Estimation of the Mean Function with Panel Count Data Using Monotone Polynomial Splines

By MINGGEN LU, YING ZHANG

Department of Biostatistics, The University of Iowa,

200 Hawkins Drive, C22 GH

Iowa City, IA 52242, U.S.A.

minggen-lu@uiowa.edu ying-j-zhang@uiowa.edu

AND JIAN HUANG

Department of Statistics and Actuarial Science, The University of Iowa,

221 Schaeffer Hall, Iowa City, IA 52242, U.S.A.

jian-huang@uiowa.edu

Summary

We study the nonparametric likelihood-based estimators of the mean function of counting processes with panel count data using monotone polynomial splines. The generalised Rosen algorithm, proposed by Zhang & Jamshidian (2004), was used to compute the estimators. We show that the proposed spline likelihood-based estimators are consistent and their rate of convergence can be faster than $n^{1/3}$. Simulation studies with moderate samples show that the estimators have smaller variances and mean square errors than their counterparts proposed in Wellner & Zhang (2000). A real example from bladder tumor clinical trial is used to illustrate the method.

Some key words: Counting process; Empirical process; Maximum likelihood estimator; Maximum pseudolikelihood estimator; Monotone polynomial splines; Monte Carlo; Isotonic regression.

1 Introduction

This article considers estimation of the mean function of counting processes with panel count data using monotone polynomial splines. In many long-term clinical trials or epidemiological studies, the subjects were observed at several discrete times during the study. The only available information was the number of recurrent events occurring before each observation time. The exact event times were unknown. The number of observations and observation times may vary from individual to individual. This kind of data is referred to as panel count data.

Panel count data arises frequently in clinical studies, in which each patient visits the clinical center at some subsequent follow-up times after the initial visit. An example is the National Cooperative Gallstone Study, a 10-year multicenter double-blinded placebo-controlled trial investigating the effects of the natural bile acid chenodeoxycholic acid on the dissolution of cholesterol gallstones (Thall & Lachin, 1988). The individuals were asked to report the total number of symptoms such as nausea and diarrhea that had occurred during the intervals between successive clinic visits. One of the primary objectives of this study was to assess the impact of the treatment on the incidence of biliary (digestive) symptoms commonly associated with gallstone disease. Another example is the bladder tumour randomized clinical trial conducted by the Veterans Administration Cooperative Urological Research Group. All patients had superficial bladder tumours when they entered the trial, and they were randomly assigned to one of three arms: placebo, pyridoxine, and thiotepa. Many patients had multiple recurrences of the tumour, and new tumours were removed at each visit. The goal of this study was to determine the effect of treatment on the frequency of tumour recurrence; see for example, Byar (1980), Wei et al. (1989), Wellner & Zhang (2000), Sun & Wei (2000), and Zhang (2002).

Several authors have considered methods for analysing panel count data (Kalbfleisch & Lawless, 1985; Thall & Lachin 1988; Thall, 1988 and Lee & Kim, 1998). Sun & Kalbfleisch (1995) appears to be the first in the literature studying the nonparametric estimation of the mean function with panel count data. Their method was based on the isotonic regression by pulling the observations together and taking into account the monotonicity of the mean function. Wellner & Zhang (2000) studied nonparametric maximum pseudolikelihood and nonparametric maximum likelihood estimators based on a 'working model' of nonhomogeneous Poisson process. They showed that the maximum pseudolikelihood estimator was exactly the one described in Sun & Kalbfleisch (1995) and they also studied the asymptotic properties of both maximum pseudolikelihood and maximum likelihood estimators. The maximum likelihood estimator is more efficient than the maximum pseudolikelihood estimator, but the maximum likelihood estimator is more difficult to compute.

Many investigators have studied spline estimation of an unknown function such as hazard and survival functions in survival analysis. However, directly modelling the hazard function using the spline may not guarantee the nonnegativity of the function (Anderson & Senthilselvan, 1980 and Whittemore & Keller, 1986). To overcome this drawback, Rosenberg (1995) proposed to use the spline with coefficients expressed in exponential form; Kooperberg et al. (1995) and Cai & Betensky (2003) modelled the log hazard function using linear spline in their corresponding applications. Doing so, the nonnegativity of the hazard function is automatically satisfied.

In our application, the pseudolikelihood and likelihood functions with panel count data are functions of the cumulative mean function of event numbers. Although we can reparameterise the likelihood functions in terms of the intensity function and use the spline method, for example, proposed by Kooperberg et al. (1995), it will incur unnecessary complication in estimation procedure computationally, especially when the estimation of the mean function is the primary interest. In this article, the monotone cubic spline studied in Ramsay (1988) is applied to directly approximating the true mean function $\Lambda_0(t)$ of the counting process by

$$\Lambda(t) = \sum_{j=1}^{q_n} \alpha_j I_j(t) \quad \text{subject to} \quad \alpha_j \geq 0 \quad \text{for } j = 1, 2, \cdots, q_n$$

This is referred to as I-spline by Ramsay (1988) and the monotonicity of the Λ is guaranteed by imposing the nonnegative constraint on the coefficients of the linear combination.

We express both the pseudolikelihood and likelihood functions given in Wellner & Zhang

(2000) using the I-spline functions and estimate the spline coefficients using the generalised Rosen algorithm proposed by Zhang & Jamshidian (2004). Our approach has two attractive features: (i) it is much less demanding in computing the spline likelihood estimator compared to computing the nonparametric maximum likelihood estimator in Wellner & Zhang (2000) based on the iterative convex minorant algorithm proposed by Jongbloed (1988) and (ii) the spline estimators may have a higher convergence rate than their counterparts proposed by Wellner & Zhang (2000).

The rest of the paper is outlined as follows: In section 2, both the spline maximum pseudolikelihood and likelihood estimators, $\hat{\Lambda}^{ps}$ and $\hat{\Lambda}$, are characterised and the generalised Rosen's algorithm is introduced to compute the spline estimators. In section 3, the main asymptotic results, consistency and rate of convergence, are stated. In section 4, two sets of simulation are carried out to study the properties of the spline-based estimators. The methods are illustrated with the bladder tumour example. In section 5, future extension of the monotone spline estimator is discussed. Finally, proofs of the asymptotic consistency and rate of convergence of the spline estimators are sketched in the Appendix.

2 Monotone Spline Estimators of the Mean Function

Let $\{\mathbb{N}(t) : t \geq 0\}$ be a counting process with mean function $E\mathbb{N}(t) = \Lambda_0(t)$. The total number of observation K on the counting process is an integer-valued random variable and $T = (T_{K,1}, T_{K,2}, \cdots, T_{K,K})$ is a sequence of random observation times with $0 < T_{K,1} < T_{K,2} < \ldots < T_{K,K}$. The cumulative numbers of recurrent events up to these times $\mathbb{N} = \{\mathbb{N}(T_{K,1}), \mathbb{N}(T_{K,2}), \ldots, \mathbb{N}(T_{K,K})\}$ with $0 \leq \mathbb{N}(T_{K,1}) \leq \mathbb{N}(T_{K,2}) \leq \cdots \leq \mathbb{N}(T_{K,K})$ are observed accordingly. The panel count data of the counting process consist of $\mathrm{X} = (K, T, \mathbb{N})$. We assume that the number of observations and the sequence of observation times are distributed independently of the underlying process. That is, $\mathbb{N}$ and (K, T) are independent.

