A note on the large sample properties of estimators based on generalized linear models for correlated pseudo-observations

Martin Jacobsen
Torben Martinussen

Research Report 14/04

Department of Biostatistics
University of Copenhagen
A note on the large sample properties of estimators based on generalized linear models for correlated pseudo-observations

Martin Jacobsen
Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5, 2100 København Ø
email: martin@math.ku.dk

Torben Martinussen
Department of Biostatistics
University of Copenhagen
Øster Farimagsgade 5B, 1014 Copenhagen K, Denmark
email: tma@sund.ku.dk

Summary
Pseudo-values have proven very useful in censored data analysis in complex settings such as multi state models. It was originally suggested by Andersen et al. (2003) that also suggested to estimate standard errors using classical generalized estimating equation results. These results were studied more formally in Graw et al. (2009) that derived some key results based on a second order von Mises expansion. However, results concerning large sample properties of estimates based on regression models for pseudo-values
still seem unclear. In this paper we study these large sample properties in the simple setting of survival probabilities and show that the estimating function can be written as a U-statistic of second order giving rise to an additional term that does not vanish asymptotically. We further show in the one dimensional case that previously advocated standard error estimates will typically be too large although in many practical applications the difference will be of minor importance. We show how to estimate correctly the variability of the estimator. This is further studied in some simulation studies.

Keywords: Pseudo-observations; Survival analysis; U-statistic; Von Mises expansion.

1 Introduction

When analyzing survival data the life time of interest is often right censored and hence only fully observed for a subset of the subjects at study. In practice, such analyses are typically based on models for the hazard function such as the Cox model (Cox, 1972). By doing so, (independent) censoring is easily handled as it is merely a question of modifying the at risk indicator, see for instance Andersen et al. (1993). The drawback, however, is that the hazard function may be difficult to model and to understand, and there may therefore be a wish of a more direct modeling of the response of interest. Pseudo-observations (Andersen et al., 2003) were proposed as such an alternative tool, the idea being to try somehow to replace the true but unknown survival time by a pseudo-observation. For a survival time, $\tilde{T}$, if we want to estimate the parameter $\theta = E\{f(\tilde{T})\}$ for some function $f$ then if all survival times were observed this could be achieved using the sample mean $n^{-1} \sum_k f(\tilde{T}_k)$. In the more realistic case where the are censorings this is no longer possible, but we may still have an unbiased (asymptotically) estimator $\hat{\theta}$;
for instance the Kaplan-Meier estimator if $\theta = S(t)$ is the survival function (evaluated at the specific time-point $t$), that is by letting $f(\tilde{T}) = I(\tilde{T} > t)$, where $I()$ is the indicator function. The pseudo-observations are then defined in the following way. The $k$th pseudo-observation $\hat{\theta}_k$ (Andersen et al. 2003) is given by

$$\hat{\theta}_k = n\hat{\theta} - (n - 1)\hat{\theta}_{-k},$$

where $\hat{\theta}_{-k}$ is the estimator $\hat{\theta}$ but calculated based on the sample where we leave out the $k$th observation. Andersen et al. (2003) suggested to take this a step further and to build regression models using the pseudo-observations as responses in the generalized linear model

$$g[E\{f(\tilde{T})|Z\}] = Z^T\beta,$$

where $Z$ is a $p$-dimensional covariate vector, $\beta$ is a $p$-vector of regression parameters, and $g$ is a link-function. Andersen et al. (2003) further suggested to use standard results from generalized estimating equations theory to estimate the variance of the estimator $\hat{\beta}$ exploiting that the pseudo-observations are marginally unbiased of $\theta$. However, when conditioning on the covariates it is not clear whether the unbiasedness remain. This was investigated in more detail by Graw et al. (2009) using a second order von Mises expansion. They showed that the suggested estimating equation in fact has zero mean, and they also argued that the variance estimator previously suggested is valid. However, inspecting their proof, this conjecture is unclear.

In this paper we revisit the simple situation where the parameter of interest is the conditional survival function and show that the previously suggested variance estimator is not consistent as an additional term is needed in the variance of the estimating function. This is seen by establishing the estimating function as a U-statistic of second order. We show in the one-dimensional case that the usual variance estimator will produce a too
large estimate of the variance leading to conservative tests. Based on numerical studies it also appears, however, that the bias is typically small in many settings.

The paper is structured as follows. In the next section we describe estimation based on pseudo-observations and give the results of Andersen et al. (2003). In Section 3 we study the asymptotic behavior of the normed estimating function and show that it can be written as a U-statistic of order 2. In Section 4 we study the one-dimensional case in detail showing that the variance term of the estimating function corresponding to the second order influence functions cannot be ignored, and that the usual variance estimator is biased upwards. Section 5 gives the new estimator of the variance of the regression estimator, and in Section 6 we study by simulations the behavior of this new estimator and compare it to the usual estimator. Section 7 contains some closing remarks and technical details are deferred to the Appendix in Section 8.

2 Estimation based on pseudo-observations

Let $Z$ denote a $p$-dimensional covariate vector, and let $\tilde{T}$ denote the survival time that may be censored by $U$. Put $T = \tilde{T} \wedge U$, $\Delta = I(\tilde{T} \leq U)$ and $X = (T, \Delta)$. It is assumed that $\tilde{T}$ and $U$ are conditionally independent given $Z$, and that $U$ and $Z$ are independent. The marginal hazard function of $\tilde{T}$ is denoted by $\lambda(t)$ and the hazard function of $U$ is denoted by $b(t)$.

Let $(X_k, Z_k), k = 1, \ldots, n$ denote $n$ iid replicates. For simplicity, we focus on the following model parameter (fixing time $t_0$)

$$\theta_k = P(\tilde{T}_k > t_0 | Z_k) = S(t_0 | Z_k),$$

and assume that there exists a link function $g$ so that

$$g\{S(t_0 | Z_k)\} = Z_k^T \beta,$$
where $\beta$ is the vector of regression parameters of interest. The covariate vector may contain an intercept term. Andersen et al. (2003) suggested to estimate $\beta$ using pseudo-observations as the responses in a generalized estimating equation approach in the following way. Let $\theta = S(t_0) = P(\hat{T}_k > t_0)$ and let $\hat{\theta} = \hat{S}(t_0)$ denote the Kaplan-Meier estimator (at time point $t_0$). The $k$th pseudo-observation $\hat{\theta}_k$ is given by

$$\hat{\theta}_k = n\hat{\theta} - (n - 1)\hat{\theta}_{-k},$$

where $\hat{\theta}_{-k}$ is the estimator $\hat{\theta}$ but calculated based on the sample where we leave out the $k$th observation. Andersen et al. (2003) suggested to estimate $\beta$ based on the estimating equation

$$U(\beta) = \sum_k (\hat{\mu}_k)^T V_k^{-1} (\hat{\theta}_k - \theta_k) = 0,$$

where $\hat{\mu}_k$ denotes the derivative with respect to $\beta$ of

$$\mu_k = \mu(Z_k^T \beta) = g^{-1}(Z_k^T \beta),$$

and $V_k$ is some weight matrix usually chosen as the identity matrix. The estimator $\hat{\beta}$ is the solution to estimating equation $U(\hat{\beta}) = 0$, and it is suggested to use a sandwich estimator to estimate the variance of $\hat{\beta}$ in the following way (Andersen et al., 2003; Graw et al., 2009). Let

$$I(\beta) = \sum_k (\hat{\mu}_k)^T V_k^{-1} (\hat{\mu}_k),$$

$$\text{vâr}\{U(\beta)\} = \sum_k U_k(\hat{\beta}) U_k(\hat{\beta})^T,$$

then

$$\text{vâr}(\hat{\beta}) = I(\hat{\beta})^{-1} \text{vâr}\{U(\beta)\} I(\hat{\beta})^{-1}.$$  

