Lifetime Data Analysis, 7, 39–54, 2001

Marginal Likelihood Estimation for Proportional Odds Models with Right Censored Data

K. F. LAM hrntlkf@hkucc.hku.hk
Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong.

T. L. LEUNG
Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong.

Received February 10, 1999; Revised February 18, 2000; Accepted April 18, 2000

Abstract. One major aspect in medical research is to relate the survival times of patients with the relevant covariates or explanatory variables. The proportional hazards model has been used extensively in the past decades with the assumption that the covariate effects act multiplicatively on the hazard function, independent of time. If the patients become more homogeneous over time, say the treatment effects decrease with time or fade out eventually, then a proportional odds model may be more appropriate. In the proportional odds model, the odds ratio between patients can be expressed as a function of their corresponding covariate vectors, in which, the hazard ratio between individuals converges to unity in the long run. In this paper, we consider the estimation of the regression parameter for a semiparametric proportional odds model at which the baseline odds function is an arbitrary, non-decreasing function but is left unspecified. Instead of using the exact survival times, only the rank order information among patients is used. A Monte Carlo method is used to approximate the marginal likelihood function of the rank invariant transformation of the survival times which preserves the information about the regression parameter. The method can be applied to other transformation models with censored data such as the proportional hazards model, the generalized probit model or others. The proposed method is applied to the Veteran's Administration lung cancer trial data.

Keywords: censoring, marginal likelihood, Monte Carlo method, proportional odds model, rank invariant transformation

1. Introduction

In cancer research, the investigators might wish to ascertain the relationship between the survival times and various clinical characteristics that the patients have at diagnosis, such as age, size and number of tumors, and the treatment received. This can help to determine the prognosis of the disease under study, identify the pertinent risk factors, and control the relevant confounders by the appropriate treatment. The term survival time does not necessarily mean the time to death from the start of a diagnosis. It may represent the response time to a particular medical treatment, the time to recurrence of cancer tumor or the time to occurrence of any preassigned event.

The most commonly used model in the context of survival analysis is the proportional hazards model proposed by Cox (1972). In this model, the hazard rates for separate groups of subjects are in proportion, independent of time. Covariate effects are assumed to act multiplicatively on the baseline hazard rate in an exponential scale. Parametric analysis can be carried out if a known function is assigned to the baseline hazard rate while without any specification of the baseline hazard function, a semiparametric analysis can be performed

by the partial likelihood approach. The large sample properties of the partial likelihood estimator was proved by Cox in 1975.

The assumption of a fixed ratio between the hazard rates of separate groups of patients is sometimes inappropriate when homogeneity between groups increases with time. In such cases, a proportional odds model may be more appropriate. For the class of proportional odds models, the ratio of the hazard functions for different groups of patients converges to unity over time. This is a reasonable assumption when the heterogeneity property between groups disappears over time. When testing the effectiveness of a treatment, it is sometimes more appropriate to assume that patients in the treatment group and the control group have a similar mortality rate in the long run as the effect of the treatment may fade out gradually.

The proportional odds model has been widely used to model random variables measured with nominal or ordinal scales, especially in social science research. The odds ratio is also a commonly used measure in the biomedical field, and provides a good approximation to the relative risk in the case of rare disease. However, it is not a popular choice in the analysis of continuous survival data, possibly due to the fact that no simple, globally accepted estimation procedure is yet available.

The class of proportional odds models is a flexible class of distributions when the baseline odds function is left unspecified, and the resulting model is a semiparametric one. Bennett (1983a) has suggested the estimation of the regression parameter of the proportional odds model parametrically by assuming the time to event has a log-logistic distribution. Bennett (1983b), assuming that the baseline odds function as strictly continuous, proposed a profile likelihood for the regression parameter with the baseline odds function being profiled out. Following the work of Bennett (1983b), Murphy et al. (1997) showed that the maximum profile likelihood estimator of the regression parameter is consistent, asymptotically efficient and normally distributed. Shen (1998) proposed to approximate the baseline odds function by monotone splines and use the sieve maximum likelihood method to estimate the baseline odds function and the regression parameter simultaneously. However, in the strictly continuous case, the large number of parameters increases the difficulty of the estimation and hence, treating the baseline odds function as nuisance and eliminate it from the analysis is more preferable to simultaneous estimation. McCullagh (1984) considered a sequential conditioning approach to eliminate the nuisance parameters. Approximations to the conditional moments are necessary due to computational difficulties, but the behavior of the approximations is unknown. Cheng et al. (1995) derived the estimator of the regression parameter for a class of semiparametric transformation models using the generalized estimation equation approach with the proportional odds model as a special case.

In this paper, a marginal likelihood approach to semiparametric analysis of the proportional odds model is suggested by using the ranks only. This marginal approach has been studied by Pettitt (1983, 1984, 1987), Doksum (1987), Dabrowska and Doksum (1988) and Lam and Kuk (1997). Pettitt's method (1983) is based on a Taylor series expansion on the logarithm of the marginal likelihood about zero, and the resulting estimator is biased and inconsistent. Pettitt (1987) compared several estimators using rank information with a very good review on regression using ranks. Doksum (1987) approximated the marginal likelihood of the rank by the Monte Carlo method in the absence of censoring and the Monte

Carlo method was extended to accommodate censored data by Dabrowska and Doksum (1988). As will be shown later that the bias of the extended Monte Carlo method is rather severe and cannot be removed or reduced by increasing the size of Monte Carlo simulation. The analysis based on incomplete rankings caused by random censoring has been studied by Pettitt (1983) and Lam and Kuk (1997). We shall modify the Monte Carlo method of Doksum to accommodate cases with incomplete rankings. The bias of the proposed estimator is negligible. The proposed method can be applied to the transformation models and all regression problems using complete or incomplete rankings, with the illustrative example being a semiparametric proportional odds model.