Suppose we observe n independently and identically distributed copies of X, $X_i = (K_i, T_i, \mathbb{N}^{(i)})$ with $T_i = (T_{K_i,1}, T_{K_i,2}, \ldots, T_{K_i,K_i})$ and

$\mathbb{N}^{(i)} = \left\{\mathbb{N}^{(i)}(T_{K_i,1}), \mathbb{N}^{(i)}(T_{K_i,2}), \ldots, \mathbb{N}^{(i)}(T_{K_i,K_i})\right\}$ for $i = 1, 2, \ldots, n$. We denote the observed data by $D = (X_1, X_2, \ldots, X_n)$. Wellner & Zhang (2000) proposed two nonparametric estimation methods for the mean function of the counting process. Assuming the underline counting process being nonhomogeneous Poisson process and ignoring the correlations among the cumulative counts, they established the log pseudolikelihood by omitting the parts irrelevant to the mean function Λ,

$$l_n^{ps}(\Lambda|D) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \left\{\mathbb{N}^{(i)}(T_{K_i,j}) \log \Lambda(T_{K_i,j}) - \Lambda(T_{K_i,j})\right\}. \tag{1}$$

In parallel, they also established the log likelihood for Λ using the independence of the increments of $\mathbb{N}(t)$,

$$l_n(\Lambda|D) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \{\Delta N_{K_i,j} \log(\Delta\Lambda_j)\} - \sum_{i=1}^{n} \Lambda(T_{K_i,K_i}) \tag{2}$$

where $\Delta N_{K_i,j} = \mathbb{N}^{(i)}(T_{K_i,j}) - \mathbb{N}^{(i)}(T_{K_i,j-1})$ and $\Delta\Lambda_j = \Lambda(T_{K_i,j}) - \Lambda(T_{K_i,j-1})$ for $j = 1, 2, \ldots, K_i$; $i = 1, 2, \ldots, n$.

As described in Wellner & Zhang (2000), the nonparametric maximum pseudolikelihood estimator can be computed in one step via the max-min formula, and the computation of the maximum likelihood estimator involves the iterative convex minorant algorithm which can be computationally demanding when the sample size is large.

Polynomial regression splines defined on the interval $[L, U]$ is a piecewise polynomial with specified continuity constraints. $[L, U]$ is subdivided by a mesh Δ consisting of points $L = t_0 < t_1 < t_2 < \cdots < t_{q_n} < t_{q_n+1} = U$. The spline is a polynomial of degree $l - 1$ within the subinterval $[t_j, t_{j+1})$, and the polynomials in the adjacent intervals have the same derivatives up to $l - 2$ at the joint points. The spline for $l = 4$ is piecewise cubic polynomial with second order continuous derivative. As a special case, the spline with $l = 1$ is a step function discontinuous at each knot. Let $t = (t_1, t_2, \cdots, t_{q_n})$ be a sequence of knots. The M-spline is a family of basis spline, M_i, which is positive in (t_i, t_{i+1}), zero elsewhere, and is normalised to $\int M_i(x)dx = 1$. For $t_i \le x \le t_{i+1}$, $M_i(x)$ is defined by the recursion

$$\begin{aligned} M_i(x|1,t) &= \frac{1}{t_{i+1}-t_i}, \quad \text{and 0 otherwise} \\ M_i(x|l,t) &= \frac{l\{(x-t_i)M_i(x|l-1,t) + (t_{i+l}-x)M_{i+1}(x|l-1,t)\}}{(l-1)(t_{i+l}-t_i)} \end{aligned}$$

Further discussion of polynomial splines and M-splines can be found in Schumaker (1981) and Ramsay (1988). In this article, we consider I-splines constructed by Ramsay (1988): $I_i(x|l,t) = \int_L^x M_i(u|l,t)du$. A linear combination of I_i's with nonnegative coefficients yields a monotone nondecreasing function (Ramsay 1988, pp. 428).

Let $q_n = n^\nu$ with $0 < \nu < 1/2$ being a positive integer such that $\max_{1\le i\le q_n+1} |t_i - t_{i-1}| = O(n^{-\nu})$. Stone (1986) showed that when n is sufficiently large, any smooth function can be approximated by a linear combination of polynomial spline functions with number of knots chosen in this fashion. Therefore, we approximate the smooth monotone mean function $\Lambda_0(t)$ by $\sum_{j=1}^{q_n} \alpha_j I_j(t)$ and estimate the coefficients $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_{q_n})$ through maximising the approximated pseudolikelihood and likelihood functions, respectively.

Let $\hat{\alpha}_p^{ps}$ for $p = 1, 2, \cdots, q_n$ denote the spline pseudolikelihood estimators that maximise

$$l_n^{ps}(\alpha|D) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}\left[\mathbb{N}^{(i)}(T_{K_i,j}) \log\left\{\sum_{p=1}^{q_n} \alpha_p I_p(T_{K_i,j})\right\} - \sum_{p=1}^{q_n} \alpha_p I_p(T_{K_i,j})\right] \tag{3}$$

subject to the constraints $\alpha_p \ge 0$ for $p = 1, 2, \cdots, q_n$. Similarly, let $\hat{\alpha}_p$ for $p = 1, 2, \cdots, q_n$ denote the spline likelihood estimators that maximise

$$l_n(\alpha|D) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}\left[\Delta N_{K_i,j} \log\left\{\sum_{p=1}^{q_n} \alpha_p \Delta I_p(T_{K_i,j})\right\}\right] - \sum_{i=1}^{n}\sum_{p=1}^{q_n} \alpha_p I_p(T_{K_i,K_i}) \tag{4}$$

with $\Delta I_p = I_p(T_{K_i,j}) - I_p(T_{K_i,j-1})$ for $p = 1, 2, \ldots, q_n$, subject to the same constraints as above.

We use the cubic spline to approximate the mean function $\Lambda_0(t)$. In fact, the two estimators proposed by Wellner & Zhang (2000) can be viewed as the special spline of $l = 1$ that uses all the distinct observation times as knots. The estimators constructed here can, therefore, be treated as the extension of those proposed by Wellner & Zhang (2000) with respect to the smoothness of estimators. However, the number of coefficients to be estimated is reduced. As a result, the spline estimators are expected to be less computationally demanding.

We note that both the spline pseudolikelihood and spline likelihood functions are concave with respect to the unknown coefficients. So the spline likelihood estimation problem is equivalent to a nonlinear convex programming problem subject to linear inequality constraints. Specifically, the spline estimation problems (3) and (4) can be formulated as the linearly inequality constrained maximisation problem

$$\max_{\alpha \in \Theta_\alpha} l(\alpha|X) \tag{5}$$

where $\Theta_\alpha = \{\alpha : \alpha_p \geq 0,\ p = 1, 2, \cdots, q_n\}$. Rosen (1960) proposed a now well-known iteratively generalised gradient method for optimising an objective function with linear constraints. Rosen's algorithm was formulated based on the Euclidean metric in the optimisation literature. Jamshidian (2004) developed a general algorithm based on the generalised Euclidean metric $\|x\| = x^T W x$ where W is a positive definite matrix and can vary from iteration to iteration. Zhang & Jamshidian (2004) applied this generalised Rosen algorithm to computing the nonparametric maximum likelihood estimator of failure function with various types of censored data. Zhang & Jamshidian (2004) used $W = -D_H$, the diagonal elements of the negative Hessian matrix H, to avoid the possible storage problem in updating H for a large-scale nonparametric maximum likelihood estimation problem.