A proof of this results was given in Graw et al. (2009), but a closer look at the proof of their Theorem 2 makes it unclear whether this really holds true. This will be investigated in the following under the specific considered setting.
3 Asymptotic distribution of the normed estimation function

In this section we study the asymptotic distribution of the normed estimation function $n^{-1/2}U(\beta)$. Let the survival functions of $\tilde{T}_k$ and $U_k$ be denoted by $S$ and $G$, respectively. Let $H = SG$ and

$$N_k(t) = I(T_k \leq t, \Delta_k = 1), \quad Y_k(t) = I(t \leq T_k).$$

Let $\psi(P)$ denote the statistical parameter $S(t)$, and let $P_n$ denote the empirical distribution. As in Graw et al. (2009) we can use the Von Mises expansion to write the pseudo-observation as

$$\hat{\theta}_k = n\psi(P_n) - (n - 1)\psi(P_n(k))$$

$$= \theta + \dot{\psi}(X_k) - \frac{1}{2n(n - 1)} \sum_{i,j} \ddot{\psi}(X_i, X_j) + \frac{1}{n - 1} \sum_i \dddot{\psi}(X_i, X_k)$$

$$- \frac{1}{2(n - 1)} \dot{\psi}(X_k, X_k) + R_n \quad (2)$$

In the latter display, $\dot{\psi}(X_k)$ and $\ddot{\psi}(X_i, X_k)$ denote the first and second order influence functions. For more details on von Mises expansion and influence functions, see Serfling (1980) Ch. 6. The remainder term $R_n$ contains higher order terms which we will return to later, but for the moment assume that they can be ignored. The first order influence function is given by

$$\dot{\psi}(X_k) = -\theta \int_0^t \frac{1}{H(s)} \, dM_k(s),$$

where

$$M_k(t) = N_k(t) - \int_0^t Y_k(s) \, d\Lambda(s), \quad k = 1, \ldots, n,$$

are the counting process martingales w.r.t. $\mathcal{F}_t$ the history spanned by the counting processes that is without conditioning on the covariates. Here we use the notation
\[ \Lambda(t) = \int_0^t \lambda(s) \, ds \] for the marginal cumulative hazard function; the conditional one is denoted by \( \Lambda_Z(t) \). For further details see the Appendix where we also give the explicit expression for the second order influence function. Using the representation of \( \dot{\psi} \), see the Appendix, it is easy to see that all the second order terms in the latter display converge in probability to zero as \( n \) tends to infinity. Thus

\[ \hat{\theta}_k = \theta + \dot{\psi}(X_k) + o_p(1). \]

It further follows, see Proposition 1 in the Appendix, that

\[ E(\dot{\psi}(X_k)|Z_k) = \theta_k - \theta, \]

which is what is needed to conclude that the estimating function has mean zero. We may now study the asymptotic behavior of the normed estimating function:

\[ n^{-1/2}U(\beta) = n^{-1/2} \sum_k A_k(Z_k)(\hat{\theta}_k - \theta_k), \]

where we use the general notation \( A_k(Z_k) \), which can be any function of \( Z_k \) such as \((\mu_k)^T V_k^{-1}\). Using (2), we can write the normed estimating function as

\[ n^{-1/2}U(\beta) = n^{-1/2} \sum_k A_k(Z_k)(\theta + \dot{\psi}(X_k) - \theta_k) + W_n^1 + W_n^2, \]

where

\[ W_n^1 = n^{-1/2} \sum_k A_k(Z_k)(n-1)^{-1} \sum_i \ddot{\psi}(X_i, X_k), \]

and where \( W_n^2 \) contains the remaining terms coming from (2). We argue in the Appendix that \( W_n^2 \) converges in probability to zero, and thus can be ignored. If \( W_n^1 \) also converges in probability to zero then the usual variance estimator would be in order. However, as we shall see, this is not the case as \( W_n^1 \) converges in distribution to a normal variate with variance larger than zero.
3.1 Writing the estimating function as a U-statistic

We use the notation \( X^*_k = (X_k, Z_k) \) and \( \tilde{\psi}(X^*_k) = \theta + \psi(X_k) - \theta(t_0|Z_k) \), and hence \( \mathbb{E}\{A(Z_k)\tilde{\psi}(X^*_k)\} = 0 \). We now write

\[
n^{-1}U(\beta) = \frac{1}{n-1} \sum_k A(Z_k)\tilde{\psi}(X^*_k) + n^{-2} \sum_k \sum_i A(Z_k)\tilde{\psi}(X_i, X_k)
\]

as a U-statistic of order 2 with kernel \( h \) given by

\[
2h(X^*_1, X^*_2) = A(Z_1)\tilde{\psi}(X^*_1) + A(Z_2)\tilde{\psi}(X^*_2) + \{A(Z_1) + A(Z_2)\}\tilde{\psi}(X_1, X_2).
\]

This follows since

\[
n^{-1}U(\beta) = n^{-1} \sum_k A(Z_k)\tilde{\psi}(X^*_k) + n^{-2} \sum_k \sum_i A(Z_k)\tilde{\psi}(X_i, X_k) = \sum_a \tilde{h}(X^*_{a_1}, X^*_{a_2}),
\]

where the sum is taken over the set of all unordered subsets \( a \) of 2 different integers chosen from \( \{1, \ldots, n\} \). In the latter display,

\[
\tilde{h}(X^*_1, X^*_2) = A(Z_1)\tilde{\psi}(X^*_1) - \frac{n}{(n-1)n^2} + A(Z_2)\tilde{\psi}(X^*_2) - \frac{n}{(n-1)n^2} + \{A(Z_1) + A(Z_2)\}\tilde{\psi}(X_1, X_2) \frac{1}{n^2}
\]

\[
+ A(Z_1)\tilde{\psi}(X_1, X_1) \frac{1}{(n-1)n^2} + A(Z_2)\tilde{\psi}(X_2, X_2) \frac{1}{(n-1)n^2},
\]

and where the latter two terms can be ignored as they are of lower order. It now follows from Van der Vaart (1998) that \( n^{-1/2}U(\beta) \) is asymptotical normal with mean zero and variance \( 4\zeta_1 \), where

\[
\zeta_1 = \text{cov}(h(X^*_1, X^*_2), h(X^*_1, X^*_3)).
\]

