} \left[\frac{\sum_{j \in \mathcal{R}_t} Z_{jt}^{\otimes 2} \exp(\hat{\beta}' Z_{jt})}{\sum_{j \in \mathcal{R}_t} \exp(\hat{\beta}' Z_{jt})} - \left(\frac{\sum_{j \in \mathcal{R}_t} Z_{jt} \exp(\hat{\beta}' Z_{jt})}{\sum_{j \in \mathcal{R}_t} \exp(\hat{\beta}' Z_{jt})} \right)^{\otimes 2} \right].$$

Software for Cox-regression will fit this model, but the conservative $\hat{\Gamma}^{-1}$ will be reported for the covariance matrix. A lazy adjustment to this estimate, valid when $\beta = 0$, can consist of multiplying $\hat{\Gamma}^{-1}$ by $1 - \sum_t \sum_{i \in \mathcal{R}_t} Y_{it} / \sum_t n_t$ where n_t is the number at risk right before time t .

Also in this case weighting that may improve efficiency is possible. Likewise Section 3.1 we may replace Y_{it} by $Y'_{it} = Y_{it} \hat{w}_{it}$ where $\hat{w}_{it} = 1/(1 - \hat{\lambda}_{it})$ and refit the model with these weights. This corresponds to maximizing

$$M^* = \prod_t \prod_{i \in \mathcal{R}_t} \left\{ \frac{\exp(\beta' Z_{it}) \hat{w}_{it}}{\sum_{j \in \mathcal{R}_t} \exp(\beta' Z_{jt}) \hat{w}_{jt}} \right\}^{Y_{it} \hat{w}_{it}}.$$

The variances reported in the examples below are obtained from the inverse of the observed information $-\partial^2 \log(M^*) / \partial \beta^2$ inserted the estimator β^* maximizing M^* .

Example

A small simulation study was carried out with $n = 1000$ individuals, two covariates Z_{1i} binary with outcome probability of 0.5, $Z_{2i} = Z_{1i} + V_i$ where V_i uniform on $(-0.5, 0.5)$, regression coefficients $\beta_1 = \beta_2 = 1$ and $\alpha_t = -2.5 - \log(2) - 0.1 * (t - 10)$ for $t = 1, \dots, 10$. Since $\max(Z_{1i} + Z_{2i}) \leq 2.5$ this gives $\lambda_{it} \leq 0.5$. Censoring was uniform on $\{1, \dots, 10\}$. Both the unweighted estimators $\hat{\beta}$, the weighted estimators β^* and the joint MLE of $(\alpha_1, \dots, \alpha_{10}, \beta_1, \beta_2)$ were computed. The simulation was repeated 1000 times.

There was no observable bias on either estimate of β_1 or β_2 . The empirical variance of $\hat{\beta}_1$ was 0.053 compared to an average of unadjusted variances of 0.054 and of adjusted variances of 0.049. The empirical variances of weighted estimates and the MLE's of β_1 was also 0.053 both with averages of variance estimates of 0.048. For β_2 all estimators had empirical variances of 0.035. The average unadjusted variance of $\hat{\beta}_2$ was 0.037 whereas the average adjusted variance was 0.031. The average variance estimates on β_2 for the weighted estimators and the MLE were respectively 0.031 and 0.030. Estimates of α_t did not show serious bias except for α_{10} . Likely this bias was caused by occasionally no failures for the maximal $t = 10$.

Example

Data from the Norwegian Study of Sexual Behavior in 1987 on age at first intercourse (Sundet et al., 1992) were reanalysed. Only the 3107 women aged 18 to 60 were included in the analysis. The event times were given in years. Since the main part of the population experienced their first intercourse between the early teens and mid twenties the data were severely tied, as a maximum age 16 was reported by 529 women. Since not everyone had experienced intercourse about 9% of the times were censored.