The idea of the proportional odds model will be elaborated in the next section. The construction of the marginal likelihood function based on the rank information will be discussed in Section 3. Since it is not possible to evaluate the marginal likelihood function analytically, a Monte Carlo approximation using techniques in importance sampling is proposed in Section 4. Simulation studies and a real case application will be examined in Section 5 and Section 6, and some concluding remarks are given in Section 7.

2. Proportional Odds Model

Let t_i $(i = 1, \ldots, n)$ denote the failure time of the ith individual and $\boldsymbol{z}_i$ be the corresponding vector of covariates. Also denote the probability of failure by time t for individual i by $F_i(t|\boldsymbol{z}_i)$ or, for simplicity, by $F_i(t)$. The $F(t)$'s for different individuals may not be identical as they may be related to the vector of explanatory variables $\boldsymbol{z}$.

For the proportional odds model, the logarithm of the odds ratio of individuals i and j is given by

$$\log\left\{\frac{F_i(t)}{1-F_i(t)}\right\} - \log\left\{\frac{F_j(t)}{1-F_j(t)}\right\} = (\boldsymbol{z}_i - \boldsymbol{z}_j)^T\boldsymbol{\beta},$$

where $\boldsymbol{\beta}$ is a vector of unknown regression parameters. This is the main feature of this class of models that the difference of the log odds function is a linear function of the known covariate vectors, which is time independent. For individual i, we can write the logarithm of the odds as

$$\log\left\{\frac{F_i(t)}{1-F_i(t)}\right\} = h(t) + \boldsymbol{z}_i{}^T\boldsymbol{\beta}, \tag{1}$$

where $h(t) = \log\left\{\frac{F_0(t)}{1-F_0(t)}\right\}$ is the baseline log odds function and $F_0(t)$ is the probability of failure by time t for an individual with $\boldsymbol{z} = \boldsymbol{0}$. It is obvious that $h(t)$ is a non-decreasing function of t. When $h(t)$ is left unspecified, the resulting model is a rich class of semiparametric models and is extremely flexible in modeling a wide range of data with the characteristic as specified in (1). The distribution function $F_i(t)$ for individual i is given by

$$F_i(t) = \frac{\exp\{h(t) + \boldsymbol{z}_i{}^T\boldsymbol{\beta}\}}{1 + \exp\{h(t) + \boldsymbol{z}_i{}^T\boldsymbol{\beta}\}}.$$

The hazard function of T_i is thus

$$\lambda_i(t) = \frac{\exp\{h(t) + \boldsymbol{z_i}^T\boldsymbol{\beta}\}h'(t)}{1 + \exp\{h(t) + \boldsymbol{z_i}^T\boldsymbol{\beta}\}},$$

where $h'(t) = \frac{dh(t)}{dt}$. The ratio of the hazard functions of individuals i and j is then

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\exp\{h(t)\} + \exp\left(-\boldsymbol{z_j}^T\boldsymbol{\beta}\right)}{\exp\{h(t)\} + \exp\left(-\boldsymbol{z_i}^T\boldsymbol{\beta}\right)}.$$

Since $h(t)$ is non-decreasing and the covariate vectors are time independent, the ratio of two hazard functions converges to unity as time increases.

The log-logistic distribution belongs to the class of proportional odds models with

$$h(t) = \varphi \log(t)$$

for φ is the measure of precision.

3. Marginal Likelihood Function

Suppose the failure times t_i $(i = 1, \ldots, n)$ satisfy the semiparametric model of (1) with regression parameter $\boldsymbol{\beta}$ and unknown baseline log odds function $h(t)$. In the presence of censoring, the observed data are $(y_i, \boldsymbol{z_i}, \delta_i)$ $(i = 1, \ldots, n)$ where $\delta_i = 1$ means that $t_i = y_i$ is uncensored and $\delta_i = 0$ means that $t_i > y_i$ is right censored. Denote $t_{(1)} < \cdots < t_{(k)}$ the ordered uncensored failure times in the sample and let $t_{(0)} = 0$ and $t_{(k+1)} = \infty$. Further let R be the observed incomplete rank vector with elements r_i $(1 \le r_i \le k, i = 1, \ldots, n)$, where r_i is the rank of $t_i = y_i$ among the k uncensored observations, or y_i has rank r_i if y_i is censored in the interval $[t_{(r_i)}, t_{(r_i+1)})$. Note that if t_i is censored, then $t_{(r_i)}$ represents the largest preceding uncensored time unless $r_i = 0$ at which we define $t_{(0)} = 0$. Finally, define R^* be the set of all possible complete rankings of the n failure times consistent with R and R^{**} be the true unobserved complete ranking as if there is no censoring.

In most clinical research, the main objective is to investigate whether a certain treatment is effective in suppressing the disease under study. In such cases, instead of estimating $h(t)$ simultaneously, we prefer to treat it as nuisance and eliminate it from the likelihood function by the marginal likelihood approach as in Kalbfleisch and Prentice (1973) and Pettitt (1983, 1984). As remarked by Lam and Kuk (1997), the marginal likelihood of rank is a true likelihood under the progressive type II censoring mechanism, and can be viewed as an approximate likelihood under more general censoring scheme.