It further follows that

\[
n^{-1/2}U(\beta) = n^{-1/2} \sum_k h_1(X^*_k) + o_p(1), \quad (3)
\]
where we in the Appendix show that
\[ h_1(X^*_t) = h_{11}(X^*_t) + h_{12}(X^*_t). \] (4)

In (4),
\[ h_{11}(X^*_t) = A(Z_1)\tilde{\psi}(X^*_t), \]
and
\[ h_{12}(X^*_t) = \int_0^t \phi_1(s) dM_1(s) - \theta \int_0^t \frac{H(s) - Y_1(s)}{H(s)} d\mu(s) + \phi_0 \int_0^t \frac{1}{H(s)} dM_1(s) \]
with
\[ \phi_1(t) = \theta E\left\{A(Z_k)\frac{\theta_k(t) - \theta(t)}{\theta(t)}\right\}, \quad \theta_k(t) = S(t | Z_k), \quad \theta(t) = S(t), \]
\[ \phi_0 = -E\{A(Z_k)(\theta_k - \theta)\}, \]
and \(d\mu(s) = E\{J(s, Z)\} ds,\) where
\[ J(t, Z) = A(Z)e^{-(\Lambda_Z(t) - \Lambda(t))}\{\lambda_Z(t) - \lambda(t)\}.\]

In the appendix it is further shown that
\[ \phi_0 = \theta \mu([0, t_0]), \quad \phi_1(t) = -\theta \mu([0, t]). \]

The variance of \(h_1(X^*_t)\) is given by \(E[h_1(X^*_t)h_1(X^*_t)^T]\), where \(E[h_{11}(X^*_t)h_{11}(X^*_t)^T]\) gives the term coming from the first order term that is used in the usual variance estimator of Andersen et al. (2003).

4 The one-dimensional case

Here we study more closely whether the second order term of (4) can be ignored in the one-dimensional case. Hence we assume for the moment that \(p = 1\). The first result concerns the variance of the second order term of (4).
Theorem 1 The variance \( \text{var}\{h_{12}(X^*_1)\} \) is \( O(n^{-1}) \) if and only if either there is no censoring on \([0,t_0]\) \((b \equiv 0 \text{ Lebesgue a.e on } [0,t_0])\) or

\[
E\{J(s,Z)\} = 0
\]  

for \( b \)-almost every \( s \in [0,t_0] \).

Remark. Because \( S(t) = E\{P(T > t|Z)\} \) it is seen directly that (5) is satisfied if the function \( A \) is constant. If \( Z \) has no effect on \( \hat{T} (\beta = 0) \) then (5) is also satisfied.

The remarkable connection between the first and second order variances is the following:

Theorem 2 For all \( t_0 \), all choices of hazard functions \( \lambda, \lambda_z, b \) and functions \( A \) with \( E\{|A(Z)|\} < \infty \)

\[
cov\{h_{11}(X^*_1), h_{12}(X^*_1)\} = -\text{var}\{h_{12}(X^*_1)\}. \tag{6}
\]

In particular

\[
\text{var}\{h_1(X^*_1)\} = \text{var}\{h_{11}(X^*_1)\} - \text{var}\{h_{12}(X^*_1)\}.
\]

From Theorem 2 we see that the usual variance estimator is upward biased. In Section 6 we investigate the magnitude of the bias in a numerical study.

5 Estimating the variance of \( \hat{\beta} \)

The variance, \( \Sigma_U \) say, of the estimating function evaluated at the true \( \beta \) can be estimated by

\[
\hat{\Sigma}_U = n^{-1}\sum_k \hat{h}_1(X^*_k)^T \hat{h}_1(X^*_k),
\]

where

\[
\hat{h}_1(X^*_1) = \hat{h}_{11}(X^*_1) + \hat{h}_{12}(X^*_1)
\]
with

$$h_{11}(X_i^*) = A(Z_1)\{\hat{\theta} - \hat{\theta} \int_0^{t_0} \frac{1}{H(s)} d\hat{M}_1(s) - \mu(Z_1^T \hat{\beta})\},$$

$$h_{12}(X_i^*) = \int_0^{t_0} \frac{\hat{\phi}_1(s)}{H(s)} d\hat{M}_1(s) - \hat{\theta} \int_0^{t_0} \frac{\hat{H}(s) - Y_1(s)}{H(s)} d\hat{\mu}(s) + \hat{\phi}_0 \int_0^{t_0} \frac{1}{H(s)} d\hat{M}_1(s).$$

In the latter display,

$$d\hat{M}_1(t) = dN_1(t) - Y_1(t) d\hat{\Lambda}(t), \quad \hat{\phi}_0 = -n^{-1} \sum_k A(Z_k)\{\mu(Z_k^T \hat{\beta}) - \hat{\theta}\},$$

$$\hat{\phi}_1(t) = -\hat{\theta} n^{-1} \sum_k A(Z_k) \frac{\hat{H}(t) - Y_k(t)}{\hat{H}(t)}, \quad d\hat{\mu}(t) = n^{-1} \sum_k A(Z_k) \frac{1}{\hat{H}(t)} d\hat{M}_k(t).$$

The variance of $\hat{\beta}$ can thus be estimated by

$$\hat{U}(\hat{\beta})^{-1} \hat{\Sigma} U(\hat{\beta})^{-1}.$$

### 6 Numerical results

To investigate the properties of our proposed methods and the size of the bias of the usual variance estimator, we conducted a simulation study. We generated 10000 data sets for various combinations of sample size and model parameters as described below. Let $V$ and $Z$ be standard normals with correlation $\rho$, and let $\tilde{T} = -\log(1 - \Phi(V))$ giving us that $\tilde{T} \sim \text{Exp}(1)$. One can then calculate

$$P(\tilde{T} > t_0 | Z) = 1 - \Phi(k - \beta Z) = \mu(Z, \beta)$$

where

$$k = \Phi^{-1}(1 - e^{-t_0}) / \sqrt{1 - \rho^2}, \quad \beta = \rho / \sqrt{1 - \rho^2}.$$

