Nuisance parameter elimination for proportional likelihood ratio models with nonignorable missingness and random truncation

BY KWUN CHUEN GARY CHAN

Department of Biostatistics, University of Washington, Seattle, Washington 98195, U.S.A.

kcgchan@u.washington.edu

SUMMARY

We show that the proportional likelihood ratio model proposed recently by Luo & Tsai (2012) enjoys model-invariant properties under certain forms of nonignorable missing mechanisms and randomly double-truncated data, so that target parameters in the population can be estimated consistently from those biased samples. We also construct an alternative estimator for the target parameters by maximizing a pseudo-likelihood that eliminates a functional nuisance parameter in the model. The corresponding estimating equation has a U-statistic structure. As an added advantage of the proposed method, a simple score-type test is developed to test a null hypothesis on the regression coefficients. Simulations show that the proposed estimator has a small-sample efficiency similar to that of the nonparametric likelihood estimator and performs well for certain nonignorable missing data problems.

Some key words: Double truncation; Nonignorable missingness; Pairwise pseudolikelihood; U-statistic.

1. INTRODUCTION

Regression modelling is popular in data analysis. However, challenges arise when the response has a finite range with a substantial portion of data taking boundary values, when its distribution is multimodal and asymmetric, or when the relationship between response and covariates is nonlinear. To handle regression modelling with an unknown response distribution and other such practical concerns, Luo & Tsai (2012) proposed a semiparametric proportional likelihood ratio model, which is a parsimonious alternative to existing nonparametric models. The semiparametric model relates a response $Y$ and a $p \times 1$ covariate vector $X$ in the following way:

$$f(y \mid x) = \frac{\exp(\beta^\top y x) g(y)}{\int \exp(\beta^\top y x) dG(y)},$$

(1)

where $\beta$ is a Euclidean parameter of interest and $G(y)$ is an unspecified baseline distribution function with density function $g(y)$ with respect to some dominating measure. For arbitrary $y_1$ and $y_2$ within the support of $G$, we have

$$\log \frac{f(y_2 \mid x)}{f(y_1 \mid x)} = \log \frac{g(y_2)}{g(y_1)} + \beta^\top x(y_2 - y_1).$$

(2)

If $X$ is one-dimensional and the response $Y$ is binary, then

$$\beta = \log \frac{f(1 \mid x + 1)/f(0 \mid x + 1)}{f(1 \mid x)/f(0 \mid x)},$$

which is exactly the log odds ratio. Therefore, $\beta$ is a generalization of the log odds ratio for an arbitrary response. For vector-valued $X$, it follows from (2) that the $j$th component of $\beta$ is the loglikelihood ratio that the response increases by one unit given that the $j$th component of $x$ increases by one unit, while all...
the other components of \( x \) are held fixed. There is no intercept term in this model, because an intercept would be contained in \( G \).

Luo & Tsai (2012) discussed extensively the connection of model (1) with generalized linear models (Nelder & Wedderburn, 1972), density ratio models (Qin & Zhang, 1997), biased sampling models (Gilbert et al., 1999), single-index models (Ichimura, 1993) and exponential tilt regression models (Rathouz & Gao, 2009). Huang & Rathouz (2012) explicitly modelled the mean function for a proportional likelihood ratio model with discrete covariates. We will show that model (1) is a robust extension of the generalized linear model for certain missing data problems and doubly truncated data. For model (1), Luo & Tsai (2012) proposed a nonparametric maximum likelihood method that estimates \( \beta \) and \( G \) simultaneously by an iterative algorithm, as usual likelihood arguments do not lead to elimination of the nuisance function \( G(y) \). We present an alternative method for estimating \( \beta \), which eliminates the unknown function \( G \) based on conditioning on the rank of responses (Kalbfleisch, 1978) in a pairwise fashion (Liang & Qin, 2000).

2. Invariance Properties Under Missingness and Truncation

Missing data are common in practice for reasons that are often beyond the control of investigators. Suppose that we fail to observe response \( Y \) for some observations and covariates \( X \) for some possibly different observations. Let \( R_Y \) and \( R_X \) be the indicators that \( Y \) and \( X \) are being observed, and let \( R = R_Y R_X \) be the indicator of both being observed. We refer to an observation with \( R = 1 \) as a complete case; for such an observation, the conditional density is

\[
f(y_i | x_i, r_i = 1) = \frac{\text{pr}(r_i = 1 | y_i, x_i) f(y_i | x_i)}{\int \text{pr}(r_i = 1 | y, x_i) f(y | x_i) \, dy}.
\]

(3)

As seen from (3), inference for \( \beta \) from complete case observations often requires correct specification of the conditional probability of \( R \) given \( (Y, X) \). However, when \( \text{pr}(R = 1 | Y, X) \) depends on the unobserved data, the missing mechanism is nonignorable (Rubin, 1976). When \( \text{pr}(R | Y, X) \) is arbitrary, inference based on (3) requires untestable assumptions on the missing data mechanism.