Define $W_i = \exp\left(\bar{\boldsymbol{z}}^T\boldsymbol{\beta}\right)F_0(T_i)/\{1 - F_0(T_i)\}$, where $\bar{\boldsymbol{z}} = \sum_{i=1}^n \boldsymbol{z_i}/n$ is the average of the vectors $\boldsymbol{z_i}$. Let $\boldsymbol{x_i} = \boldsymbol{z_i} - \bar{\boldsymbol{z}}$ be the standardized covariate vector for individual i so that $\sum_{i=1}^n \boldsymbol{x_i} = \boldsymbol{0}$. The survivor and the density functions of W_i are given by

$$G(w_i; \boldsymbol{x_i}) = \frac{1}{1 + \exp\left(\boldsymbol{x_i}^T\boldsymbol{\beta}\right)w_i} \tag{2}$$

and

$$g(w_i; \boldsymbol{x}_i) = \frac{\exp(\boldsymbol{x}_i{}^T \boldsymbol{\beta})}{\{1 + \exp(\boldsymbol{x}_i{}^T \boldsymbol{\beta}) w_i\}^2}. \tag{3}$$

It is obvious that W is a rank invariant transformation of the survival times T and the maximal rank-invariant information about the parameter $\boldsymbol{\beta}$ contained in the rankings of T is preserved. The marginal likelihood of rank for $\boldsymbol{\beta}$ can be written as

$$\begin{aligned} L_R(\boldsymbol{\beta}) &= P(R) \\ &= \sum_{r^* \in R^*} P(R^{**} = r^*) \\ &= \int_{0<w_{(1)}<\cdots<w_{(k)}<\infty} \cdots \int \prod_{i=1}^{n} \left\{g(w_{(r_i)}; \boldsymbol{x}_i)\right\}^{\delta_i} \left\{G(w_{(r_i)}; \boldsymbol{x}_i)\right\}^{1-\delta_i} dw_{(k)} \cdots dw_{(1)}. \end{aligned} \tag{4}$$

It can be seen that, for those $y_i < t_{(1)}$, their likelihood contributions to the marginal likelihood function $G(0) = 1$ is non-informative. For convenience, these observations would be eliminated from the dataset.

Typically, there is no explicit solution to (4). A Monte Carlo method would be suggested to approximate the marginal likelihood function by simulations.

4. Monte Carlo Approximation

As the marginal likelihood in (4) cannot be solved explicitly, the technique of importance sampling is employed to express the marginal likelihood as an expectation with respect to some distribution. Specifically, we can multiply and divide the integrand in (4) by

$$c \prod_{i=1}^{n} \left\{g_0(w_{(r_i)})\right\}^{\delta_i} \left\{G_0(w_{(r_i)})\right\}^{1-\delta_i}, \tag{5}$$

where $g_0(w)$ and $G_0(w) = 1 - \bar{G}_0(w)$ are the density and survivor functions of the standard log-logistic distribution given by

$$g_0(w) = \frac{1}{\{1 + w\}^2}$$

and

$$G_0(w) = \frac{1}{1 + w}, \tag{6}$$

respectively. The constant c is just the total number of possible rankings among the n survival times which are consistent with the observed incomplete ranking R. Let d_l denote the number of censored observations in the interval $[t_{(l)}, t_{(l+1)})$ then the number of individuals

at risk just prior to $t_{(l)}$ is given by

$$c_l = (n - l + 1) - \sum_{i=1}^{l-1} d_i$$

for $2 \le l \le k$ with $d_0 = 0$ and $c_1 = n$. The total number of possible rankings is given by

$$c = \prod_{l=1}^{k} \left\{ \prod_{j=1}^{d_l} (c_l - j) \right\}^{I(d_l > 0)},$$

where $I(d_l > 0)$ is the indicator function. It is easily seen that (5) is just the density function of the uncensored order statistics $W_{(1)} < \cdots < W_{(k)}$ under a progressive type II censored sample with fixed $d_l (l = 1, \ldots, k)$ and the sample size being n.

The marginal likelihood function (4) can then be expressed as

$$\text{Ł}_R(\beta) = \frac{1}{c} E_{\bar{G}_0} \left(E_{R^{**}} \left[\frac{\prod_{i=1}^{n} \left\{ g(w_{(r_i)}; \boldsymbol{x_i}) \right\}^{\delta_i} \left\{ G(w_{(r_i)}; \boldsymbol{x_i}) \right\}^{1-\delta_i}}{\prod_{i=1}^{n} \left\{ g_0(w_{(r_i)}) \right\}^{\delta_i} \left\{ G_0(w_{(r_i)}) \right\}^{1-\delta_i}} \middle| R \right] \right), \tag{7}$$

where the inner expectation is taken with respect to the unobserved complete ranking R^{**}, conditioned on the observed incomplete ranking R and $w_1, \ldots, w_n$, while the outer expectation is taken with respect to the uncensored sample of order statistics $W_{(1)} < \cdots < W_{(k)}$ under a progressive type II censoring scheme and $W_1, \ldots, W_n$ are independently and identically distributed as $\bar{G}_0$.

As the W's are independently and identically distributed in the expectations, it can be seen easily that for any particular complete ranking $r^* \in R^*$,

$$P(R^{**} = r^* | R) = 1/c$$

which is just a constant. Therefore, a 2-step Monte Carlo resampling procedure is proposed to approximate the expectation in (7) and is described below:

1. Simulate M_0 sets of $(W_1, \ldots, W_n)$ from $\bar{G}_0$, namely $(w_{1,j}, \ldots, w_{n,j})$ for $j = 1, \ldots, M_0$ and order them in ascending order to give $w_{(1),j} < \cdots < w_{(n),j}$.