Censoring was generated as $U \sim \text{Exp}(b)$, where $b$ is chosen so that there is about 75% censoring. To stay in the one dimensional case we only estimated $\beta$ in the above model.
thus setting $k$ to its true value. In the estimation of $\beta$ we will consider two estimating functions, one where we let $A(Z) = Z$, and the other equal to one suggested by Andersen et al. (2003), i.e. the usual generalized estimating function. Estimation of s.e. was done using the usual estimator that corresponds to only using the first order term denoted ($se_1$) and the one where we also use the second order terms ($se_2$). Table 1 gives the result for various combinations of $(n, \rho, \beta)$, and where we took $A(Z) = Z$ and $t_0 = 2/3$. Results where $A(Z) = \mu(Z\beta)$ (GEE) are given in Table 2. It is seen from Table 1 and 2 that when (5) is not violated ($\beta = 0$) then both estimators agree (as they should) and give correct coverage. This also holds when $\beta$ is not too large. It also seen, however, that when $\beta$ is large then, as predicted, the usual variance estimator produces a too large estimate giving too large coverage; this is especially seen then $A(Z) = Z$, where there can be a huge difference between the variability of the score and the estimator based only on first order terms.
Table 1: The weight function is $A(Z) = Z$. Results are based on 10000 runs. The $\overline{\beta}$ gives the mean of the estimated $\hat{\beta}$, $\var_U$ is the empirical variance of the estimating function evaluated at true $\beta$, $\var_U\beta$ is the mean of the estimated variance of the score based only on first order term ($h_{11}$), $\var_U\beta$ is the mean of the estimated variance of the score based on first and second order terms ($h_1$), sd is standard deviation of the $\hat{\beta}$'s, $se_1$ is the mean of the estimated s.e. of $\hat{\beta}$ using only first order term, $se_2$ is the mean of the estimated s.e. of $\hat{\beta}$ using first and second order terms, $cov_1\%$ gives 95% coverage probability using $se_1$ and $cov_2\%$ gives 95% coverage probability using $se_2$.
Table 2: The weight function is $A(Z) = GEE$. Results are based on 10000 runs. The mean $\hat{\beta}$ gives the mean of the estimated $\hat{\beta}$, $\text{var}_U$ is the empirical variance of the estimating function evaluated at true $\beta$, $\text{var}1_U$ is the mean of the estimated variance of the score based only on first order term ($h_{11}$), $\text{var}2_U$ is the mean of the estimated variance of the score based on first and second order terms ($h_1$), sd is standard deviation of the $\hat{\beta}$’s, $se_1$ is the mean of the estimated s.e. of $\hat{\beta}$ using only first order term, $se_2$ is the mean of the estimated s.e. of $\hat{\beta}$ using first and second order terms, $cov_1\%$ gives 95% coverage probability using $se_1$ and $cov_2\%$ gives 95% coverage probability using $se_2$.

<table>
<thead>
<tr>
<th>$(n, \rho, \beta)$</th>
<th>mean</th>
<th>$\text{var}_U$</th>
<th>$\text{var}1_U$</th>
<th>$\text{var}2_U$</th>
<th>sd</th>
<th>$se_1$</th>
<th>$se_2$</th>
<th>$cov_1%$</th>
<th>$cov_2%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(800,0,0)$</td>
<td>0</td>
<td>0.144</td>
<td>0.137</td>
<td>0.133</td>
<td>0.085</td>
<td>0.084</td>
<td>0.083</td>
<td>95.4</td>
<td>94.8</td>
</tr>
<tr>
<td>$(1600,0,0)$</td>
<td>0</td>
<td>0.140</td>
<td>0.138</td>
<td>0.136</td>
<td>0.059</td>
<td>0.059</td>
<td>0.059</td>
<td>95.3</td>
<td>95.0</td>
</tr>
<tr>
<td>$(3200,0,0)$</td>
<td>0</td>
<td>0.141</td>
<td>0.139</td>
<td>0.138</td>
<td>0.042</td>
<td>0.042</td>
<td>0.042</td>
<td>94.9</td>
<td>94.8</td>
</tr>
<tr>
<td>$(800,0.2,0.20)$</td>
<td>0.20</td>
<td>0.123</td>
<td>0.118</td>
<td>0.114</td>
<td>0.089</td>
<td>0.088</td>
<td>0.086</td>
<td>95.4</td>
<td>94.8</td>
</tr>
<tr>
<td>$(1600,0.2,0.20)$</td>
<td>0.20</td>
<td>0.117</td>
<td>0.119</td>
<td>0.116</td>
<td>0.061</td>
<td>0.062</td>
<td>0.061</td>
<td>95.8</td>
<td>95.4</td>
</tr>
<tr>
<td>$(3200,0.2,0.20)$</td>
<td>0.20</td>
<td>0.119</td>
<td>0.119</td>
<td>0.118</td>
<td>0.043</td>
<td>0.044</td>
<td>0.043</td>
<td>95.1</td>
<td>94.8</td>
</tr>
<tr>
<td>$(1600,0.7,0.98)$</td>
<td>0.99</td>
<td>0.021</td>
<td>0.024</td>
<td>0.021</td>
<td>0.117</td>
<td>0.124</td>
<td>0.116</td>
<td>96.2</td>
<td>94.9</td>
</tr>
<tr>
<td>$(3200,0.7,0.98)$</td>
<td>0.98</td>
<td>0.022</td>
<td>0.024</td>
<td>0.022</td>
<td>0.083</td>
<td>0.087</td>
<td>0.082</td>
<td>96.2</td>
<td>95.1</td>
</tr>
<tr>
<td>$(3200,0.9,2.06)$</td>
<td>2.09</td>
<td>0.0043</td>
<td>0.0050</td>
<td>0.0042</td>
<td>0.224</td>
<td>0.234</td>
<td>0.214</td>
<td>96.6</td>
<td>94.8</td>
</tr>
<tr>
<td>$(6400,0.9,2.06)$</td>
<td>2.08</td>
<td>0.0043</td>
<td>0.0050</td>
<td>0.0042</td>
<td>0.155</td>
<td>0.164</td>
<td>0.151</td>
<td>96.7</td>
<td>94.9</td>
</tr>
</tbody>
</table>
7 Closing remarks

Pseudo observations have become popular within survival analysis as they allow for the use of standard regression techniques that otherwise cannot be used because of censored observations. It is hence important to have results on the asymptotic behavior of estimates derived from this approach. In this paper we studied the case based on the survival function and gave a detailed analysis of the asymptotic behavior showing that some second order terms cannot always be ignored. The same technique should be possible to use when other estimands are of interest but the specific calculations will change according to the specific influence functions.

The iid-representation of the estimating function where we let \( \epsilon_k = h_1(X_k^*) \), see (3),

\[
n^{-1/2}U_n(\beta) = n^{-1/2} \sum_{k=1}^{n} \epsilon_k + o_p(1)
\]

may be used to perform goodness-of-fit. One might for instance consider the following process

\[
V_n(z) = n^{-1/2} \sum_k \left\{ \hat{\theta}_k - \theta_k(Z_k; \hat{\beta}) \right\} I(Z_{kj} \leq z)
\]

that can be written in the following way

\[
V_n(z) = n^{-1/2} \sum_k \left\{ \hat{\theta}_k - \theta_k(Z_k; \beta) \right\} I(Z_{kj} \leq z) - \left[ n^{-1} \sum_i D_{\beta} \theta_k(Z_i; \beta) I^{-1}(\hat{\beta}) I(Z_{ij} \leq z) \right] \epsilon_k
\]

We thus have the iid-representation of \( V_n(z) \):

\[
V_n(z) = n^{-1/2} \sum_k \nu_k(z) + o_p(1),
\]

where

\[
\nu_k(z) = \delta_k(z) - \left[ \sum_i n^{-1} D_{\beta} \theta_k(Z_i; \beta) I^{-1}(\hat{\beta}) I(Z_{ij} \leq z) \right] \epsilon_k
\]
with $\delta_k(z) = \{\hat{\theta}_k - \theta_k(Z_k; \beta)\}I(Z_k \leq z)$. It furthermore follows, given data, that

$$\hat{V}_n(z) = n^{-1/2} \sum_{k=1}^n \hat{\nu}_k(v)G_k,$$

converges to the same limit distribution as $V_n(z)$. In the latter display $G_1, \ldots, G_n$ are iid standard normals and $\hat{\nu}_k(v)$ is obtained from $\nu_k(v)$ inserting empirical quantities.