We assume the following factorization of the missing data model:

\[
\text{pr}(R = 1 | Y = y, X = x) = h_1(x)h_2(y)
\]

(4)

for arbitrary functions \( h_1(\cdot) \) and \( h_2(\cdot) \). This assumption will be satisfied for the following examples, which include cases where missingness is nonignorable.

**Example 1.** Suppose that response \( Y \) is subject to missingness but covariate \( X \) is completely observed; that is, \( R_X \equiv 1 \).

(a) Suppose that the missingness depends only on unobserved \( Y \), \( \text{pr}(R_Y = 1 | Y, X) = \pi_{1a}(Y) \); then (4) is satisfied with \( h_1(x) = 1 \) and \( h_2(y) = \pi_{1a}(y) \).

(b) Suppose that the missingness depends only on observed \( X \), \( \text{pr}(R_Y = 1 | Y, X) = \pi_{1b}(X) \); then (4) is satisfied with \( h_1(x) = \pi_{1b}(x) \) and \( h_2(y) = 1 \).

**Example 2.** Suppose that response \( Y \) is completely observed but covariate \( X \) is subject to missingness; that is, \( R_Y \equiv 1 \).

(a) Suppose that the missingness depends only on observed \( Y \), \( \text{pr}(R_X = 1 | Y, X) = \pi_{2a}(Y) \); then (4) is satisfied with \( h_1(x) = 1 \) and \( h_2(y) = \pi_{2a}(y) \).

(b) Suppose that the missingness depends only on unobserved \( X \), \( \text{pr}(R_X = 1 | Y, X) = \pi_{2b}(X) \); then (4) is satisfied with \( h_1(x) = \pi_{2b}(x) \) and \( h_2(y) = 1 \).

**Example 3.** Here both the response \( Y \) and the covariate \( X \) may be missing.

(a) Suppose that the missing data mechanism for both \( Y \) and \( X \) depends only on response \( Y \), i.e. \( \text{pr}(R_X = 1, R_Y = 1 | Y, X) = \pi_{3a}(Y) \); then (4) is satisfied with \( h_1(x) = 1 \) and \( h_2(y) = \pi_{3a}(y) \).
(b) Suppose that the missing data mechanism for both Y and X depends only on covariate X, i.e., 
\[ \Pr(R_X = 1, R_Y = 1 \mid Y, X) = \pi_{3X}(X); \] then (4) is satisfied with \( h_1(x) = \pi_{3X}(X) \) and \( h_2(y) = 1 \).

(c) Suppose that the missing data mechanism for Y depends only on Y and the missing data mecha-
nism for X depends only on X, i.e., \( \Pr(R_Y = 1 \mid Y, X) = \pi_{3y}(Y) \) and \( \Pr(R_X = 1 \mid Y, X) = \pi_{3x}(X) \).

Suppose further that \( R_X \) and \( R_Y \) are conditionally independent given \( (Y, X) \). Then (4) is satisfied 
with \( h_1(x) = \pi_{3x}(x) \) and \( h_2(y) = \pi_{3y}(y) \).

Except for Examples 1(b) and 2(a), all the cases have nonignorable missing data. We now state an 
invariance property for the proportional likelihood ratio model.

**Property 1.** If (4) is satisfied, then 
\[ f(y \mid x, r = 1) = \frac{\exp(\beta^T xy) \, dG_1(y)}{\int \exp(\beta^T xy) \, dG_1(y)} \]

where \( G_1(y) = \int_{-\infty}^y h_2(y') \, dG(y') \) and \( \beta \) is the same as in (1).