2. List out all $r^* \in R^*$, namely $r_1^*, \ldots, r_c^*$, and map the rankings of r_ℓ^*, $(\ell = 1, \ldots, c)$ to $(w_{(1),j} < \cdots < w_{(n),j})$ to obtain a realization of the uncensored sample of order statistics $w_{(1),j,r_\ell^*} < \cdots < w_{(k),j,r_\ell^*}$ under the progressive type II censoring scheme.

Therefore, the marginal likelihood function can be approximated by

$$\begin{aligned} \widehat{L_R}(\beta) &= \frac{1}{c} \frac{1}{M_0} \left[\sum_{j=1}^{M_0} \sum_{\ell=1}^{c} P(R^{**} = r_\ell^* | R) \frac{\prod_{i=1}^{n} \left\{ g(w_{(r_i),j,r_\ell^*}; \boldsymbol{x_i}) \right\}^{\delta_i} \left\{ G(w_{(r_i),j,r_\ell^*}; \boldsymbol{x_i}) \right\}^{1-\delta_i}}{\prod_{i=1}^{n} \left\{ g_0(w_{(r_i),j,r_\ell^*}) \right\}^{\delta_i} \left\{ G_0(w_{(r_i),j,r_\ell^*}) \right\}^{1-\delta_i}} \right] \\ &= \frac{1}{c} \frac{1}{M_0} \left[\sum_{j=1}^{M_0} \sum_{\ell=1}^{c} \frac{\prod_{i=1}^{n} \left\{ g(w_{(r_i),j,r_\ell^*}; \boldsymbol{x_i}) \right\}^{\delta_i} \left\{ G(w_{(r_i),j,r_\ell^*}; \boldsymbol{x_i}) \right\}^{1-\delta_i}}{c \prod_{i=1}^{n} \left\{ g_0(w_{(r_i),j,r_\ell^*}) \right\}^{\delta_i} \left\{ G_0(w_{(r_i),j,r_\ell^*}) \right\}^{1-\delta_i}} \right]. \end{aligned} \tag{8}$$

As the value of c increases dramatically with increasing number of censored observations, instead of listing out all the possible rankings in the set R^*, only B such rankings will be generated at random and the marginal likelihood can further be approximated by

$$\widehat{L_R}(\beta) = \frac{1}{c}\frac{1}{M_0}\sum_{j=1}^{M_0}\left[\sum_{b=1}^{B}\frac{1}{B}\frac{\prod_{i=1}^{n}\left\{g(w_{(r_i),j,b};\boldsymbol{x_i})\right\}^{\delta_i}\left\{G(w_{(r_i),j,b};\boldsymbol{x_i})\right\}^{1-\delta_i}}{\prod_{i=1}^{n}\left\{g_0(w_{(r_i),j,b})\right\}^{\delta_i}\left\{G_0(w_{(r_i),j,b})\right\}^{1-\delta_i}}\right]. \tag{9}$$

Note that the same set of simulated $(w_{1,j},\ldots,w_{n,j})$ is used in each iteration. This is slightly different from the usual Monte Carlo estimation methods. The logarithm of the marginal likelihood function $\log\widehat{L_R}(\boldsymbol{\beta})$ is then maximized with respect to $\boldsymbol{\beta}$ to obtain the maximum likelihood estimates $\hat{\boldsymbol{\beta}}$. The maximization can easily be done by numerical methods such as the Newton Raphson method. As noted by Lam and Kuk (1997) that using the same set of $w_{1,j},\ldots,w_{n,j}$ will save a lot of computing effort, and it also implies that the first and second derivatives of $\log\widehat{L_R}(\boldsymbol{\beta})$ can be obtained analytically to provide an estimate for the information matrix.

The Monte Carlo method of Dabrowska and Doksum (1988) generates random samples of size k from $\bar{G}_0$ rather than the uncensored samples of order statistics $w_{(1)}<\cdots<w_{(k)}$ under a progressive type II censoring scheme. In those experiments with moderate to heavy censoring, the bias is quite severe and does not seem to be removable by increasing the simulation size. The proposed algorithm provides an efficient way to obtain marginal estimates in the semiparametric model.

To estimate the variance covariance matrix of $\hat{\boldsymbol{\beta}}$, the usual asymptotic variance covariance matrix, which is transformed from the Fisher information matrix, would be used. As our estimator depends on the simulated data, the variance induced by the Monte Carlo simulation can be added according to

$$Var(\hat{\boldsymbol{\beta}}) = Var\{E(\hat{\boldsymbol{\beta}}|D)\} + E\{Var(\hat{\boldsymbol{\beta}}|D)\}, \tag{10}$$

where D is the set of simulated data. The first term in (10) can be estimated by the inverse of the Fisher information matrix given the simulated data. The second term in (10) represents the simulation induced variance. To estimate the simulation induced variance, the estimation procedure can be repeated m times to obtain m estimates of $\boldsymbol{\beta}$, namely $\hat{\boldsymbol{\beta}}_1,\ldots,\hat{\boldsymbol{\beta}}_m$. Sample variance covariance matrix $\boldsymbol{V}$ can be computed based on the m estimates to approximate the simulation induced variance. The mean of the m estimates $\bar{\hat{\boldsymbol{\beta}}}$ is used as a more precise estimator for $\boldsymbol{\beta}$. The variance of the estimator can thus be estimated by the asymptotic covariance matrix evaluated at $\bar{\hat{\boldsymbol{\beta}}}$ plus the sample variance of the mean conditioned on the simulated data to give

$$Var(\bar{\hat{\boldsymbol{\beta}}}) = -\boldsymbol{l}''^{-1}\big|_{\beta=\bar{\hat{\beta}}} + \frac{\boldsymbol{V}}{m}, \tag{11}$$

where $\boldsymbol{l}''$ is the Fisher information.