The data were analysed with the covariate $Z_i = \text{year of birth} / 30$. Thus Z_i is a transformation of the censoring time. First the data were analysed with a log-link. Using the Peto-Breslow method the estimate of β was 0.695 and standard error (se) reported from the program was 0.050. The adjusted se became 0.044. The weighted estimate β^* was 0.634 also with a se of 0.044. The MLE of β was 0.639 with a se of 0.042.

The MLE with a complementary log-log link was 0.755 with a se of 0.050. This corresponds well to the estimate obtained using the Efron method for tied data of 0.752 again with a se of 0.050. With a logit link the MLE became 0.881 with a se of 0.056. Numerical problems prevented computation of the conditional MLE, however an approximation suggested by Ryan et al. (1999) gave an estimate of 0.909.

In the methods so far discussed the event times have been considered stemming from a discrete distribution which is not very realistic. The Efron method can be seen as an exception in this respect, however also this method uses the convention in survival analysis of letting censored event times follow all those not censored. For this data set this also seems unrealistic. When originally analysing these data the Peto-Breslow estimates were compared to the simple approach of adding random

numbers between zero and one to all the possibly censored event times, thus giving a random ordering of the individuals. Applying this method to the data 100 times gave estimates of β ranging from 0.782 to 0.801 with a mean of 0.792. A shortcoming of this method is that it does not take into consideration that individuals with a high value of Z_i are more likely to have small event times. However drawing random numbers in accordance with the previous estimate of β again repeated 100 times gave the range 0.782 to 0.805 with a mean of 0.794. Although a very modest adjustment this was significantly higher.

3.3 Repeated measurements with random intercept

Consider $i = 1, \dots, n$ matched sets with $t = 1, \dots, T_i$ subjects in set no. i . Denote the binary outcome for subject t in set i by Y_{it} . Random intercept models for binary data with a logistic link have been discussed by for instance Diggle et al. (1994) using the conditional likelihood for logistic regression. Some alternative estimation methods are suggested by Ryan et al. (1999). Here it is instead assumed that

$$P(Y_{it} = 1|U_i) = p_{it} = \exp(U_i + \beta'Z_{it}) \quad (3.2)$$

where U_i is a random intercept for matched set i and the Z_{it} are covariates for subject t in set i . Then β may be estimated by maximizing

$$M = \prod_{i=1}^n \prod_{t=1}^{T_i} \left(\frac{p_{it}}{\sum_{s=1}^{T_i} p_{is}} \right)^{Y_{it}} = \prod_{i=1}^n \prod_{t=1}^{T_i} \left\{ \frac{\exp(\beta'Z_{it})}{\sum_{s=1}^{T_i} \exp(\beta'Z_{is})} \right\}^{Y_{it}}, \quad (3.3)$$

which is again an application of the general objective function (2.5) and hence leads to unbiased estimating equations and under regularity assumptions consistent and approximately normal estimators. Note that the random intercept U_i drops out of M and it is not necessary to specify a distribution for it. Note also that for a component of β to be identifiable with this objective function the corresponding component of Z_{it} must vary within the matched sets. Since M is also on the same form as Cox-likelihood it is easy to obtain the estimates $\hat{\beta}$.

With respect to estimation of variances note that, with E_{U_i} denoting the conditional expectation given U_i ,

$$\text{var} \left[\sum_{i=1}^n E_{U_i} \sum_{t=1}^{T_i} Y_{it} \left\{ Z_{it} - \frac{\sum_{s=1}^{T_i} Z_{is} \exp(\beta'Z_{is})}{\sum_{s=1}^{T_i} \exp(\beta'Z_{is})} \right\} \right] = 0,$$

and thus covariance matrix of the score becomes

$$\Delta = E \sum_{i=1}^n \sum_{t=1}^{T_i} p_{it}(1 - p_{it}) \left\{ Z_{it} - \frac{\sum_{s=1}^{T_i} Z_{is} \exp(\beta'Z_{is})}{\sum_{s=1}^{T_i} \exp(\beta'Z_{is})} \right\}^{\otimes 2}$$