In this article, we directly use the negative Hessian matrix H because the dimension of unknown space is usually not large in our methods. The use of the full Hessian matrix substantially reduces the number of iterations and thus save the computing time. The detailed description of the computation method and algorithm coded in R can be obtained from the first author.

3 Asymptotic Results

We study the asymptotic properties of both the spline pseudolikelihood and spline likelihood estimators with the same L_2 metric d defined in Wellner & Zhang (2000); that is

$$d(\Lambda_1, \Lambda_2) = \|\Lambda_1 - \Lambda_2\|_2 = \left\{ \int |\Lambda_1(t) - \Lambda_2(t)|^2 d\mu(t) \right\}^{1/2}$$

where

$$\mu(t) = \sum_{k=1}^{\infty} \mathrm{pr}(K = k) \sum_{j=1}^{k} \mathrm{pr}(T_{K,j} \leq t | K = k)$$

for any $\Lambda_1, \Lambda_2 \in \mathcal{F}$ with $\mathcal{F} = \{\Lambda : \Lambda \text{ is nondecresing. } \Lambda(0) = 0\}$

We show the consistency and rate of convergence of both the spline pseudolikelihood and spline likelihood estimators, based on the following regularity conditions on the true mean function and the underlying distribution of observation times.

Condition 1. For some interval $O[T] = [\sigma, \tau]$ with $\sigma > 0$ and $\Lambda_0(\sigma) > 0$, $\mathrm{pr}(\cap_{j=1}^{K}\{T_{K,j} \in [\sigma, \tau]\}) = 1$

Condition 2. There exists a positive integer M_0 such that $\mathrm{pr}(K \leq M_0) = 1$. That is, the number of observations is finite.

Condition 3. The true mean function Λ_0 is $(r-1)$th bounded differentiable in $O[T]$ with $r \geq 2$. Moreover, the first derivative has a positive lower bound in $O[T]$. That is, there exists a constant $C_0 > 0$ such that $\Lambda_0'(t) \geq C_0$ for $t \in O[T]$.

Condition 4. $E\left\{e^{C\mathbb{N}(t)}\right\}$ is uniformly bounded for $t \in S[T] = \{t : 0 < t < \tau\}$ for some $\tau > 0$. The τ can be viewed as the termination time in a clinic follow-up study.

Condition 5. The observation times points are γ-separated. That is, there exists a $\gamma > 0$ such that $\mathrm{pr}(T_{k,j} - T_{k,j-1} \geq \gamma) = 1$ for all $j = 1, 2, \cdots, K$.

The conditions related to the observation schemes (Conditions 1, 2 and 5) are mild and easily justified in view of applications in clinical trials. Condition 4 is true if the underlying counting process is uniformly bounded or if it is a Poisson/Mixed Poisson process. The smoothness assumption of true mean function Λ_0 (Condition 3) is standard in the nonparametric smoothing literature.

We denote the spline pseudolikelihood estimator by $\hat{\Lambda}_n^{ps} = \sum_{j=1}^{q_n} \hat{\alpha}_j^{ps} I_j(t)$ and the spline likelihood estimator by $\hat{\Lambda}_n = \sum_{j=1}^{q_n} \hat{\alpha}_j I_j(t)$, respectively.

Theorem 1 (*Consistency*) *Suppose Conditions 1-4 hold, then*

$$d(\hat{\Lambda}_n^{ps}, \Lambda_0) \to 0 \qquad \text{as } n \to \infty.$$

in probability. In addition, if Condition 5 holds, then

$$d(\hat{\Lambda}_n, \Lambda_0) \to 0 \qquad \text{as } n \to \infty$$

in probability.

Theorem 2 (*Rate of convergence*) *Suppose Conditions 1-4 hold, then*

$$n^{r/(1+2r)} d(\hat{\Lambda}_n^{ps}, \Lambda_0) = O_p(1)$$

In addition, if Condition 5 holds, then

$$n^{r/(1+2r)} d(\hat{\Lambda}_n, \Lambda_0) = O_p(1)$$

The proofs of these theorems are sketched in appendix. Theorem 2 shows that the spline estimators can have a higher convergence rate than their counterparts studied in Wellner & Zhang (2000), because $r/(1+2r) \geq 1/3$ when $r \geq 2$.

4 Numerical Results

4.1 Simulation Studies

We conduct simulation studies to compare the statistical properties and computation complexity among the spline pseudolikelihood/likelihood estimators and their counterparts studied in Wellner & Zhang (2000). Two Monte Carlo simulation studies designed in Zhang &

Jamshidian (2004) are carried out here. For each simulation study, we generate n independently and identically distributed observations $X_i = (K_i, T_i, \mathbb{N}^{(i)})$ for $i = 1, 2, \ldots, n$. For each i, K_i samples randomly from the discrete uniform distribution $\{1, 2, 3, 4, 5, 6\}$. Given K_i, the random panel observation times $T_i = (T_{K_i,1}, T_{K_i,2}, \ldots, T_{K_i,K_i})$ are K_i ordered random draws from Un(0,10) and rounded to the second decimal point. The two simulations differ in the method of generating the panel counts $\mathbb{N}^{(i)} = \{\mathbb{N}^{(i)}(T_{K_i,1}), \mathbb{N}^{(i)}(T_{K_i,2}), \ldots, \mathbb{N}^{(i)}(T_{K_i,K_i})\}$, given (K_i, T_i). They are described as follows:

Simulation 1. The panel counts are generated from Po$(2t)$. That is,

$$\mathbb{N}^{(i)}(T_{K_i,j}) - \mathbb{N}^{(i)}(T_{K_i,j-1}) \sim \text{Po}\{2(T_{K_i,j} - T_{K_i,j-1})\} \text{ for } j = 1, 2, \ldots, K_i.$$

Simulation 2. The panel counts are generated from a mixed Poisson process. We first generate a random sample $\alpha_1, \alpha_2, \ldots, \alpha_n \sim \{-0 \cdot 4, 0, 0 \cdot 4, 0\}$ with pr$(\alpha_i = -0 \cdot 4)$=pr$(\alpha_i = 0 \cdot 4)$=1/4 and pr$(\alpha_i = 0)$=1/2 for $i = 1, 2, \ldots, n$. Given α_i, the panel counts for the ith subject are generated according to Po$\{(2 + \alpha_i)t\}$. That is,

$$\mathbb{N}^{(i)}(T_{K_i,j}) - \mathbb{N}^{(i)}(T_{K_i,j-1})|\alpha_i \sim \text{Po}\{(2 + \alpha_i)(T_{K_i,j} - T_{K_i,j-1})\} \text{ for } j = 1, 2, \ldots, K_i.$$

This counting process is not a Poisson process unconditionally since the mean function of the process, $E\{\mathbb{N}^{(i)}(t)\} = 2t$, is not equal to the variance function of the process, var$\{\mathbb{N}^{(i)}(t)\} = 2t + 0.08t^2$.

Cubic I-splines are used in the simulations. The number of knots q_n is selected as cubic root of the number of distinct observation times plus 1. We consider two methods for selecting the knots. Let $T_{\min}$ and $T_{\max}$ be the two endpoints of the collection of total observation time points in the data. Method 1 equally divides the interval $[T_{\min}, T_{\max}]$ into $q_n + 1$ subintervals, and the endpoints of the subintervals are chosen to be the knots. Method 2 is a data-driven knot selection method: given the number of knots q_n, the knots $t_j, j = 1, 2, \cdots, q_n$, are chosen to be the $j/(q_n + 1) \times 100\%$ - quantile of the collection of total observation time points in the data.