Property 1 can be shown by combining (1), (3) and (4), and it generalizes the model-invariant property of 
the logistic regression model under case-control sampling (Prentice & Pyke, 1979). Suppose that a binary 
response \( Y \) follows a logistic regression model \( \logit[\Pr(Y = 1 \mid X = x)] = \alpha + \beta^T x \). This model is a special 
case of model (1) where the baseline distribution follows a Bernoulli distribution with success probability 
\( \exp(\alpha) / (1 + \exp(\alpha)) \). Case-control sampling is a special case of Example 2(a) with \( \Pr(R_X = 1 \mid Y = y) = \pi_y \) \( (y = 0, 1) \). It is well known that case-control data also follow a logistic regression model with the same 
\( \beta \) coefficients as in the population model but a different intercept \( \alpha^* = \alpha + \log(\pi_1 / \pi_0) \).

The second invariance property is for randomly truncated data, which are often collected in medical 
studies. Suppose that the response \( Y \) is observed only when \( T_1 \leq y \leq T_2 \), where \( T_1 \) and \( T_2 \) are random 
variables indicating the left and right truncation times. When \( Y \) is a time-to-event response, left truncation 
is often observed in cross-sectional sampling, where prevalent subjects have to be surviving at the 
recruitment time to be eligible to enter the study. Right truncation often occurs in retrospective sampling, 
where subjects are only observed if the failure event happens before the sampling time. Let \( f_{T_1, T_2}(t_1, t_2) \) 
be the joint distribution of \( (T_1, T_2) \), and let \( f_S(y \mid x) \) be the sampling conditional density for the randomly 
truncated data.

**Property 2.** If \( (T_1, T_2) \) is independent of \( (Y, X) \), then 
\[ f_S(y \mid x) = \frac{\exp(\beta^T xy) \, dG_2(y)}{\int \exp(\beta^T xy) \, dG_2(y)} \]

where \( G_2(y) = \int_{-\infty}^y \int_{-\infty}^{t_2} f_{T_1, T_2}(t_1, t_2) \, dt_2 \, dt_1 \, dG(y') \) and \( \beta \) is the same as in (1).

A derivation of Property 2 is given in the Appendix. If the population follows a proportional likelihood 
ratio model, randomly truncated data will also follow such a model, with the same target parameters as in 
the population model. Randomly truncated data can be used directly for estimating \( \beta \), even though the sample 
is biased. Property 2 holds under an independent truncation condition and is not true in general when the 
truncation distribution is covariate-dependent. We will discuss covariate-dependent truncation in § 5.

3. **PAIRWISE CONDITIONING AND ESTIMATION**

In this section, we propose an alternative method for estimating \( \beta \) in model (1), which eliminates the 
functional nuisance parameter \( G(y) \). By the invariance properties 1 and 2 discussed in § 2, the method 
also estimates the target parameters \( \beta \) in certain nonignorable missing data problems and for randomly 
truncated responses.
Suppose that we observe \((y_i, x_i) (i = 1, \ldots, n)\), which are independent and identically distributed as the random variables \((Y, X)\) where the conditional density function of \(Y\) given \(X\) follows model (1). Let \(y = (y_1, \ldots, y_n)\) and \(x = (x_1, \ldots, x_n)\), and let \(y_{(i)} = (y_{(1)}, \ldots, y_{(n)})\) be the order statistics for \((y_1, \ldots, y_n)\). Kalbfleisch (1978) showed that

\[
f(y \mid x, y_{(i)}) = \frac{\prod_{i=1}^{n} f(y_i \mid x_i)}{\sum_{j \in \sigma} \prod_{i=1}^{n} f(y_{(j)} \mid x_{(j)})},
\]

where \(j = (j_1, \ldots, j_n)\) and \(\sigma\) is the set of permutations of the integers \([1, \ldots, n]\). For model (1), we get

\[
f(y \mid x, y_{(i)}) = \frac{\exp(\beta^T \sum_{i=1}^{n} x_i y_i)}{\sum_{j \in \sigma} \exp(\beta^T \sum_{i=1}^{n} x_i y_{(j)}^i)}, \quad (5)
\]

which contains only the target parameters \(\beta\) but not the nuisance parameter \(G\). The conditional likelihood (5) provides a legitimate way of constructing likelihood inference for \(\beta\), but the computational burden is of order \(n!\). By applying (5) to a parametric generalized linear model with missing data, Liang & Qin (2000) proposed the pairwise pseudolikelihood

\[
L_{PL}(\beta) \propto \prod_{i < k} \frac{1}{1 + \exp(-\beta^T (x_i - x_k)(y_i - y_k))},
\]