while the expected information matrix can be written as

$$\Gamma = E \sum_{i=1}^n \sum_{t=1}^{T_i} p_{it} \left[\frac{\sum_{s=1}^{T_i} Z_{is}^{\otimes 2} \exp(\beta' Z_{is})}{\sum_{s=1}^{T_i} \exp(\beta' Z_{is})} - \left\{ \frac{\sum_{s=1}^{T_i} Z_{is} \exp(\beta' Z_{is})}{\sum_{s=1}^{T_i} \exp(\beta' Z_{is})} \right\}^{\otimes 2} \right]$$

which can be estimated inserting $\hat{\beta}$ for β and $\hat{p}_{it} = \exp(\hat{U}_i + \hat{\beta}' Z_{it})$ for p_{it} where $\hat{U}_i = \log\{\sum_t Y_{it} / \sum_t \exp(\hat{\beta}' Z_{it})\}$. With respect to Γ this will not differ from using Y_{it} in place of \hat{p}_{it} and is thus equivalent to using the observed information matrix. Again inverse of the observed information matrix is conservative estimate for the covariance matrix of $\hat{\beta}$, and it is possible to find lazy adjustments similar to those in Sections 3.1 and 3.2.

Also similar to the previous two subsections one might improve on the estimator by weighting the outcomes Y_{it} by $\hat{w}_{it} = 1/(1 - \hat{p}_{it})$. For the weighted estimator the observed information matrix will be used as the estimated covariance matrix in the example below.

Example

Similarly to Ryan et al. (1999) interpret Y_{it} as the indicator of an allergy attack for individual i at day t . The chance that an attack occur may depend on $Z_t =$ pollen level at day t and the indicator x_i that individual i is sensitive to pollen. With a model $P(Y_{it} = 1|U_i) = \exp(U_i + \beta_0 + \beta_1 x_i + \beta_2 Z_t + \beta_3 Z_t x_i)$ maximizing M will not give an estimate of β_1 since x_i is constant over time. Since we need not model the U_i it does not create a problem to include the effect of being sensitive into the random component, that is to let $U_i' = U_i + \beta_0 + \beta_1 x_i$. The effects of Z_t and of the interaction between pollen and being sensitive $Z_t x_i$ are however directly identifiable.

In this simulation Z_t was uniformly distributed on $\langle 0, 4 \rangle$, $P(x_i = 1) = 0.25$, and U_i was uniformly distributed on $\langle -3, 3 \rangle$. The parameters values were $\beta_0 = -7.2, \beta_1 = 0, \beta_2 = \beta_3 = 0.5$. The simulations were repeated 1000 times. The population size was set to 500 and the individuals were assumed to be observed for a full pollen season of 100 days. A full MLE could not be fitted since that model would require 500 individual parameters to be estimated. With this model many individuals will never experience an allergy attack and the proportion with no attacks varied from 51% to 65%. The average number of attacks among those with attacks varied from 2.5 to 4.6 and the maximum number of attacks from 14 to 35.

Both β_2 and β_3 were estimated without noticeable bias. The unweighted estimators had empirical variances of 0.00313 and 0.00590 for the estimators β_2 and β_3 respectively. The corresponding average of unadjusted variances were 0.00315 and 0.00674 whereas the average adjusted variances were 0.00305 and 0.00614. The weighted estimators had empirical variances of 0.00313 and 0.00586 whereas the average variance estimates were 0.00307 and 0.00610.

4 Discussion

This paper has established that using the Peto-Breslow approximation for tied survival data and interpreting the survival time distribution as discrete one is fitting a conditional model with a log-link instead of the logit link suggested by Cox (1972). It is also shown how standard errors of parameter estimates should be adjusted with the Peto-Breslow approximation.