In our studies, we generated 1000 Monte Carlo samples with n=100 and 200 for each case, respectively. We found that the two knots selection methods yielded very similar results. Therefore, we only report the results based on the quantile knots selection method.

Fig.1 plots the four estimators with n=100 along with the true mean function $\Lambda_0(t) = 2t$ in *Simulation 1*. It is clear that, while all these estimators converge to the true mean function, the spline estimators appear closer to the true mean function than their counterparts. To compare these estimators in detail, we calculate the estimates of the mean function at time points $t = 1 \cdot 5, 2 \cdot 0, 2 \cdot 5, \ldots, 9 \cdot 5$ and summarise the results in Tables 1 and 2 for n=100 and 200, respectively. From the tables, we can see that the biases of all these estimators are negligible compared to the estimated values. The pointwise standard deviations and mean square errors of the spline estimators are remarkably smaller than their counterparts. While the spline likelihood estimator appears to be the most efficient estimator among the four, the spline pseudolikelihood estimator performs almost as well as the nonparametric maximum likelihood estimator. When sample size doubles, both the pointwise standard deviations and mean square errors drop substantially which supports the asymptotic consistency of these estimators.

In *Simulation 2*, we similarly plot the four estimators along with the true mean function in Fig.2 for n=100 and summarise the Monte Carlo simulation results in Tables 3 and 4 for n=100 and 200, respectively. The figure and tables reveal the same pattern of the four estimators as in *Simulation 1*. These simulation studies also reinforce the conclusion made in Wellner & Zhang (2000) that the likelihood method based on Poisson process is robust against the underlying counting process. However, the standard deviations and mean square errors of theses estimators are elevated when the Poisson process model is misspecified from the true underlying counting process.

We also compare the computing time among the four estimators and summarise the results in Table 5. The nonparametric maximum pseudolikelihood estimator appears to be the lest computationally demanding estimator. However it is lest efficient as shown in Tables 1-4. The spline likelihood estimator, while demonstrating its estimation efficiency through the simulation studies, also shows the computation efficiency. Computing time for the spline likelihood estimator is less than 1/12 of that for its counterpart on average and the saving of computing time for the spline likelihood estimator over its counterpart becomes more significant when sample size increases, as shown in Table 5.

4.2 A Real Example: Bladder Tumour Trial

The proposed methods are illustrated using the bladder tumour example described in the introduction. The data were extracted from the book by Andrew & Herzberg (1985). In this randomised clinical trial, a total of 116 patients were randomly assigned into one of three treatment groups, placebo (47), pyridoxine (31) and thiotepa (38). The number of follow-ups and follow-up times varied greatly from patient to patient which gave arise a perfect example of panel count data described in this article. The data set has been extensively studied in the literatures; see for example, Byar et al. (1977), Byar (1980), Sun & Wei (2000), Zhang (2002).

In these analyses, researchers were interested in the efficacy of two treatments: pyridoxine pill and thiotepa installation in terms of suppressing the recurrence of bladder tumour. We compute the spline pseudolikelihood and likelihood estimators of the cumulative mean function for three treatment groups. The spline estimators along with nonparametric estimators proposed in Wellner & Zhang (2000) are plotted in Figs. 3 and 4. As shown in the figures, the difference of the mean function between thiotepa group and placebo group is quite substantial. Having observed the difference between the treatments, we also noticed the big discrepancy between the pseudolikelihood estimators and the likelihood estimators. This may be due to the fact that the sample for each treatment is relatively small, in particular, the observations at later times are scarce.

5 Discussion

Our studies show that the spline likelihood estimators outperform the conventional nonparametric likelihood estimators in terms of finite-sample statistical efficiency. They are also easier to compute. These advantages motivate the use of monotone spline estimator in applications involving the estimation of a monotone function while the conventional nonparametric likelihood estimator of the function is difficult to compute.

In semiparametric regression problems, joint estimation of both nonparametric nuisance parameter and parametric regression parameter is often a challenging task. For example, in an unpublished technical report of Department of Statistics at University of Washington, Wellner and Zhang considered estimation in the semiparametric proportional mean model with panel count data, namely

$$E\{\mathbb{N}(t)|Z\} = \exp(\beta_0' Z)\Lambda_0(t), \tag{6}$$

where Z is a vector of covariates and Λ_0 is the baseline mean function. Although the asymptotic properties of semiparametric maximum pseudolikelihood and maximum likelihood estimators were studied and the normality of the estimator of β_0 was established, the asymptotic variances of these estimators are difficult to estimate directly. The bootstrap inference procedure was implemented in their technical report with an enormous effort in computation. Therefore, it is worth making further effort toward more computationally efficient estimators. Spline estimators appear to be a possible candidate due to its advantages aforementioned. The semiparametric model (6) can be reformulated through approximating the $\log \Lambda_0(t)$ by a linear combination of normalized B-splines introduced by Schumaker (1981),

$$E\{\mathbb{N}(t)|Z\} = \Lambda_0(t)\exp(\beta_0' Z) = \exp\left\{\sum_1^{q_n} \alpha_i B_i(t) + \beta_0' Z\right\}$$

subject to the constraint $\alpha_1 < \alpha_2 < \cdots < \alpha_{q_n}$. By Theorem 5·9 of Schumaker (1981), the monotonicity of $\log \Lambda_0(t)$ is guaranteed by the monotonicity of the coefficients α's. The joint estimation of α's and β's can be implemented in the same way as described in this article and is expected to be a well-manageable task computationally.

Appendix

The proofs for the two asymptotic results are sketched in this section. The modern empirical process theory is the major technical tool. The notations used in this section follow those given in van der Vaart & Wellner (1996) and Huang (1999). Here we only sketch the proofs for the spline pseudolikelihood estimator, since the proofs for the spline likelihood estimator are almost parallel with Condition 5 included.

Proof of Theorem 1. A pseudolikelihood function for Λ can be written as

$$\mathbb{M}_n(\Lambda) = l_n^{ps}(\Lambda|D) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}\left\{\mathbb{N}^{(i)}(T_{K_i,j})\log\Lambda(T_{K_i,j}) - \Lambda(T_{K_i,j})\right\}$$

Let Φ denote the collection of functions h on $O[T]$ whose $(r-1)$th derivative $h^{(r-1)}$ for $r \geq 1$ exists and satisfies the Lipschitz conditions

$$|h^{(r-1)}(s) - h^{(r-1)}(t)| \leq C|s-t| \qquad \text{for } s,t \in O[T]$$

By Condition 3, the true mean function $\Lambda_0 \in \Phi$. Let $\xi = t_0 < t_1 < \cdots < t_{q_n} < t_{q_n+1} = \tau$ be a partition of $O[T] = [\xi, \tau]$ into $q_n + 1$ subintervals where $q_n = n^\nu$ with $0 < \nu < 1/2$ and φ_n be the set of polynomial spline functions on $O[T]$. Condition 3 guarantees the assumptions of Lemma A·5 of Huang (1999). Therefore for $\Lambda_0 \in \Phi$, there exists a $\Lambda_n \in \varphi_n$ such that $\|\Lambda_n - \Lambda_0\|_\infty = O(n^{-\nu r} + n^{-(1-\nu)/2})$. So, for $\eta > 0$, $\sup_{t\in O[T]} |\Lambda_n(t) - \Lambda_0(t)| \leq \eta$ for large n. Therefore, we can find $\Lambda_n \in \varphi_n$ such that $\Lambda_n > \Lambda_0$ and $\|\Lambda_n - \Lambda_0\|_\infty = O(n^{-\nu r} + n^{-(1-\nu)/2})$. Choose a positive $h_n \in \varphi_n$ such that $h_n \geq c(\Lambda_n - \Lambda_0)$ and $\|h_n\|_2^2 = O(n^{-\nu r} + n^{-(1-\nu)/2})$. Therefore, for any $\alpha > 0$, $\|\Lambda_n - \Lambda_0 + \alpha h_n\|_2^2 = O(n^{-\nu r} + n^{-(1-\nu)/2})$ and $\inf(\Lambda_n - \Lambda_0 + \alpha h_n) > 0$ for sufficiently large n.