**Acknowledgement**

We thank Per Kragh Andersen and Thomas Gerds for helpful discussions and comments.

**Appendix**

The first and second order influence functions for the parameter $\Lambda(t)$ are given by (James, 1997)

$$\dot{\Gamma}(X_i) = T^{(1)}(X_i, \Lambda(t)) = \int_0^t \frac{1}{H(s)}dM_i(s)$$

$$\ddot{\Gamma}(X_i, X_j) = T^{(2)}(X_i, X_j, \Lambda(t)) = \int_0^t \frac{H(s) - Y_j(s)}{H^2(s)}dM_i(s) + \int_0^t \frac{H(s) - Y_i(s)}{H^2(s)}dM_j(s),$$

where

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)d\Lambda(s), \quad i = 1, \ldots, n,$$

are the counting process martingales w.r.t. $\mathcal{F}_t$ the history spanned by the counting processes. The first and second order influence functions for the parameter $S(t)$ are given by (James, 1997)

$$T^{(1)}(X_i, S(t)) = -S(t)T^{(1)}(X_i, \Lambda(t))$$

$$T^{(2)}(X_i, X_j, S(t)) = -S(t)\{T^{(2)}(X_i, X_j, \Lambda(t)) - T^{(1)}(X_i, \Lambda(t))T^{(1)}(X_j, \Lambda(t))$$

$$+ \sum_{s \leq t} \Delta T^{(1)}(X_i, \Lambda(t))\Delta T^{(1)}(X_j, \Lambda(t))I(i = j)\}$$
since, for $i \neq j$,
\[
\Delta T^{(1)}(X_i, \Lambda(t))\Delta T^{(1)}(X_j, \Lambda(t)) = 0
\]
when $\Lambda(t) = \int_0^t \lambda(s) \, ds$, and
\[
\sum_{s \leq t} \Delta T^{(1)}(X_i, \Lambda(t))\Delta T^{(1)}(X_i, \Lambda(t)) = \int_0^t \frac{1}{H^2(s)} \, dN_i(s).
\]

We denote $T^{(1)}(X_i, S(t))$ and $T^{(2)}(X_i, X_j, S(t))$ as $\dot{\psi}(X_i)$ and $\ddot{\psi}(X_i, X_j)$, respectively.

We now argue that $W_n^2$ converges to zero in probability. Let $\phi_3 = E\{A_k(Z_k)\}$. The leading term of the second term of $W_n^2$ is
\[
\theta n^{-1/2} \sum_k A_k(Z_k) \frac{1}{2n(n-1)} \sum_{i,j} \ddot{\Gamma}(X_i, X_j)
\]
that is asymptotically equivalent to
\[
\theta \phi_3 n^{-1/2} \sum_{k=1}^n \int_0^t \hat{\phi}_2(s) \, dM_k(s)
\]
which converges to zero in probability by use of Lenglart’s inequality (Martinussen and Scheike, 2006, p. 41). In the latter display,
\[
\hat{\phi}_2(t) = \frac{1}{n} \sum_{i=1}^n \frac{H(t) - Y_i(t)}{H(t)}.
\]

Likewise one can show that the higher order terms can be dropped. The only non-trivial terms are the third order terms that can be dealt with using the expressions given on p. 1615 in James (1997).

**U-statistic**

\[
h_1(x) = E\{2h(x, X_2^*)\}.
\]
We now calculate the latter one.

\[ h_1(x) = A(z) \tilde{\psi}(x) + E \left[ \{ A(z) + A(Z_2) \} \tilde{\psi}(x, X_2) \right] \]

\[ = A(z) \tilde{\psi}(x) + E \{ A(Z_2) \tilde{\psi}(x, X_2) \} \]

Further, with \( T^{(1)}(X_2, \Lambda(t)) = \int_0^t \frac{1}{H(s)} dM_2(s) \),

\[ E \{ A(Z_2) \tilde{\psi}(x, X_2) \} = -\theta E(A(Z_2) \int_0^t \frac{H(s) - \theta(s|Z_2)G(s)}{H^2(s)} dM(s) \]

\[ - \theta E(A(Z_2) \int_0^t \frac{H(s) - Y(s)}{H^2(s)} dM_2(s)) \]

\[ + \theta T^{(1)}(x, \Lambda(t)) E(A(Z_2)T^{(1)}(X_2, \Lambda(t)), \]

where \( M(t) \) and \( Y(t) \) in the latter display are non-stochastic because of the \( x \) in \( \tilde{\psi}(x, X_2) \).

Here,

\[ -\theta E(A(Z_2) \frac{H(s) - \theta(s|Z_2)G(s)}{H^2(s)} = \phi_1(s)/H(s), \]

\[ \theta E(A(Z_2)T^{(1)}(X_2, \Lambda(t))) = -E(A(Z_2)(\theta(t|Z_2) - \theta)) = \phi_0, \]

and

\[ -\theta E(A(Z_2) \int_0^t \frac{H(s) - Y(s)}{H^2(s)} dM_2(s)) = -\theta \int_0^t \frac{H(s) - Y(s)}{H(s)} d\mu(s) \]

with

\[ d\mu(s) = E \{ A(Z)e^{-\Lambda_2(s)\Lambda(s)}d(\Lambda_2(s) - \Lambda(s)) \}. \]

Hence,

\[ h_1(X^*_1) = A(Z_1) \tilde{\psi}(X^*_1) + \int_0^t \frac{\phi_1(s)}{H(s)} dM_1(s) - \theta \int_0^t \frac{H(s) - Y_1(s)}{H(s)} d\mu(s) + \phi_0 \int_0^t \frac{1}{H(s)} dM_1(s). \]
First and second order variances and the covariance

The proofs of the two theorems are based on finding the second order variance $Eh_{12}^2(X^*)$ and the covariance $Eh_{11}(X^*)h_{12}(X^*)$. We rewrite $h_{11}$ and $h_{12}$ as follows

$$h_{11}(X^*) = A(Z)S(t_0) \left( 1 - \frac{S_Z(t_0)}{S(t_0)} \right) - \int_0^{t_0} \frac{1}{H} dM,$$