The paper also presents a very general class of unbiased estimating equations which allow addressing a much wider class of problems than the examples on binary data. The estimating equations are likely of most interest when there are many nuisance parameters in a model that act multiplicatively on the expected values, that is under a model $EY_{ij} = U_i V_{ij}$ where the interest is on the part V_{ij} . It is well-known that when there are many nuisance parameters U_i the maximum likelihood estimates can be seriously biased. The examples have used log-linear structures $V_{ij} = \exp(\beta' Z_{ij})$, but could simply be modified to other risk functions $V_{ij} = \Psi(\beta' Z_{ij})$.

An extreme example could be the mixed model $Y_{ij} = \beta' Z_{ij} + U_i + \epsilon_{ij}$ where U_i and ϵ_{ij} are random coefficients usually assumed to be normally distributed. The conditional expectation of $Y'_{ij} = \exp(Y_{ij})$ given U_i equals $\exp(U_i + \sigma^2/2 + \beta' Z_{ij})$ if $\text{var}(\epsilon_{ij}) = \sigma^2$ and maximizing the objective function M in (2.5) for Y'_{ij} give consistent estimates of β . Under the traditional assumptions of this mixed model one would of course expect such estimators to be severely inefficient, but when they do not hold the picture is less clear. In particular if $\text{var}(Y'_{ij}) \propto EY'_{ij}$ one could expect good properties.

The paper also suggests to improve on the objective function by weighting by estimates of $EY'_{ij}/\text{var}(Y'_{ij})$. In the examples on binary data the improvements were only slight, but it appeared that this removes the need to adjust variances. It also seems that the weighting may lead to estimators that are closer to the MLE. This however is not necessarily an advantage since the MLE can be biased with many nuisance parameters. More research is needed to fully understand the impact of the weighting.

The method of the paper seems to have relations to some recently proposed methods. Lin et al. (1998) surprisingly arrive at Cox-regression for current status data under an additive hazard model $\lambda_i(t) = \lambda_0(t) + \beta' Z_i$. If survival \bar{D}_i of individual i is inspected at time X_i the model implies the loglinear structure $E\bar{D}_i = \exp\{-\Lambda_0(X_i) - \beta' Z_i X_i\}$ where $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ are nuisance parameters. Estimators of β , possibly different from those of Lin et al., may be constructed from the framework discussed in this paper. Since the methods of Lin et al. are not efficient such efforts may lead to improvement.

Self et al. (1991) discuss case-only studies, that is studies where data on genes are obtained only for cases of a disease and the parents of the cases. This way one may determine the distribution of genes for a hypothetical series of offspring of the parents. Self et al. suggest a log-linear model for the probability of disease for case no i that may be rewritten as $\exp(U_i + \beta' Z_i)$. Here U_i can be considered a random component

incorporating environmental influences and Z_i the genes of the case. Their objective function can be written on the form

$$M = \prod_i \frac{\exp(\beta' Z_i)}{E \exp(\beta' Z_i)} = \prod_i \left\{ \frac{E Y_i | U_i, Z_i}{E Y_i | U_i} \right\}^{Y_i}$$

where expectation is taken over the distribution determined by the genes of the parents. Similarly to (2.5) it can be shown that this objective function also leads to unbiased estimating equations.

Acknowledgements

This paper was written while the author was visiting the Department of Biostatistics at Harvard School of Public Health. The author would like to thank the department for their hospitality and in particular Louise Ryan and Matt Wand for discussion that led to this work. The author also acknowledges support from The Norwegian Research Council and Johan and Mimi Wessmann's foundation.

Appendix

Let $Y'_i = w_i Y_i$ where $w_i = E Y_i / \text{var}(Y_i)$ with expectations and variances evaluated at the true parameters and U_{0w} be defined by (2.4). The corresponding information matrix becomes

$$I_{0w} = -\partial U_{0w} = -\sum_{i=1}^n Y'_i \left\{ \frac{\partial^2 E Y'_i}{E Y'_i} - \left(\frac{\partial E Y'_i}{E Y'_i} \right)^2 - \frac{\sum_{j=1}^n \partial^2 E Y'_j}{\sum_{j=1}^n E Y'_j} + \left(\frac{\sum_{j=1}^n \partial E Y'_j}{\sum_{j=1}^n E Y'_j} \right)^2 \right\}.$$

We will show $\text{var}(U_{0w}) = E I_{0w}$. For notational convenience we only consider a one parameter case, the matrix generalization is straightforward.