Table 1. Estimation results with $\beta = \log 1/2$ and $m = 1$.

resampling size M_0	sample size n	$Z \sim Bern(0.6)$ bias	$\hat{\sigma}$	S	$Z \sim Bern(0.8)$ bias	$\hat{\sigma}$	S
100	100	−0.015	0.364	0.356	−0.054	0.448	0.469
	300	−0.017	0.208	0.220	0.009	0.254	0.271
	500	0.006	0.161	0.164	0.022	0.197	0.185
500	100	0.042	0.364	0.367	−0.037	0.447	0.487
	300	−0.019	0.208	0.213	−0.003	0.254	0.264
	500	−0.005	0.161	0.163	0.014	0.197	0.205
1000	100	−0.037	0.365	0.397	0.017	0.445	0.449
	300	−0.001	0.208	0.215	0.025	0.252	0.227
	500	0.014	0.161	0.177	0.020	0.196	0.205

5. Simulation Study

To illustrate the performance of the proposed estimation method in the class of proportional odds models, three sets of simulation were carried out. In the first set, a medical experiment is simulated in the absence of censoring. Suppose a new treatment is introduced in a clinical trial to combat a certain disease. Patients are allocated to the treatment and the control groups at random but in different proportion. A one dimensional covariate vector is assigned to each subject, indicating the type of treatment they are receiving. In our simulation, the variable Z is simulated from a Bernoulli(p) distribution with $p = 0.6$ and $p = 0.8$. Subjects with $z = 1$ are allocated to the treatment group while those with $z = 0$ are allocated to the control group. The survival time for each subject is then simulated from the log-logistic distribution, the parametric form of the proportional odds model, by the probability integral transform to give

$$t_i = \frac{u_i}{(1 - u_i)\exp(z_i\beta)}$$

where u_i is a random number generated from a uniform(0, 1) distribution.

The regression parameter β here is just the logarithm of the odds ratio between patients in the treatment group and the control group. Using the ranks of the simulated survival times and the covariate information, the proposed estimation method is used to estimate the regression parameter β with different combinations of the resampling size M_0 and the sample size n. For each combination, 200 datasets are simulated and $\hat{\beta}_1, \ldots, \hat{\beta}_{200}$ are estimated. The mean and the sample variance S^2 of the 200 estimates are computed. The theoretical variance σ^2 of the estimate is approximated by the mean of the asymptotic variances evaluated at each estimate $\hat{\beta}$, namely $\hat{\sigma}^2$. Results are summarized in Table 1. It can be seen that the estimation bias is very small compare to its standard error. Moreover, the size of M_0 does not seem to affect the performance of the estimation procedure and a small to moderate value of M_0 would be adequate in general. Without taking the simulation induced variance into account, the inverse of the Fisher information still gives a good approximation to the variance of the estimate.

To illustrate the performance of the proposed estimation method in the presence of censoring, we consider three censoring proportions, 30%, 20% and 10% under two different censoring schemes in the second set of simulation. In the first case, the survival time and the censoring time for each subject are simulated independently so that censoring is under a random process. While in the second case, we adopt a type II censoring mechanism so that the largest 10 to 30% of the observations are censored. Using the incomplete ranking, the proposed estimation method and the method of Dabrowska and Doksum (1988) are used to estimate the regression parameter β.

Two hundred data sets are generated, each with $n = 200$ observations. The setup is similar to that in the first set of simulation. For the proposed method, we let the resampling size $M_0 = 200$ and each set of simulated data will be reordered 20 times ($B = 20$) for the random censoring case. For the type II censoring case, we only need the k smallest t from each simulated sample of size n, and hence no reordering is required ($B = 1$). For the Monte Carlo method of Dabrowska and Doksum (1988), we try two different resampling sizes, namely $M_0 = 200$ and $M_0 = 4000$. The summary of the results are given in Table 2.

Results showed that the proposed estimation method can accommodate censored observations well. Although the random censoring scheme used in the simulation is not exactly a progressive type II censoring mechanism, the estimation bias is still small and stable for even up to 30% censoring. Results also suggested that a small value of B, say $B = 20$, is sufficient to give a good approximation to the marginal likelihood function. The computation times required are comparable when we use $M_0 = 4000$ in Dabrowska and Doksum's method and $M_0 = 200$ and $B = 20$ in the proposed method. The bias induced by the Monte Carlo method of Dabrowska and Doksum (1988) is quite large in magnitude, which increases with the censoring proportion. Besides, the magnitude of the bias seems unable to be reduced by increasing the simulation size. The problem is much more severe in the case of type II censoring.