Let $H_n(\alpha) = \mathbb{M}_n(\Lambda_n + \alpha h_n)$. The first and second derivatives of H_n are

$$H_n'(\alpha) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}\left\{\frac{\mathbb{N}^{(i)}(T_{K_i,j})}{\Lambda_n(T_{K_i,j}) + \alpha h_n(T_{K_i,j})} - 1\right\} h_n(T_{K_i,j})$$

and

$$H_n''(\alpha) = -\sum_{i=1}^{n}\sum_{j=1}^{K_i}\frac{\mathbb{N}^{(i)}(T_{K_i,j})h_n^2(T_{K_i,j})}{\{\Lambda_n(T_{K_i,j}) + \alpha h_n(T_{K_i,j})\}^2} < 0$$

Thus $H_n'(\alpha)$ is non-increasing function. Therefore, to prove Theorem 1, it is sufficient to show that for some $\alpha = \alpha_0 > 0$ such that $H_n'(\alpha_0) < 0$ and $H_n'(-\alpha_0) > 0$ in probability. Then $\hat{\Lambda}_n^{ps}$ must be between $\Lambda_n - \alpha_0 h_n$ and $\Lambda_n + \alpha_0 h_n$ in probability, so $\|\hat{\Lambda}_n - \Lambda_n\|_2 \leq \alpha_0\|h_n\|_2$ in probability.

$nH_n'(\alpha_0)$ can be written as

$$\begin{aligned}\mathbb{P}_n \sum_{j=1}^{K}\left\{\frac{\mathbb{N}^{(i)}(T_{K,j})}{\Lambda_n+\alpha_0 h_n}-1\right\}h_n &= (\mathbb{P}_n - P)\sum_{j=1}^{K}\left\{\frac{\mathbb{N}^{(i)}(T_{K,j})}{\Lambda_n+\alpha_0 h_n}-1\right\}h_n + P\sum_{j=1}^{K}\left\{\frac{\mathbb{N}^{(i)}(T_{K,j})}{\Lambda_n+\alpha_0 h_n}-1\right\}h_n \\ &= I_{n1}+I_{n2}\end{aligned}$$

According to Lemma A·2 of Huang (1999), for $\eta > 0$ and any $\varepsilon \leq \eta$,

$$\log N_{[\,]}(\varepsilon, \varphi_n, L_2(\mu)) \leq cq_n \log(\eta/\varepsilon) \qquad \text{and} \qquad J_{[\,]}(\eta, \varphi_n, L_2(\mu)) \leq c_0 {q_n}^{1/2}\eta$$

For each $\varepsilon > 0$, since the bracket number of class φ_n is no more than $(\eta/\varepsilon)^{cq_n}$, there exists a set of brackets $\{[\Lambda_i^L, \Lambda_i^R] : i = 1, 2, \cdots, (\eta/\varepsilon)^{cq_n}\}$ such that for each $\Lambda \in \varphi_n$

$$\Lambda_i^L(t) \leq \Lambda(t) \leq \Lambda_i^R(t) \quad \text{for all } t \in O[T] \text{ and some } i$$

with $d^2(\Lambda_i^R, \Lambda_i^L) = \int \left\{\Lambda_i^R(t) - \Lambda_i^L(t)\right\}^2 d\mu(t) \leq \varepsilon^2$.

For any $\eta > 0$, define class

$$\mathcal{F}_\eta = \left\{\sum_{j=1}^{K}\left(\frac{\mathbb{N}(T_{K,j})}{\Lambda}-1\right)(\Lambda-\Lambda_n) : \quad \Lambda \in \varphi_n \text{ and } d(\Lambda, \Lambda_n) \leq \eta\right\}$$

By Cauchy-Schwartz inequality and Conditions 2-4, we can show that $\mathcal{F}_\eta$ is a Donsker class. Hence, $I_{n1} = O_p(n^{-1/2})$

Define $m(s) = \dfrac{\Lambda_0}{\Lambda_0 + s\Delta_n}$ where $\Delta_n = \Lambda_n - \Lambda_0 + \alpha_0 h_n$, $0 \leq s \leq 1$. By Taylor expansion,

$$m(s) = 1 + \left(-\frac{\Delta_n}{\Lambda_0}\right)s + \frac{\Lambda_0\Delta_n^2}{(\Lambda_0+\xi\Delta_n)^3}s^2 \quad \text{where } \xi \text{ between 0 and 1}$$

Since Λ_0 and Δ_n are bounded on $O[T]$, there exist constants c_1 and c_2 such that

$$c_1 E\sum_{j=1}^{K}\Delta_n^2 \leq E\sum_{j=1}^{K}\frac{\Lambda_0\Delta_n^2}{(\Lambda_0+\xi\Delta_n)^2} \leq c_2 E\sum_{j=1}^{K}\Delta_n^2$$

Therefore $E\sum_{j=1}^{k}\frac{\Lambda_0\Delta_n^2}{(\Lambda_0+\xi\Delta_n)^2} = O(n^{-\nu r} + n^{-(1-\nu)/2})$. Hence,

$$\begin{aligned} I_{2n} &\leq E\sum_{j=1}^{K}(-c_1\Delta_n + c_2\Delta_n^2)h_n \\ &\leq -\frac{c_1}{2}E\sum_{j=1}^{K}\Delta_n^2 = -\frac{c_1}{2}p_n^{-1}\end{aligned}$$

where $p_n^{-1} = n^{-\nu r} + n^{-(1-\nu)/2}$. Since $n^{-\nu r} + n^{-(1-\nu)/2} \geq n^{-1/(1+2r)} > n^{-1/2}$ for $0 < \nu < 1/2$, therefore

$$H_n'(\alpha_0) \leq O_p(n^{-1/2}) - cp_n^{-1} < 0$$

in probability. Same augments can apply for $H_n'(-\alpha_0) > 0$ in probability.

Proof of Theorem 2 . Let $m_\Lambda(x) = \sum_{j=1}^{K} \{\mathbb{N}(T_{K,j}) \log \Lambda(T_{K,j}) - \Lambda(T_{K,j})\}$ and define $M(\Lambda) = Pm_\Lambda(x)$ and $\mathbb{M}_n(\Lambda) = \mathbb{P}_n m_\Lambda(x)$. Then the log pseudolikelihood can be written as $\mathrm{n}\mathbb{P}_n m_\Lambda(x)$.