(7)

$$h_{12}(X^*) = S(t_0) \left( \int_0^{t_0} \frac{\mu([u,t])}{H(u)} dM(u) - \int_0^{t_0} \left( 1 - \frac{Y(s)}{H(s)} \right) d\mu(s) \right),$$

(8)

$$= S(t_0) \int_0^{t_0} \left\{ \int_0^s \frac{1}{H(u)} dM(u) - \left( 1 - \frac{Y(s)}{H(s)} \right) \right\} d\mu(s),$$

(9)

where we have used the notation $S_Z(t) = S(t|Z)$. Here the last two formulas follow by noting that

$$\phi_1(t) = S(t_0)EA(Z) \left( -1 + \frac{S_Z(t)}{S(t)} \right) = -S(t_0)\mu([0,t])$$

$$\phi_0 = -S(t_0)EA(Z) \left( \frac{S_Z(t_0)}{S(t_0)} - 1 \right) = S(t_0)\mu([0,t_0]).$$

We shall write

$$\delta_Z = \lambda_Z - \lambda, \quad q_Z = \delta_Z \frac{H_Z}{H}$$

so that

$$\mu(ds) = f_\mu(s) \, ds$$

with the density $f_\mu$ given by

$$f_\mu(s) = EA(Z) q_Z(s).$$

Note that

$$\left( \frac{H_Z}{H} \right)' = \left( \frac{S_Z}{S} \right)' = -q_Z.$$

(10)

We shall frequently switch from the basic martingale $M$ to the conditional (given $Z$) martingale $M_Z$ through the identity

$$dM = dM_Z + \delta_Z Y \, ds$$
and first use this to derive an identity employed frequently below: for all \(0 \leq s, u \leq t_0\),
\[
E \left[ Y(u) \int_0^s \frac{1}{H} \, dM_Z \mid Z \right] = -H_Z(u) \int_0^{s \wedge u} \frac{\lambda_Z}{H} \, dv. \tag{11}
\]
To see this, first use the martingale property to see that on the left one may replace \(s\) by \(s \wedge u\). But if \(Y(u) = 1\) we have
\[
\int_0^{s \wedge u} \frac{1}{H} \, dM_Z = -\int_0^{s \wedge u} \lambda_Z Y \, dv
\]
and it only remains to use that \(E[Y(u) \mid Z] = H_Z(u)\). ((11) has an obvious generalisation where the deterministic function \(H\) is replaced by an arbitrary deterministic function (or even a suitably integrable predictable process). This is used below in the proof of Theorem 1.)

The first result we establish is the identity \(E \left( \psi(X) \mid Z \right) = \theta_k - \theta\). It is equivalent to (12).

**Proposition 1**  It holds that
\[
E \left[ \int_0^{t_0} \frac{1}{H} \, dM \mid Z \right] = 1 - \frac{S_Z(t_0)}{S(t_0)}. \tag{12}
\]

**Proof.**
\[
E \left[ \int_0^{t_0} \frac{1}{H} \, dM \mid Z \right] = E \left[ \int_0^{t_0} \frac{1}{H} \, dM_Z + \int_0^{t_0} \frac{1}{H} \delta_Z Y \, ds \mid Z \right] = \int_0^{t_0} \frac{H_Z}{H} \delta_Z ds = 1 - \frac{S_Z(t_0)}{S(t_0)} \tag{13}
\]
using (10) for the last step. \(\square\)

It is useful also to have an expression for \(Eh_{11}^2(X^*):\)

**Proposition 2**  The first order variance is given by
\[
Eh_{11}^2(X^*) = S^2(t_0)EA^2(Z) \left( - \left( 1 - \frac{S_Z(t_0)}{S(t_0)} \right)^2 \right) + \int_0^{t_0} \frac{H_Z}{H^2} \lambda_Z \, ds - 2 \int_0^{t_0} q_Z(s) \int_0^s \frac{\lambda}{H} \, du \, ds.
\]
Proof. By (13) the conditional expectation of the interesting part of $h_{11}^2 (X^*)$ (see (7)) is given by

$$E \left[ \left( 1 - \frac{S_Z (t_0)}{S (t_0)} - \int_0^{t_0} \frac{1}{H} \, dM \right)^2 | Z \right]$$

$$= - \left( 1 - \frac{S_Z (t_0)}{S (t_0)} \right)^2 + E \left[ \left( \int_0^{t_0} \frac{1}{H} \, dM \right)^2 | Z \right].$$

(14)

Here

$$E \left[ \left( \int_0^{t_0} \frac{1}{H} \, dM \right)^2 | Z \right] = E \left[ \left( \int_0^{t_0} \frac{1}{H} \, dM \right)^2 | Z \right] + 2E \left[ \int_0^{t_0} \frac{1}{H} \, dM \int_0^{t_0} \frac{1}{H} \, \delta Z \, ds | Z \right] + E \left[ \left( \int_0^{t_0} \frac{1}{H} \, \delta Z \, ds \right)^2 | Z \right].$$

The first term equals

$$E \left[ \int_0^{t_0} \frac{1}{H^2} \lambda Z \, ds | Z \right] = \int_0^{t_0} \frac{H_Z}{H^2} \lambda Z \, ds,$$

the second term equals, using (11)

$$-2 \int_0^{t_0} q_Z (s) \int_0^s \frac{\lambda Z}{H} \, du \, ds,$$

while the third equals

$$E \left[ \int_0^{t_0} \int_0^{t_0} \frac{\delta Z (s)}{H (s)} \frac{\delta Z (u)}{H (u)} H_Z (s \vee u) \, du \, ds | Z \right] = 2 \int_0^{t_0} q_Z (s) \int_0^s \frac{\delta Z (u)}{H (u)} \, du \, ds.$$

Thus

$$E \left[ \left( \int_0^{t_0} \frac{1}{H} \, dM \right)^2 | Z \right] = \int_0^{t_0} \frac{H_Z}{H^2} \lambda Z \, ds - 2 \int_0^{t_0} q_Z (s) \int_0^s \frac{\lambda}{H} \, du \, ds$$

and the proof is completed inserting this in (14).\qed

21
In conclusion we present the proofs of Theorems 1 and 2. The following analytic result is needed for the proof of Theorem 1.

**Lemma 1** Let \( f : [0, t] \rightarrow \mathbb{R} \) be Lebesgue integrable on \([0, t]\) and let \( c : [0, t] \rightarrow \mathbb{R} \) be differentiable and increasing with \( c(0) = 0 \). Then

\[
\int_0^t \int_0^t f(s)f(u)c(s \wedge u) \, du \, ds \geq 0
\]

with equality if and only if

\[
f \equiv 0 \quad c - a.e
\]

with \( c \) the positive measure on \([0, t]\) determined by the function \( c \).