Comparison is made between the efficiency of the profile likelihood estimator of Bennett (1983b) and the proposed estimator in the last set of simulation. Following the work of Murphy et al. (1997), we consider different design matrices and censoring mechanisms. A 2-dimensional covariate vector, $\boldsymbol{Z} = (Z_1, Z_2)$, is used where Z_1 is a Bernoulli($p = 0.5$) variable and Z_2 follows either a uniform, exponential or gamma distribution. As in Murphy et al. (1997), the three variables are standardized to have mean and variance equal 1 and the coefficients of skewness equal 0, 2 and 4, respectively. Combining with the 3 design matrices, 3 censoring mechanisms are used to generate the incomplete rankings. In the independent censoring, failure times are censored at a fixed time so that approximately 20% of the failure times are censored, independent of the covariate pattern. For the dependent cases, censoring depends on the value of Z_1. In the dependent 1 situation, observations with $Z_1 = 1$ and $Z_1 = 0$ are censored separately at distinct fixed times so that about 10% of the failure times in each group are censored, while in the dependent 2 censoring, only observations with $Z_1 = 1$ will be censored and the censoring proportion is about 20%. Two different sample sizes, $n = 100$ and $n = 200$, are considered. Two hundred datasets are simulated for each of the combinations. For each method, the mean of the 200 estimates, the empirical mean square error, together with the average CPU time required are tabulated in Table 3. As demonstrated in Murphy et al. (1997), the performance between

Table 2. Comparison between the proposed method and the Dabrowska & Doksum's (D & D) method with $\beta = \log 1/2$, $n = 200$ and $m = 1$.

Censoring Mechanism	proposed method $M_0 = 200^*$			D & D's method $M_0 = 200$			D & D's method $M_0 = 4000$		
	bias	$\hat{\sigma}$	S	bias	$\hat{\sigma}$	S	bias	$\hat{\sigma}$	S
				$Z \sim Bern(0.6)$					
Random 30%	0.02628	0.27081	0.25755	−0.09263	0.27968	0.28404	−0.09431	0.28041	0.28548
Random 20%	0.00780	0.27188	0.28080	−0.10842	0.28084	0.32185	−0.10938	0.28135	0.32261
Random 10%	−0.00795	0.25799	0.25933	−0.07527	0.26448	0.28538	−0.07555	0.26473	0.28679
Type II 30%	−0.01996	0.25799	0.27503	−0.43414	0.30671	0.44251	−0.43398	0.30682	0.44256
Type II 20%	−0.02190	0.25632	0.25938	−0.29370	0.28534	0.36134	−0.29473	0.28611	0.36241
Type II 10%	−0.01298	0.25591	0.24224	−0.14559	0.26920	0.28690	−0.14618	0.26931	0.28737
				$Z \sim Bern(0.8)$					
Random 30%	0.00836	0.33032	0.31502	−0.09568	0.33832	0.34440	−0.09650	0.33870	0.34514
Random 20%	0.01616	0.32009	0.34446	−0.09423	0.33799	0.38832	−0.09434	0.33833	0.38897
Random 10%	−0.00348	0.31610	0.30290	−0.06284	0.32249	0.32212	−0.06309	0.32251	0.32291
Type II 30%	−0.03016	0.31359	0.30736	−0.37325	0.36089	0.44340	−0.37518	0.36152	0.44730
Type II 20%	−0.03398	0.31520	0.33380	−0.39015	0.36513	0.48856	−0.38146	0.36520	0.48264
Type II 10%	0.01054	0.31220	0.34395	−0.09788	0.32498	0.39148	−0.09833	0.32519	0.39045

* $B = 20$ for the random censoring samples

Table 3. Comparison between the proposed estimator and the profile likelihood estimator with $\beta_1 = 0$, $\beta_2 = 1$ with $n = 100$ and $n = 200$.

$n = 100$									
	proposed estimator*				profile likelihood estimator				
	$\hat{\beta}_{1A}$		$\hat{\beta}_{2A}$		$\hat{\beta}_{1B}$		$\hat{\beta}_{2B}$		
Censoring mechanism	bias	MSE	bias	MSE	bias	MSE	bias	MSE	$\frac{MSE(\hat{\beta}_{2B})}{MSE(\hat{\beta}_{2A})}$
independent, Uniform	0.0368	0.1322	−0.0408	0.0296	0.0401	0.1468	0.0110	0.0415	1.4030
CPU time#				1.07				54.92	
independent, Exponential	−0.0626	0.1496	0.0017	0.0400	−0.0621	0.1611	0.0430	0.0559	1.3975
CPU time#				1.12				54.86	
independent, Gamma	−0.0195	0.1177	−0.0378	0.0342	−0.0200	0.1269	0.0468	0.0516	1.5088
CPU time#				7.21				56.06	
dependent1, Uniform	0.0315	0.1353	−0.0338	0.0296	0.0335	0.1472	0.0111	0.0402	1.3581
CPU time#				30.05				120.94	
dependent1, Exponential	0.0120	0.1386	−0.0195	0.0293	0.0123	0.1516	0.0403	0.0491	1.6758
CPU time#				29.97				109.11	
dependent1, Gamma	0.0329	0.1210	−0.0320	0.0414	0.0346	0.1325	0.0652	0.0693	1.6739
CPU time#				29.99				119.41	
dependent2, Uniform	−0.0242	0.1423	−0.0468	0.0332	0.0177	0.1524	0.0038	0.0416	1.2530
CPU time#				22.25				684.66	
dependent2, Exponential	−0.0445	0.1151	−0.0443	0.0387	0.0167	0.1270	0.0277	0.0520	1.3437
CPU time#				24.69				615.32	
dependent2, Gamma	−0.0404	0.1115	−0.0512	0.0425	0.0126	0.1215	0.0349	0.0587	1.3812
CPU time#				23.34				676.84	

(*continued*)

Table 3. (*continued*).