For any $\eta > 0$, define the class

$$\mathcal{F}_\eta = \{\Lambda | \Lambda \in \varphi_n,\ d(\Lambda, \Lambda_0) \leq \eta\}$$

By Theorem 1, $\hat{\Lambda}_n^{ps} \in \mathcal{F}_\eta$ for any $\eta > 0$ and sufficiently large n.

Next, define the class

$$\mathcal{M}_\eta = \{m_\Lambda(x) - m_{\Lambda_0}(x) :\ \Lambda \in \mathcal{F}_\eta\}$$

Using the results of Lemma A·2 of Huang (1999), we can easily establish

$$\log N_{[\ ]}(\varepsilon, \mathcal{M}_\eta, ||\cdot||_{P,B}) \leq cq_n \log(\eta/\varepsilon),$$

where $\|\cdot\|_{P,B}$ is *Bernstein Norm* defined to be $\|f\|_{P,B} = \{2P(e^{|f|} - 1 - |f|)\}^{1/2}$. Moreover, we have $||m_\Lambda(X) - m_{\Lambda_0}(X)||_{P,B}^2 \leq C\eta^2$ for any $m_\Lambda(X) - m_{\Lambda_0}(X) \in \mathcal{M}_\eta$ using the same arguments. Therefore, by Lemma 3·4·3 of van der Vaart & Wellner (1996), we obtain

$$E_P||n^{1/2}(\mathbb{P} - P)||_{\mathcal{M}_\eta} \leq CJ_{[\ ]}(\eta, \mathcal{M}_\eta, ||\ ||_{P,B}) \left\{1 + \frac{J_{[\ ]}(\eta, \mathcal{M}_\eta, ||\cdot||_{P,B})}{\eta^2 n^{1/2}}\right\} \quad \text{(A.1)}$$

where

$$J_{[\ ]}(\eta, \mathcal{M}_\eta, ||\cdot||_{P,B}) = \int_0^\eta \{1 + \log N_{[\ ]}(\varepsilon, \mathcal{F}, ||\ ||_{P,B})\}^{1/2} d\varepsilon \leq c_0 {q_n}^{1/2}\eta$$

The right hand side of (A.1) yields $\phi_n(\eta) = C({q_n}^{1/2}\eta + q_n/n^{1/2})$. It is easy to see that $\phi(\eta)/\eta$ is decreasing in η, and

$$r_n^2 \phi(\frac{1}{r_n}) = r_n {q_n}^{1/2} + r_n^2 q_n/n^{1/2} \leq n^{1/2}$$

for $r_n = n^{(1-\nu)/2}$ and $0 < \nu < 1/2$. Hence, $n^{(1-\nu)/2} d(\hat{\Lambda}_n^{ps}, \Lambda_0) = O_p(1)$ by Theorem 3·2·5 of van der Vaart & Wellner (1996). The choice of $\nu = 1/(1+2r)$ yields the rate of convergence of $r/(1+2r)$ which completes the proof.

References

Anderson, J. A. & Senthilselvan, A. (1980). Smooth estimates for the hazard function. *J. R. Statist. Soc.* B **42**, 322-327.

Andrews, D. F. & Herzberg, A. M. (1985). *Data; A Collection of Problems from Many Fields for the Student and Research Worker.* New York: Spring-Verlag.

Byar, D. P. (1980). The Veterans administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa. In *Bladder Tumours and Other Topics in Urological Oncology*, Ed. M. Macaluso, P. H. Smith and F. Edsmyr, pp. 363-370. New York: Plenum.

Byar, D. P., Blackard, C. & The Veterans Administration Cooperative Urological Research Group (1977). Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage I bladder cancer. *Urology* **10**, 556-561.

Cai, T. & Betensky, R. A. (2003). Hazard regression for interval-censored data with penalized spline. *Biometrics* **59**, 570-579.

Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *Ann. Statist.* **27**, 1536-1563.

Jamshidian, M. (2004). On algorithms for restricted maximum likelihood estimation. *Computational Statistics and Data Analysis* **45**, 137-157.

Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *Journal of Computational and Graphical Statistics* **7**, 310-321.

Kalbfleisch, J. D. & Lawless, J. F. (1985). The analysis of panel count data under a Markov assumption. *J. Am. Statist. Assoc.* **80**, 863-871.

Kooperberg, C., Stone, C. J. & Truong, Y. K. (1995). Hazard regression. *J. Am. Statist. Assoc.* **90**, 78-94.

Lee, E, W. & Kim, M. Y. (1998). The analysis of correlated panel count data using a continuous-time Markov model. *Biometrics* **54**, 1638-1644.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statist. Sci.* **3**, 425-441.

Rosen, J. B. (1960). The gradient projection method for nonlinear programming. *Journal of the society for industrial and applied mathematics* **8**, 181-217.

Rosenberg, P. S. (1995). Hazard function estimation using B-splines. *Biometrics* **51**, 874-887.

Schumaker, L. (1981). *Spline Functions: Basic Theory.* New York: Wiley.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14**, 590-606.

Sun, J. & Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statist. Sinica* **5** 279-290.

Sun, J. & Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *J. R. Statist. Soc.* B **62**, 293 -302.

Thall, P. F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics* **44**, 197-209.

Thall, P. F. & Lachin, J. M. (1988). Analysis of recurrent events: nonparametric methods for random-interval count data. *J. Am. Statist. Assoc.* **83**, 339-347.

Van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Application to Statistics.* New York: Springer-Verlag.

Wei, L. J., Lin, D. Y., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Am. Statist. Assoc.* **84**, 1065-1073.

Wellner, J. A. & Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28**, 779-814

Whittemore, A. S. & Keller, J. B. (1986). Survival estimation using splines. *Biometrics* **42**, 495-506.

Zhang, Y. & Jamshidian, M. (2004). On algorithms for NPMLE of the failure function with censored data. *The Journal of Computational and Graphical Statistics* **13**, 123-140.

Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika* **89**, 39-48.

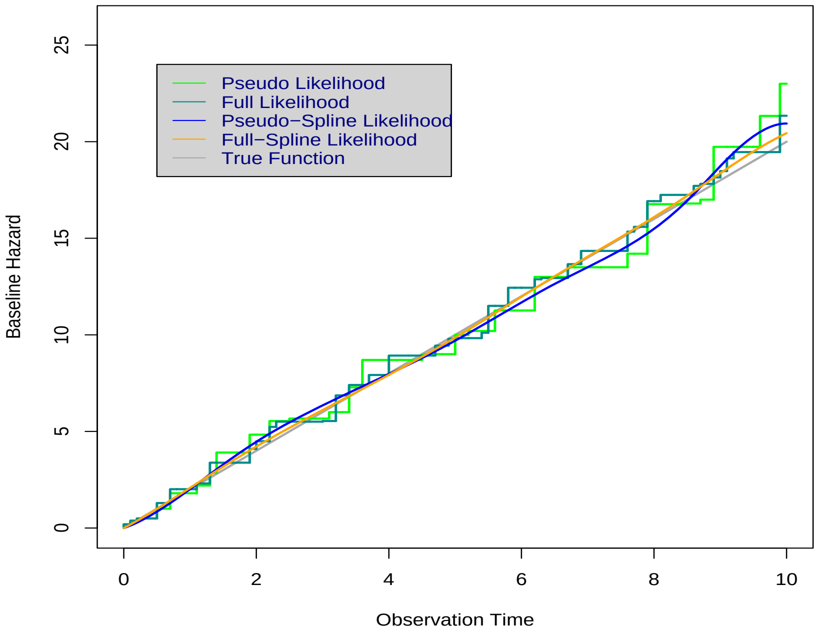

Figure 1. The four estimators of the mean function ($\Lambda_0(t) = 2t$) with panel count data generated from the Poisson process.