which applies the conditioning argument of Kalbfleisch (1978) to every pair of observations. The computational burden is of order \(n^2\). Upon applying this argument to model (1), we have the pairwise likelihood

\[
L_{PL}(\beta) \propto \prod_{i < k} \frac{1}{1 + \exp(-\beta^T (x_i - x_k)(y_i - y_k))},
\]

which is again free of the functional parameter \(G\). We denote by \(\hat{\beta}_{PL}\) the value of \(\beta\) that maximizes \(L_{PL}(\beta)\), which is equivalent to the solution of the pseudo-score equation

\[
\frac{2}{n(n-1)} \sum_{i < k} (x_i^T - x_k^T)(y_i - y_k) \frac{\exp(-\beta^T (x_i - x_k)(y_i - y_k))}{1 + \exp(-\beta^T (x_i - x_k)(y_i - y_k))} = 0. \quad (6)
\]

The factor \(2/(n(n-1))\) is the reciprocal of the number of terms in the summation.

Before we discuss the statistical properties of \(\hat{\beta}_{PL}\) and a score test based on (6), let us introduce some notation to simplify the discussion. Let

\[
\psi_{ik}(\beta) = -(x_i^T - x_k^T)(y_i - y_k) \frac{\exp(-\beta^T (x_i - x_k)(y_i - y_k))}{1 + \exp(-\beta^T (x_i - x_k)(y_i - y_k))},
\]

and denote the score function by \(s_{PL}(\beta) = 2/(n(n-1)) \sum_{i < k} \psi_{ik}(\beta)\). The score function is a second-order U-statistic (Serfling, 1980). It follows from Proposition 1 in Liang & Qin (2000) that \(n^{1/2} (\hat{\beta}_{PL} - \beta_0)\) converges weakly to a multivariate normal distribution with covariance matrix \(V = \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1}\), where \(\Sigma_1 = E[\psi_{12}(\beta_0) \partial\beta]\) and \(\Sigma_2 = 4E[\psi_{12}(\beta_0) \psi_{13}(\beta_0)]\). To estimate the asymptotic variance, we could use a sandwich-type estimator \(\hat{V} = \hat{\Sigma}_1^{-1} \hat{\Sigma}_2 \hat{\Sigma}_1^{-1}\) with

\[
\hat{\Sigma}_1 = \frac{2}{n(n-1)} \sum_{i < k} \frac{\partial \psi_{ik}(\hat{\beta}_{PL})}{\partial \beta},
\]
and

\[
\hat{\Sigma}_2 = \frac{4}{n-1} \sum_{i=1}^{n} \left\{ \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} \psi_{ij}(\hat{\beta}_{PL}) \right\} - s_{PL}(\hat{\beta}_{PL}) \right\} \otimes^2,
\]

where \(a \otimes^2 = aa^T\) for a column vector \(a\). This estimator follows from the variance estimator of U-statistics proposed by Sen (1960).

An added advantage of the proposed method is that a simple score-type test can be constructed to test a null hypothesis \(H_0: \beta = \beta_0\) based on the pseudo-score function (6). Under the null hypothesis, \(s_{PL}(\beta_0)\) is a zero-mean U-statistic. Suppose that \(E\{\psi_{12}(\beta_0)\} < \infty\) and \(E\{\psi_{12}(\beta_0)\psi_{13}(\beta_0)^T\}\) is nonsingular; it then follows from Serfling (1980, p. 192) that \(n^{1/2}s_{PL}(\beta_0)\) converges weakly to a normal distribution with mean zero and covariance matrix \(\Sigma_2\). Under the null hypothesis, we could estimate \(\Sigma_2\) by

\[
\hat{\Sigma}_2(\beta_0) = \frac{4}{n-1} \sum_{i=1}^{n} \left\{ \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} \psi_{ij}(\beta_0) \right\} - s_{PL}(\beta_0) \right\} \otimes^2,
\]

and the test statistic

\[
T = ns_{PL}(\beta_0)\hat{\Sigma}_2(\beta_0)^{-1}s_{PL}(\beta_0)
\]

has a limiting \(\chi^2_p\) distribution.

4. Simulations

We conducted simulation studies to evaluate the finite-sample performance of the proposed estimator. First, we compared the performance of the proposed estimator \(\hat{\beta}_{PL}\) with the nonparametric likelihood estimator \(\hat{\beta}_{NL}\) of Luo & Tsai (2012) under the same simulation setting as in that paper. Next, we compared the performance of the proposed estimator and the least-squares estimator for a Gaussian-distributed response with nonignorable missing data and doubly truncated data. We also evaluated the empirical Type I error and power for the score test (7). We generated 1000 independent datasets in each scenario.