$n = 200$									
	proposed estimator*				profile likelihood estimator				
	$\hat{\beta}_{1A}$		$\hat{\beta}_{2A}$		$\hat{\beta}_{1B}$		$\hat{\beta}_{2B}$		$\frac{MSE(\hat{\beta}_{2B})}{MSE(\hat{\beta}_{2A})}$
Censoring mechanism	bias	MSE	bias	MSE	bias	MSE	bias	MSE	
independent, Uniform	0.0007	0.0681	−0.0219	0.0139	−0.0154	0.0719	0.0291	0.0232	1.6691
CPU time#				3.29				111.49	
independent, Exponential	0.0088	0.0503	−0.0179	0.0195	0.0077	0.0551	0.0258	0.0288	1.4769
CPU time#				3.81				121.14	
independent, Gamma	−0.0161	0.0628	−0.0118	0.0201	−0.0095	0.0674	0.0441	0.0312	1.5522
CPU time#				24.15				165.31	
dependent1, Uniform	0.0199	0.0739	−0.0187	0.0143	−0.0095	0.0755	0.04405	0.0157	1.0979
CPU time#				71.49				210.57	
dependent1, Exponential	−0.0003	0.0543	−0.03446	0.0158	−0.0265	0.0623	0.0301	0.0257	1.6266
CPU time#				77.14				276.09	
dependent1, Gamma	−0.0043	0.0693	−0.0284	0.0196	−0.0201	0.0609	0.0295	0.0319	1.6276
CPU time#				77.82				325.35	

*$M_0 = 200$, $m = 1$ and $B = 10$ for the proposed estimate

Average CPU time in seconds on the mainframe IBM SP2 supercomputer system

the estimator of Cheng et al. (1995) and the profile likelihood estimator are similar in terms of the mean square error. Therefore, only the proposed estimator and profile likelihood estimator are compared in this simulation study. The computing method of the profile likelihood estimator is essentially identical to that described in Murphy et al. (1997) while Newton's method is used for the proposed estimator.

The estimation biases of the two estimation methods are relatively small while our estimator seems to be more efficient than the profile likelihood estimator empirically. For the estimation of β_1, the mean square error of the proposed estimator is, on the average, 3–10% smaller than that of the profile likelihood estimator. While for the estimation of β_2, the gain in relative efficiency is more substantial over the profile likelihood estimator, with an average of about 40%. The skewness of the covariates does not seem to have much effect on the gain in relative efficiency, but the censoring mechanisms do affect this gain. The simulation results for the case with dependent 2 censoring mechanism and $n = 200$ are not shown in Table 3 because convergence cannot be achieved for about 40% of the datasets from each of the two methods. It could be due to the moderate censoring proportion for $Z_1 = 1$. As the number of parameters increases with the sample size, a tight convergence criterion is a burden to the profile likelihood estimator. The gain in relative efficiency of the proposed Monte Carlo estimator could mainly be due to the elimination of the large number of nuisance parameters which stabilizes the estimation function. The average CPU times required vary among different censoring mechanisms for both methods. The convergence criteria for the both methods are set at less than 0.0001 for all parameters. This criterion is not too strong for the proposed estimator as there are only two parameters to be estimated. However this may be too tight for the profile likelihood estimator with over 100 parameters to be estimated which leads to the longer computing time required. A trade off is that the estimate of the baseline odds function is also provided.

The simulations only indicate that the proposed estimator has a smaller mean square error than the asymptotically efficient estimator of Murphy et al. (1997) when the sample size is fixed at $n = 100$ or 200. These gains are presumably an artifact of small samples. As the sample size goes to infinity, the profile likelihood estimator will be efficient irregardless of the censoring mechanism.

6. A Real Case Application

The above estimation method is applied to the Veteran's Administration lung cancer trial data (Prentice, 1973) for illustration purpose. In this trial, males with advanced inoperable lung cancer were randomized to either a standard treatment or chemotherapy. This data set has been considered by many authors, particularly those who worked on proportional odds models like Bennett (1983a,b), Cheng et al. (1995), Murphy et al. (1997) and Pettitt (1984). All the abovementioned authors only used the group of patients with no prior therapy ($n = 97$). The response variable of interest is the patient's survival time, and the only explanatory variables included in their studies are the performance status (PS) and the tumor size. The performance status, X_1, is assumed to be a continuous variable ranging from 0 to 100 while the tumor size is a factor with 4 levels, namely large, adeno, small and squamous.

Table 4. Estimates and standard errors of the regression coefficients for the Veteran's Administration lung cancer trial data.

Covariate	Proposed estimator	Murphy et al. (1997)	Cheng et al. (1995), w_Q
PS	−0.053(0.010)	−0.055(0.010)	−0.055(0.010)
Tumor type vs large			
Adeno	1.316(0.551)	1.339(0.556)	1.559(0.411)
Small	1.367(0.518)	1.440(0.525)	1.494(0.499)
Squamous	−0.247(0.590)	−0.217(0.589)	−0.004(0.569)
Covariate	Bennett (1983b)	Pettitt (1984)	Cheng et al. (1995), $w = 1$
PS	−0.053(0.010)	−0.055(0.010)	−0.055(0.010)
Tumor type vs large			
Adeno	1.314(0.554)	1.302(0.554)	1.556(0.414)
Small	1.383(0.524)	1.438(0.520)	1.496(0.498)
Squamous	−0.181(0.588)	−0.177(0.593)	−0.006(0.572)

The proposed estimation procedure is repeated 10 times ($m = 10$) and the mean of the 10 estimates is used to estimate the regression vector β. The monte carlo resampling size is fixed at $M_0 = 50,000$ and in each resampling sample, the data are re-ordered 50 times ($B = 50$). The variances of the parameter estimates are estimated by the information matrix evaluated at $\bar{\hat{\beta}}$ plus the simulation induced variance as in (11). In the analysis, the simulation induced variance is very small, contributing only about 1–3% of the total variation.