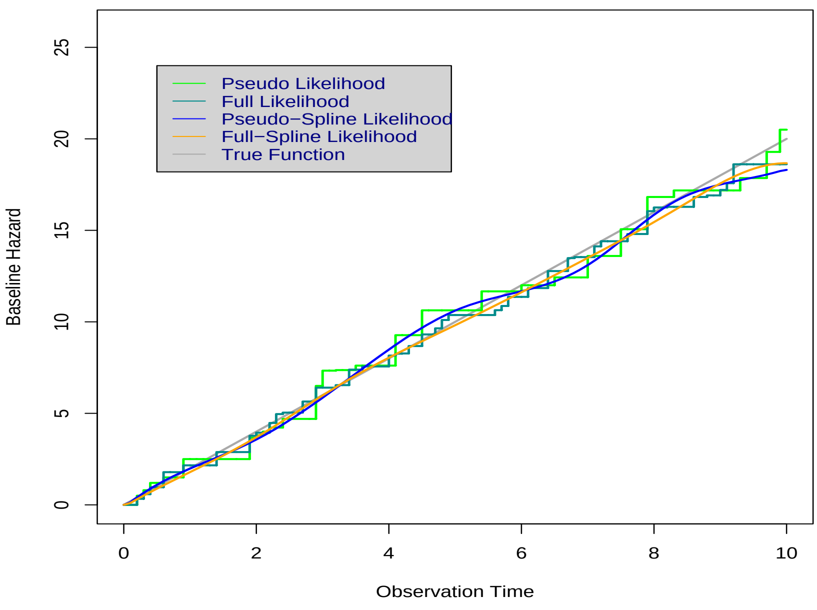

Figure 2. The four estimators of the mean function ($\Lambda_0(t) = 2t$) with panel count data generated from the mixed Poisson process.

Table 1: Monte-Carlo bias, standard deviation, and mean square error based on 1000 repeated samples for data generated from the Poisson process with sample size 100.

		Standard Deviation				**Bias**				**Square Root of MSE**			
Time	$\Lambda_0(t)$	a	b	c	d	a	b	c	d	a	b	c	d
1·5	3·0	0·452	0·397	0·305	0·231	-0·060	-0·048	-0·005	-0·006	0·456	0·400	0·305	0·231
2·0	4·0	0·526	0·417	0·334	0·262	-0·071	-0·038	-0·021	-0·016	0·531	0·419	0·335	0·262
2·5	5·0	0·570	0·452	0·338	0·270	-0·041	-0·022	-0·015	-0·012	0·571	0·453	0·338	0·270
3·0	6·0	0·623	0·461	0·396	0·299	-0·048	-0·022	-0·001	-0·003	0·625	0·462	0·396	0·299
3·5	7·0	0·656	0·480	0·427	0·321	-0·051	-0·035	0·010	0·001	0·658	0·481	0·427	0·321
4·0	8·0	0·688	0·474	0·424	0·323	-0·073	-0·049	0·009	-0·002	0·691	0·476	0·424	0·323
4·5	9·0	0·725	0·504	0·473	0·341	-0·074	-0·035	0·001	-0·007	0·728	0·505	0·473	0·341
5·0	10·0	0·730	0·516	0·522	0·365	-0·042	-0·017	-0·008	-0·012	0·732	0·516	0·522	0·365
5·5	11·0	0·786	0·526	0·521	0·371	-0·020	-0·055	-0·013	-0·012	0·786	0·529	0·521	0·372
6·0	12·0	0·812	0·542	0·553	0·386	-0·071	-0·055	-0·012	-0·009	0·815	0·544	0·553	0·386
6·5	13·0	0·826	0·545	0·610	0·411	-0·070	-0·035	-0·005	-0·006	0·829	0·546	0·617	0·411
7·0	14·0	0·829	0·559	0·599	0·420	-0·046	-0·047	0·007	-0·007	0·831	0·561	0·599	0·420
7·5	15·0	0·835	0·562	0·580	0·430	-0·066	-0·031	0·011	-0·010	0·838	0·562	0·580	0·430
8·0	16·0	0·846	0·575	0·632	0·453	-0·063	-0·059	-0·007	-0·016	0·849	0·578	0·632	0·454
8·5	17·0	0·902	0·609	0·674	0·466	-0·058	-0·078	-0·054	-0·024	0·903	0·614	0·677	0·466
9·0	18·0	0·933	0·622	0·719	0·504	-0·065	-0·073	-0·009	-0·025	0·935	0·627	0·725	0·504
9·5	19·0	1·047	0·662	0·807	0·550	-0·001	-0·064	-0·019	-0·015	1·047	0·665	0·807	0·550

a: Nonparametric pseudolikelihood estimator.

b: Nonparametric likelihood estimator.

c: Spline pseudolikelihood estimator.

d: Spline likelihood estimator.

Table 2: Monte-Carlo bias, standard deviation, and mean square error based on 1000 repeated samples for data generated from the Poisson process with sample size 200.

		Standard Deviation				**Bias**				**Square Root of MSE**			
Time	$\Lambda_0(t)$	a	b	c	d	a	b	c	d	a	b	c	d
1·5	3·0	0·388	0·314	0·232	0·182	-0·033	-0·024	-0·008	-0·009	0·389	0·315	0·232	0·182
2·0	4·0	0·435	0·340	0·240	0·190	-0·041	0·001	-0·006	-0·005	0·437	0·340	0·241	0·190
2·5	5·0	0·470	0·350	0·291	0·212	-0·022	-0·013	0·006	0·005	0·471	0·351	0·291	0·212
3·0	6·0	0·477	0·352	0·311	0·226	-0·017	-0·002	0·020	0·013	0·477	0·352	0·312	0·226
3·5	7·0	0·523	0·375	0·323	0·228	-0·012	0·026	0·025	0·014	0·523	0·376	0·324	0·228
4·0	8·0	0·541	0·370	0·361	0·244	-0·002	0·008	0·021	0·013	0·541	0·370	0·362	0·244
4·5	9·0	0·541	0·384	0·364	0·254	-0·017	-0·001	0·012	0·011	0·542	0·384	0·364	0·255
5·0	10·0	0·587	0·404	0·372	0·261	-0·027	0·003	0·005	0·012	0·588	0·404	0·372	0·261
5·5	11·0	0·589	0·410	0·414	0·278	-0·004	0·007	0·004	0·013	0·590	0·410	0·414	0·278
6·0	12·0	0·625	0·422	0·419	0·286	-0·001	0·012	0·011	0·013	0·625	0·422	0·419	0·287
6·5	13·0	0·606	0·435	0·406	0·290	0·018	0·019	0·022	0·012	0·606	0·435	0·406	0·290
7·0	14·0	0·647	0·446	0·447	0·307	0·024	0·017	0·029	0·011	0·647	0·446	0·448	0·307
7·5	15·0	0·672	0·459	0·460	0·317	-0·001	0·016	0·024	0·012	0·672	0·460	0·460	0·317
8·0	16·0	0·677	0·480	0·471	0·334	-0·034	0·009	0·005	0·016	0·678	0·480	0·471	0·334
8·5	17·0	0·725	0·491	0·526	0·358	-0·056	0·015	-0·029	0·017	0·728	0·492	0·526	0·358
9·0	18·0	0·716	0·498	0·536	0·367	-0·066	-0·005	-0·068	0·008	0·719	0·498	0·541	0·367
9·5	19·0	0·784	0·535	0·585	0·396	0·003	-0·015	-0·020	-0·001	0·784	0·535	0·586	0·396

a: Nonparametric pseudolikelihood estimator.

b: Nonparametric likelihood estimator.

c: Spline pseudolikelihood estimator.

d: Spline likelihood estimator.