**Proof.** Define the primitive \( F(t) = \int_0^t f(s) \, ds \). By symmetry

\[
\int_0^t \int_0^t f(s)f(u)c(s \wedge u) \, du \, ds = 2 \int_0^t c(s)f(s) \int_s^t f(u) \, du \, ds
\]

\[
= 2 \int_0^t c(s)f(s) (F(t) - F(s)) \, ds.
\]

Using partial integration

\[
\int_0^t c(s)f(s) \, ds = c(t)F(t) - \int_0^t c'(s)F(s) \, ds
\]

Also

\[
2 \int_0^t c(s)f(s)F(s) = c(t)F^2(t) - \int_0^t c'(s)F^2(s) \, ds
\]

so that

\[
\int_0^t \int_0^t f(s)f(u)c(s \wedge u) \, du \, ds = \int_0^t c'(s) \left(F^2(s) - 2F(t)F(s)\right) \, ds
\]

\[
= c(t)F^2(t) + \int_0^t c'(s) \left(F(s) - F(t)\right)^2 \, ds
\]

\[
- (c(t) - c(0))F^2(t)
\]

\[
= \int_0^t c'(s) \left(F(s) - F(t)\right)^2 \, ds.
\]

\(\square\)
Proof. (Theorem 1). Recall that $b$ is the hazard function for the censoring distribution. The theorem follows if we show that

$$E h_{12}^2 (X^*) = S^2 (t_0) \int_0^{t_0} \int_0^{t_0} \int_0^{s \wedge u} \frac{b(v)}{H(v)} \, dv \, d\mu(u) \, d\mu(s),$$

an expression of the form

$$S^2 (t_0) \int_0^{t_0} \int_0^{t_0} f_{\mu}(s) f_{\mu}(u) \int_0^{s \wedge u} \frac{b(v)}{H(v)} \, dv \, du \, ds$$

and Theorem 1 then follows directly using Lemma 1.

For the proof of (15) we use (8). Squaring that expression and ignoring the factor $S^2 (t_0)$ we find first that since $E Y(u) = H(u)$

$$E \left( \int_0^{t_0} \frac{1}{H(u)} \mu ([u, t_0]) \, dM(u) \right)^2 = E \int_0^{t_0} \frac{1}{H^2(u)} \mu^2 ([u, t_0]) \lambda(u) Y(u) \, du$$

$$= \int_0^{t_0} \int_0^{t_0} \int_0^{s \wedge u} \frac{\lambda(v) \mu(v)}{H(v)} \, dv \, d\mu(u) \, d\mu(s).$$

Next

$$E \left( \int_0^{t_0} \left( 1 - \frac{Y(s)}{H(s)} \right) \, d\mu(s) \right)^2$$

$$= \int_0^{t_0} \int_0^{t_0} E \left( 1 - \frac{Y(s)}{H(s)} \right) \left( 1 - \frac{Y(u)}{H(u)} \right) \, d\mu(u) \, d\mu(s)$$

$$= \int_0^{t_0} \int_0^{t_0} \left( -1 + \frac{H(s \lor u)}{H(s) H(u)} \right) \, d\mu(u) \, d\mu(s).$$

Finally, using the analogous of (11) with $M_Z$ replaced by $M$ and $\frac{1}{H}$ replaced by $\frac{\mu([v, t_0])}{H}$

$$-2 E \int_0^{t_0} \frac{1}{H(u)} \mu ([u, t_0]) \, dM(u) \int_0^{t_0} \left( 1 - \frac{Y(s)}{H(s)} \right) \, d\mu(s)$$

$$= -2 \int_0^{t_0} \int_0^{s \lor u} \frac{\lambda(v)}{H(v)} \mu ([v, t_0]) \, dv \, d\mu(s)$$

$$= -2 \int_0^{t_0} \int_0^{s \lor u} \frac{\lambda}{H} \, dv \, d\mu(u) \, d\mu(s)$$

Adding the contributions from (16), (17) and (18) the total factor on $d\mu(u) \, d\mu(s)$ becomes

$$\int_0^{s \lor u} \frac{\lambda}{H} \, dv + \left( -1 + \frac{H(s \lor u)}{H(s) H(u)} \right) - 2 \int_0^{s \lor u} \frac{\lambda}{H} \, dv$$
and evaluating this for, eg, $s \leq u$ yields
\[
-1 + \frac{1}{H(s)} \int_0^s \frac{\lambda}{H} \, dv = -1 + \frac{1}{H(s)} - \int_0^s \frac{\lambda + b}{H} \, dv + \int_0^s \frac{b}{H} \, dv = \int_0^s \frac{b}{H} \, dv
\]
as desired for (15). \hfill \Box

**Proof.** (Theorem 2). We shall show that
\[
Eh_{11}(X^*) h_{12}(X^*) = -Eh_{12}^2(X^*)
\] (19)
with the latter on the form (15). A first helpful observation is that
\[
E[1 - \frac{Y(s)}{H(s)} | Z] = 1 - \frac{H_Z(s)}{H(s)} = E \left[ \int_0^s 1 \, dM | Z \right]
\] (20)
as is seen because
\[
E \left[ 1 - \frac{Y(s)}{H(s)} | Z \right] = 1 - \frac{H_Z(s)}{H(s)} = E \left[ \int_0^s 1 \, dM | Z \right]
\] by (13). Now just refer to (9).

Because of (20)
\[
Eh_{11}(X^*) h_{12}(X^*) = -S^2(t_0) E(A(Z)) \int_0^{t_0} \frac{1}{H} \, dM \left( \int_0^{t_0} \frac{1}{H} \, dM \, d\mu(s) - \int_0^{t_0} \left( 1 - \frac{Y(s)}{H(s)} \right) \, d\mu(s) \right).
\]
We find the expectation on the right by conditioning on $Z$ and showing that
\[
E \left[ \int_0^{t_0} \frac{1}{H} \, dM \left( \int_0^s \frac{1}{H} \, dM - 1 + \frac{Y(s)}{H(s)} \right) | Z \right] = \int_0^{t_0} q_Z(u) \int_0^{s \wedge u} \frac{b}{H} \, dv \, du
\] (21)
(19) then follows by multiplying by $-S^2(t_0) E(A(Z)) \, d\mu(s)$, integrating on $s$ from 0 to $t_0$ and taking the expectation.
The integrand on the left of (21) is written as
\[
\left( \int_0^{t_0} \frac{1}{H} \, dM_Z + \int_0^{t_0} \delta_Z \frac{Y}{H} \, du \right) \left( \int_0^{s} \frac{1}{H} \, dM_Z + \int_0^{s} \delta_Z \frac{Y}{H} \, du - 1 + \frac{Y(s)}{H(s)} \right)
\]
\[(22)\]
where we first proceed to argue that
\[
E \left[ \int_0^{t_0} \frac{1}{H} \, dM_Z \left( \int_0^{s} \frac{1}{H} \, dM_Z + \int_0^{s} \delta_Z \frac{Y}{H} \, du - 1 + \frac{Y(s)}{H(s)} \right) \right] = 0.
\]
Indeed
\[
E \left[ \int_0^{t_0} \frac{1}{H} \, dM_Z \right] = \int_0^{t_0} \frac{H_Z}{H^2} \lambda_Z \, du
\]
and by (11)
\[
E \left[ \int_0^{t_0} \frac{1}{H} \, dM_Z \left( \int_0^{s} \frac{1}{H} \, dM_Z + \int_0^{s} \delta_Z \frac{Y}{H} \, du - 1 + \frac{Y(s)}{H(s)} \right) \right] = - \int_0^{s} q_Z(u) \int_0^{u} \frac{\lambda_Z}{H} \, dv \, du - \int_0^{s} \frac{H_Z(s)}{H(s)} \int_0^{s} \frac{\lambda_Z}{H} \, du = - \int_0^{s} \frac{H_Z}{H^2} \lambda_Z \, du.
\]
by partial integration.