The estimates and the corresponding standard errors obtained by Murphy et al. (1997), Cheng et al. (1995) with weight functions w_Q which mimics the quasi-likelihood approach and $w = 1$, Bennett (1983b) and Pettitt (1984) are reproduced and tabulated in Table 4 together with the estimates of the proposed estimator. The conclusions are the same for all six methods that 'squamous vs large' is the only insignificant covariate in the analysis. Our estimates closely agree with those obtained by the profile likelihood type estimators of Bennett (1983b) and Murphy et al. (1997), while the estimates by Cheng et al. (1995) seem to be consistently greater than other estimates by a small amount.

The estimate of the regression coefficient of PS is -0.0529 which indicates that for those patients with the same tumor size, a one point increment in the Performance Status will result in a reduction of $\exp(-0.0529) = 0.9485$ in the odds of dying and this estimate can be interpreted easily.

7. Conclusion

The class of semiparametric proportional odds models has received more attention recently due to its high flexibility and the property that the hazard ratio between the treated and control subjects converges to unity over time, which is more realistic in the cases of chronic diseases. A method is suggested for the estimation of the regression parameter for this class of models, as well as the generalized linear model, in modeling survival data using ranks in the presence of censoring.

Using the marginal likelihood approach, the marginal likelihood function of a rank invariant transformation of the failure times is constructed so that only the rank information is required in the estimation procedure. By the technique in importance sampling, the marginal likelihood function can be approximated and the problem of censoring can easily be tackled by re-ordering the simulated data to approximate the expectation due to the incomplete ranking. Results from the simulations showed that the proposed estimator performs rather good in estimating the regression parameter even for the cases with moderate censoring.

This method is a natural modification of the Monte Carlo method by Dabrowska and Doksum (1988) to adjust for the bias induced. It can be applied to any class of semiparametric transformation models, under which, a known, rank invariant transformation of the survival time is of particular interest. This includes the proportional hazards model, proportional odds rate model and the generalized probit model.

The proposed estimator is quite efficient and is comparable with the asymptotically efficient profile likelihood estimator. The elimination of the large number of nuisance parameters leads to the flow of information to the estimation of the regression parameter. The gain in relative efficiency over the profile likelihood estimator can be significant when the sample size is small. For the cases with large sample size, the computing time required would be less than the profile likelihood estimator as the number of parameters to be estimated would not be affected by the sample size.

Extension of the proportional odds model to model multivariate survival data can be made possible by introducing a random effect to model the correlation between observations of the same cluster. A common random effect can be used to explain the shared biological characteristics between twins, a multivariate random effects can be used to model the correlation between multiple events, while a dynamic random effects model can be used to model recurrent events. A manuscript generalizing the above method to model different types of multivariate survival data is in progress.

Acknowledgments

The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKU 7140/98H) and by a CRCG research grant 337/017/0017 of the University of Hong Kong, Hong Kong. The authors would like to thank the Editor and two referees for their helpful comments and suggestions.

References

S. Bennett, "Log-logistic regression models for survival data," *Applied Statistics* vol. 32 pp. 165–171, 1983a.

S. Bennett, "Analysis of survival data by the proportional odds model," *Statistics in medicine* vol. 2 pp. 273–277, 1983b.

S.C. Cheng, L.J. Wei and Z. Ying, "Analysis of transformation models with censored data," *Biometrika* vol. 82 pp. 835–845, 1995.

D.R. Cox, "Regression models and life-tables (with discussion)," *Journal of the Royal Statistical Society, Series B* vol. 34 pp. 187–220, 1972.
D.R. Cox, "Partial likelihood," *Biometrika* vol. 62 pp. 269–276, 1975.
D.M. Dabrowska and K.A. Doksum, "Partial likelihood in transformation models with censored data," *Scandinavian Journal of Statistics* vol. 15 pp. 1–23, 1988.
K.A. Doksum, "An extension of partial likelihood methods for proportional-hazard models to general transformation models," *Annals of Statistics* vol. 15 pp. 325–345, 1987.
J.D. Kalbfleisch and R.L. Prentice, "Marginal likelihoods based on Cox's regression and life table," *Biometrika* vol. 60 pp. 267–278, 1973.
K.F. Lam and Y.C. Kuk, "A marginal likelihood approach to estimation in frailty models," *Journal of the American Statistical Association* vol. 92 pp. 985–990, 1997.
P. McCullagh, "On the elimination of nuisance parameters in the proportional odds model," *Journal of the Royal Statistical Society, Series B* vol. 46 pp. 250–256, 1984.
S.A. Murphy, A.J. Rossini and A.W. Van Der Vaart, "Maximum likelihood estimation in the proportional odds model," *Journal of the American Statistical Association* vol. 92 pp. 968–976, 1997.
A.N. Pettitt, "Approximate methods using ranks for regression with censored data," *Biometrika* vol. 70. pp. 121–132, 1983.
A.N. Pettitt, "Proportional odds model for survival data and estimates using ranks," *Applied Statistics* vol. 33 pp. 169–175, 1984.
A.N. Pettitt, "Estimates for a regression parameter using ranks," *Journal of the Royal Statistical Society, Series B* vol. 49 pp. 58–67, 1987.
R.L. Prentice, "Exponential survivals with censoring and explanatory variables," *Biometrika* vol. 60 pp. 279–288, 1973.
R.L. Prentice, "Linear rank tests with right censored data," *Biometrika* vol. 65 pp. 167–179, 1978.
X. Shen, "Proportional odds regression and sieve maximum likelihood estimation," *Biometrika* vol. 85 pp. 165–177, 1998.