A continuous covariate \(X_2\) was generated from a normal distribution with mean zero and standard deviation 0.5. For a given \(X_2\), \(X_1\) was generated from a Bernoulli distribution with probability of success being \(\exp(1 - X_2)/(1 + \exp(1 - X_2))\). The finite-dimensional parameters are \((\beta_1, \beta_2) = (-1, -1)\), and we considered two baseline densities \(g(y) = dG(y)\). In scenario 1, \(g(y) = (2/\pi)^{1/2} \Phi(0.5)^{-1} \exp(-2(y - 1/4)^2)\); the response is a positive continuous random variable. In scenario 2, \(g(y) = (1 + y)^{3/2} \exp(-3)/(4y!);\) the response is a nonnegative integer-valued random variable.

The simulation results for both scenarios are shown in Table 1. The proposed estimator had a slightly higher sampling variability than did the maximum likelihood estimator and a similar small-sample bias. We conclude that even though the proposed estimator does not maximize a full likelihood, its performance is very close to that of the nonparametric maximum likelihood estimator for small samples. We also compared the computation times for simulation studies in scenario 1. For \(n = 50, 100\) and 200, it took on average 0.2, 0.9 and 3.4 seconds, respectively, to compute the proposed estimator, and 2.1, 11.6 and 84.7 seconds to compute the maximum likelihood estimator. Thus, using the proposed estimator led to a substantial gain in computational efficiency; we will discuss this further in § 5.

We then considered a scenario in which there were missing data and where the missing mechanism was nonignorable. We considered a Gaussian response and compared \(\hat{\beta}_{PL}\) with the least-squares estimator \(\hat{\beta}_{LS}\). Covariate data followed the same distribution as in the previous simulation, and \(g(y)\) was taken to be the standard normal density function. The response \(Y\) was observed with probability
Table 1. Comparison of $\hat{\beta}_{PL}$ and $\hat{\beta}_{NL}$

<table>
<thead>
<tr>
<th>n</th>
<th>$\hat{\beta}_{PL}$</th>
<th>$\hat{\beta}_{NS}$</th>
<th>$\hat{\beta}_{PL}$</th>
<th>$\hat{\beta}_{NS}$</th>
<th>$\hat{\beta}_{PL}$</th>
<th>$\hat{\beta}_{NS}$</th>
<th>$\hat{\beta}_{PL}$</th>
<th>$\hat{\beta}_{NS}$</th>
<th>$\hat{\beta}_{PL}$</th>
<th>$\hat{\beta}_{NS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>-145 -177</td>
<td>1059 1018</td>
<td>-170 -207</td>
<td>1090 1048</td>
<td>984 918</td>
<td>95 94</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>-56 -38</td>
<td>696 604</td>
<td>-66 -48</td>
<td>705 616</td>
<td>661 616</td>
<td>95 96</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>-40 -41</td>
<td>464 434</td>
<td>-44 -44</td>
<td>468 437</td>
<td>457 427</td>
<td>96 94</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Comparison of $\hat{\beta}_{PL}$ and $\hat{\beta}_{LS}$

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$\hat{\beta}_{PL}$</th>
<th>$\hat{\beta}_{LS}$</th>
<th>$\hat{\beta}_{PL}$</th>
<th>$\hat{\beta}_{LS}$</th>
<th>$\hat{\beta}_{PL}$</th>
<th>$\hat{\beta}_{LS}$</th>
<th>$\hat{\beta}_{PL}$</th>
<th>$\hat{\beta}_{LS}$</th>
<th>$\hat{\beta}_{PL}$</th>
<th>$\hat{\beta}_{LS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Nonignorable missingness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 100</td>
<td>-46 -70</td>
<td>451 377</td>
<td>430 367</td>
<td>90 89</td>
<td>96 95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 200</td>
<td>-53 -42</td>
<td>319 281</td>
<td>294 255</td>
<td>88 89</td>
<td>94 95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b) Double truncation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 100</td>
<td>-102 -109</td>
<td>623 610</td>
<td>601 585</td>
<td>89 88</td>
<td>94 93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 200</td>
<td>-61 -65</td>
<td>387 378</td>
<td>376 365</td>
<td>90 89</td>
<td>95 93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Percentage of hypotheses rejected by score test (7)