The expected information can then be written

$$E I_{0w} = \sum_{i=1}^n \left\{ \frac{(\partial E Y'_i)^2}{E Y'_i} - \frac{(\sum_{i=1}^n \partial E Y'_i)^2}{\sum_{i=1}^n E Y'_i} \right\} = \sum_{i=1}^n E Y'_i \left\{ \frac{\partial E Y'_i}{E Y'_i} - \frac{\sum_{j=1}^n \partial E Y'_j}{\sum_{j=1}^n E Y'_j} \right\}^2.$$

Note that the formulas for information and expected information are valid for any choice of weights including $w_i = 1$. Furthermore

$$\text{var}(U_{0w}) = \sum_{i=1}^n w_i^2 \text{var}(Y_i) \left\{ \frac{\partial E Y'_i}{E Y'_i} - \frac{\sum_{j=1}^n \partial E Y'_j}{\sum_{j=1}^n E Y'_j} \right\}^2 = \sum_{i=1}^n E Y'_i \left\{ \frac{\partial E Y'_i}{E Y'_i} - \frac{\sum_{j=1}^n \partial E Y'_j}{\sum_{j=1}^n E Y'_j} \right\}^2,$$

thus $\text{var}(U_{0w}) = E I_{0w}$.

References

- [1] Andersen, PK, Borgan, Ø, Gill, RD, and Keiding, N (1993). *Statistical Models Based on Counting Processes*. Springer Verlag, New York.
- [2] Andersen, PK, and Gill, RD (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-20.
- [3] Breslow, NE (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89-99.
- [4] Cox, DR (1972). Regression models and life tables (with Discussion). *J. Roy. Statist. Soc. B* **74**, 187-200.
- [5] Cox, DR (1975). Partial likelihood. *Biometrika* **62**, 269-76.
- [6] Diggle, PJ, Liang, K-Y and Zeger, SL (1994). *Analysis of Longitudinal Data*. Clarendon Press, Oxford.
- [7] Efron, B (1977). The efficiency of Cox's likelihood function for censored data. *J. Am. Statist. Assoc.* **72**, 557-65.
- [8] Gail, MH, Lubin, JH and Rubinstein, LV (1981) Likelihood calculations for matched case-control studies. *Biometrika*, **68**, 703-707.
- [9] Howard, S (1972) Discussion of the paper by DR Cox. *J. Roy. Statist. Soc. B* **74**, 210-211.
- [10] Lin, DY, Oakes, D and Ying, Z (1998) Additive hazard models with current status data. *Biometrika* **85**, 289-298.
- [11] McCullagh, P and Nelder, JA (1989) 2nd ed. *Generalized Linear Models* . Chapman and Hall, New York.
- [12] Oakes, D (1981) Survival analysis: aspects of partial likelihood (with Discussion). *Int. Statist. Rev.* **49**, 235-64.
- [13] Peto, R (1972) Discussion of the paper by DR Cox. *J. Roy. Statist. Soc. B* **74**, 205-207.
- [14] Ryan, LM, Samuelsen, SO and Wand, MP (1999) A cost-effective conditional panel design. (in preparation).
- [15] Self, SG, Longton, G, Kopecky, KJ and Liang, K-Y (1991) On estimating HLA / disease association with application to a study of aplastic anemia. *Biometrics* **47**, 53-61.
- [16] Sundet, JM, Magnus, P, Kvale IL, Samuelsen, SO and Bakketeig, LS (1992) Secular trends and socioeconomic regularities of coital debut age in Norway. *Archives of Sexual Behavior* **22**, 241-252.