We are left with finding two terms from (22), the first of which is
\[
E \left[ \int_0^{t_0} \delta_Z \frac{Y}{H} \, du \right] = - \int_0^{t_0} q_Z(u) \int_0^{s \wedge u} \frac{\lambda_Z}{H} \, dv \, du
\]
again using (11).

The second term required from (22) is
\[
E \left[ \int_0^{t_0} \delta_Z \frac{Y}{H} \, du \left( \int_0^{s} \delta_Z \frac{Y}{H} \, dv - 1 + \frac{Y(s)}{H(s)} \right) \right] = \int_0^{t_0} \int_0^{s} \delta_Z(u) \delta_Z(v) \frac{H_Z(u \vee v)}{H(u) H(v)} \, dv \, du - \int_0^{t_0} \delta_Z \frac{H_Z}{H} \, du + \frac{1}{H(s)} \int_0^{t_0} \delta_Z(v) \frac{H(s \vee v)}{H(v)} \, dv.
\]
Establishing (21) is now reduced to showing that for $0 \leq s \leq t_0$
\[
\int_0^{t_0} q_Z(u) \int_0^{s \wedge u} \frac{b}{H} \, dv \, du = - \int_0^{t_0} q_Z(u) \int_0^{s \wedge u} \frac{\lambda_Z}{H} \, dv \, du
\]
\[(23)\]
\[+ \int_0^{t_0} \int_0^{s} \delta_Z(u) \delta_Z(v) \frac{H_Z(u \vee v)}{H(u) H(v)} \, dv \, du
\]
\[+ \int_0^{t_0} \delta_Z \frac{H_Z}{H} \, du + \frac{1}{H(s)} \int_0^{t_0} \delta_Z(v) \frac{H(s \vee v)}{H(v)} \, dv\]
\[+ \int_0^{t_0} q_Z(u) \int_0^{s \wedge u} \frac{b}{H} \, dv \, du.
\]
and we shall achieve this by showing that the derivatives with respect to \( t_0 \) on both sides of the equality sign match and that the two sides are identical for \( t_0 = s \).

Clearly the \( t_0 \)-derivative on the left is 

\[
q_Z(t_0) \left( - \int_0^s \frac{\lambda_Z}{H} dv + \int_0^s \frac{\delta_Z}{H} dv - 1 + \frac{1}{H(s)} \right) = q_Z(t_0) \int_0^s \frac{b}{H} dv
\]

because

\[
\frac{1}{H(s)} - 1 = \int_0^s \frac{\lambda + b}{H} dv.
\]

Taking \( t_0 = s \) in (23) we finally need

\[
\int_0^s q_Z(u) \int_0^u \frac{b}{H} dv du = - \int_0^s q_Z(u) \int_0^u \frac{\lambda Z}{H} dv du + 2 \int_0^s \int_0^u \frac{\delta_Z(u) \delta_Z(v)}{H(u) H(v)} dv du - \int_0^s q_Z du + \frac{H_Z(s)}{H(s)} \int_0^s \frac{\delta_Z}{H} dv.
\]

Both sides = 0 for \( s = 0 \). The \( s \)-derivative on the left is 

\[
q_Z(s) \left( - \int_0^s \frac{\lambda_Z}{H} dv + 2 \int_0^s \frac{\delta_Z}{H} dv - 1 - \int_0^s \frac{\delta_Z}{H} dv + \frac{1}{H(s)} \right) = q_Z(s) \int_0^s \frac{b}{H} dv.
\]

\[
\square
\]

References


Research Reports available from Department of Biostatistics

http://www.pubhealth.ku.dk/bs/publikationer

11/1 Kreiner, S. Is the foundation under PISA solid? A critical look at the scaling model underlying international comparisons of student attainment.

11/2 Andersen, P.K. Competing risks in epidemiology: Possibilities and pitfalls.

11/3 Holst, K.K. Model diagnostics based on cumulative residuals: The R-package gof.


11/5 Holst, K.K., Budtz-Jørgensen, E. & Knudsen, G.M. A latent variable model with mixed binary and continuous response variables.


12/01 Keiding, N., Hansen, O.K.H., Sørensen, D.N. & Slama, R. The current duration approach to estimating time to pregnancy.

12/02 Andersen, P.K. A note on the decomposition of number of life years lost according to causes of death.
12/03 Martinussen, T. & Vansteelandt, S. A note on collapsibility and confounding bias in Cox and Aalen regression models.


12/05 Martinussen, T. & Pipper, C.B. Estimation of Causal Odds of Concordance based on the Aalen additive model.

12/06 Binder, N., Gerds, T.A. & Andersen, P.K. Pseudo-observations for competing risks with covariate dependent censoring.

12/07 Keiding, N. & Clayton, D. Standardization and control for confounding in observational studies: a historical perspective.

12/08 Hilden, J. & Gerds, T.A. Evaluating the impact of novel biomarkers: Do not rely on IDI and NRI.


13/01 Forman, J.L. & Sørensen M. A new approach to multi-modal diffusions with applications to protein folding.


13/03 Wolbers, M., Koller, M.T., Witteman, J.C.M. & Gerds, T.A. Concordance for prognostic models with competing risks.

13/04 Andersen, P.K., Canudas-Romo, V. & Keiding, N. Cause-specific measures of life lost.

13/05 Keiding, N. Event history analysis.


13/07 Hansen, S., Andersen, P.K. & Parner, E. Events per variable for risk differences and relative risks using pseudo-observations.


13/11 Olsbjerg, M. & Christensen K.B. SAS macro for marginal maximum likelihood estimation in longitudinal polytomous Rasch models.


14/01 Ambrogi, F. & Andersen, P.K. Predicting Smooth Survival Curves through Pseudo-Values.

14/02 Olsbjerg, M. & Christensen, K.B. Modeling local dependence in longitudinal IRT models.

14/03 Olsbjerg, M. & Christensen, K.B. LIRT: SAS macros for longitudinal IRT models.

14/04 Jacobsen, M. & Martinussen, T. A note on the large sample properties of estimators based on generalized linear models for correlated pseudo-observations.