<table>
<thead>
<tr>
<th>(\beta_1, \beta_2)</th>
<th>(0, 0)</th>
<th>(1, 1)</th>
<th>(2, 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 50</td>
<td>6</td>
<td>31</td>
<td>86</td>
</tr>
<tr>
<td>n = 100</td>
<td>5</td>
<td>59</td>
<td>100</td>
</tr>
<tr>
<td>n = 200</td>
<td>5</td>
<td>89</td>
<td>100</td>
</tr>
</tbody>
</table>
5. Discussion

In this paper, we have discussed invariance properties and an alternative method for estimating the target parameters $\beta$ for a proportional likelihood ratio model that was proposed recently by Luo & Tsai (2012). The invariance properties are useful in situations where the missing data mechanism is probably nonignorable and where the analysts may not have enough knowledge to model the missing data mechanism or the truncation distribution. Such unknown quantities become part of the functional nuisance parameter. Any biased sampling (Gilbert et al., 1999) in the form $f_S(y \mid x) = b(y, x) f(y \mid x) \propto b(y, x) \exp(\beta^T xy) dG(y)$, where the data are doubly truncated with covariate-dependent truncation probability, fits into this general situation. In this case, a pairwise pseudolikelihood can be constructed when $b(y, x)$ is known and does not follow a factorization as in (4):

$$L_{PL}(\beta) \propto \prod_{i < k} \frac{1}{1 + R(y_i, y_k, x_i, x_k) \exp[-\beta^T (x_i - x_k)(y_i - y_k)]}$$

where

$$R(y_i, y_k, x_i, x_k) = \frac{b(y_i, x_k) b(y_k, x_i)}{b(y_i, x_i) b(y_k, x_k)}.$$

The proposed pseudolikelihood method has a lower computational burden than the maximum likelihood estimator. As can be seen from equation (4) in Luo & Tsai (2012), the computation is dominated by the last term in that expression. The computation of the constraint maximization involves a profiling estimator (Luo & Tsai, 2012, equation 5), so the dominating term will involve summation over three indices. When both $Y$ and $X$ are continuous the computational burden will be $O(n^3)$, while the estimator proposed in this paper has a computational burden of $O(n^2)$.

Acknowledgement

The author thanks the editor, an associate editor, two reviewers and Drs Patrick Heagerty, Ali Shojaie and Mary Lou Thompson for suggestions which have greatly improved this paper.

Appendix

Derivation of Property 2

Conditioning on $T_1 = t_1$ and $T_2 = t_2$, the sampling density for the randomly truncated data is $f(y \mid x)$ if $t_1 \leq y \leq t_2$ and 0 otherwise. Averaging over the joint distribution of truncation times $(T_1, T_2)$, the joint probability of $Y = y$ and $T_1 \leq Y \leq T_2$ is $\int_{-\infty}^{y} \int_{-\infty}^{y} f(y \mid x) f_{T_1, T_2}(t_1, t_2) \, dt_1 \, dt_2$, and the probability $\Pr(T_1 \leq Y \leq T_2 \mid X = x)$ is $\int_{-\infty}^{y} \int_{-\infty}^{y} f(y \mid x) f_{T_1, T_2}(t_1, t_2) \, dt_1 \, dt_2$. For randomly truncated data, $f_S(y \mid x) = f(y \mid x), T_1 \leq Y \leq T_2$, and therefore

$$f_S(y \mid x) = \frac{\int_{-\infty}^{y} \int_{-\infty}^{y} f(y \mid x) f_{T_1, T_2}(t_1, t_2) \, dt_1 \, dt_2}{\int_{-\infty}^{y} \int_{-\infty}^{y} f(y \mid x) f_{T_1, T_2}(t_1, t_2) \, dt_1 \, dt_2}$$

$$= \frac{\exp(\beta^T xy) \int_{-\infty}^{y} \int_{-\infty}^{y} f_{T_1, T_2}(t_1, t_2) \, dG(y)}{\int \exp(\beta^T xy) \int_{-\infty}^{y} \int_{-\infty}^{y} f_{T_1, T_2}(t_1, t_2) \, dG(y)}$$

$$= \frac{\exp(\beta^T xy) \, dG_2(y)}{\int \exp(\beta^T xy) \, dG_2(y)}.$$

References


[Received February 2012. Revised August 2012]