Table 3: Monte-Carlo bias, standard deviation and mean square error based on 1000 repeated samples for data generated from the mixed Poisson process with sample size 100.

		Standard Deviation				**Bias**				**Square Root of MSE**			
Time	$\Lambda_0(t)$	a	b	c	d	a	b	c	d	a	b	c	d
1·5	3·0	0·483	0·418	0·307	0·243	-0·054	-0·029	0·005	0·007	0·487	0·419	0·307	0·243
2·0	4·0	0·540	0·456	0·348	0·276	-0·057	-0·027	-0·012	-0·007	0·543	0·457	0·348	0·276
2·5	5·0	0·606	0·467	0·373	0·284	-0·095	-0·032	-0·010	-0·009	0·613	0·468	0·373	0·285
3·0	6·0	0·642	0·495	0·447	0·311	-0·064	-0·023	-0·005	-0·005	0·645	0·496	0·447	0·311
3·5	7·0	0·702	0·527	0·482	0·336	-0·062	-0·033	-0·009	-0·004	0·704	0·528	0·482	0·336
4·0	8·0	0·737	0·565	0·471	0·346	-0·073	-0·013	-0·024	-0·005	0·741	0·565	0·472	0·346
4·5	9·0	0·743	0·597	0·512	0·369	-0·103	-0·047	-0·040	-0·007	0·750	0·599	0·513	0·369
5·0	10·0	0·784	0·630	0·558	0·397	-0·123	-0·009	-0·044	-0·007	0·793	0·630	0·560	0·397
5·5	11·0	0·845	0·651	0·554	0·411	-0·098	-0·006	-0·033	-0·004	0·851	0·651	0·555	0·411
6·0	12·0	0·875	0·685	0·592	0·434	-0·075	-0·019	-0·010	-0·001	0·878	0·685	0·592	0·434
6·5	13·0	0·907	0·688	0·664	0·464	-0·037	-0·001	0·012	0·002	0·908	0·688	0·664	0·464
7·0	14·0	0·915	0·701	0·667	0·474	-0·050	-0·029	0·025	0·004	0·916	0·702	0·667	0·474
7·5	15·0	0·943	0·722	0·659	0·490	-0·022	-0·007	0·017	0·004	0·943	0·722	0·660	0·490
8·0	16·0	0·971	0·746	0·714	0·524	-0·066	-0·032	-0·014	-0·001	0·974	0·747	0·714	0·524
8·5	17·0	1·002	0·744	0·761	0·549	-0·056	-0·036	-0·070	-0·010	1·004	0·745	0·764	0·549
9·0	18·0	1·023	0·772	0·828	0·589	-0·077	-0·050	-0·104	-0·017	1·026	0·773	0·834	0·590
9·5	19·0	1·256	0·810	0·950	0·627	0·037	-0·033	-0·001	0·001	1·257	0·811	0·950	0·627

a: Nonparametric pseudolikelihood estimator.

b: Nonparametric likelihood estimator.

c: Spline pseudolikelihood estimator.

d: Spline likelihood estimator.

Table 4: Monte-Carlo bias, standard deviation and mean square error based on 1000 repeated samples for data generated from the mixed Poisson process with sample size 200.

		Standard Deviation				**Bias**				**Square Root of MSE**			
Time	$\Lambda_0(t)$	a	b	c	d	a	b	c	d	a	b	c	d
1·5	3·0	0·386	0·311	0·241	0·179	-0·024	-0·010	0·010	0·006	0·386	0·311	0·241	0·179
2·0	4·0	0·433	0·319	0·239	0·183	-0·019	0·003	0·016	0·009	0·434	0·319	0·240	0·183
2·5	5·0	0·463	0·342	0·284	0·204	-0·004	-0·001	0·024	0·012	0·463	0·342	0·285	0·204
3·0	6·0	0·512	0·371	0·307	0·222	-0·024	-0·005	0·022	0·011	0·513	0·371	0·309	0·223
3·5	7·0	0·519	0·366	0·320	0·231	-0·001	0·007	0·015	0·009	0·519	0·366	0·321	0·231
4·0	8·0	0·546	0·394	0·367	0·248	-0·027	0·010	0·011	0·010	0·547	0·394	0·367	0·249
4·5	9·0	0·581	0·422	0·393	0·260	-0·002	0·025	0·014	0·016	0·581	0·422	0·393	0·261
5·0	10·0	0·619	0·408	0·412	0·272	-0·029	0·026	0·017	0·020	0·619	0·409	0·412	0·273
5·5	11·0	0·648	0·446	0·448	0·291	-0·028	-0·004	0·013	0·020	0·649	0·446	0·448	0·292
6·0	12·0	0·657	0·442	0·460	0·303	-0·012	0·021	0·005	0·017	0·658	0·443	0·460	0·304
6·5	13·0	0·684	0·464	0·460	0·310	-0·023	0·009	0·006	0·014	0·684	0·464	0·460	0·310
7·0	14·0	0·710	0·492	0·500	0·329	-0·019	0·019	0·019	0·014	0·710	0·492	0·501	0·329
7·5	15·0	0·731	0·512	0·508	0·344	-0·001	0·020	0·040	0·016	0·731	0·513	0·509	0·345
8·0	16·0	0·774	0·516	0·517	0·359	0·023	0·024	0·052	0·022	0·774	0·516	0·520	0·360
8·5	17·0	0·758	0·546	0·581	0·382	0·018	0·024	0·034	0·028	0·759	0·546	0·582	0·383
9·0	18·0	0·794	0·559	0·613	0·408	0·022	0·036	-0·016	0·031	0·794	0·560	0·613	0·409
9·5	19·0	0·885	0·610	0·693	0·446	0·026	0·013	0·010	0·033	0·885	0·610	0·693	0·447

a: Nonparametric pseudolikelihood estimator.

b: Nonparametric likelihood estimator.

c: Spline pseudolikelihood estimator.

d: Spline likelihood estimator.

Table 5: Comparison of computing time in seconds among the nonparametric estimators and spline estimators of the mean function, based on the data generated from the Poisson process or mixed Poisson process with sample size 100 or 200.

	Poisson Process		**Mixed Poisson Process**	
Estimators	n = 100	n = 200	n =100	n = 200
a	0.24	0.66	0.23	0.61
b	60.34	172.15	58.92	168.43
c	2.33	4.90	2.24	5.07
d	4.67	11.41	4.59	12.52

a: Nonparametric pseudolikelihood estimator.

b: Nonparametric likelihood estimator.

c: Spline pseudolikelihood estimator.

d: Spline likelihood estimator.

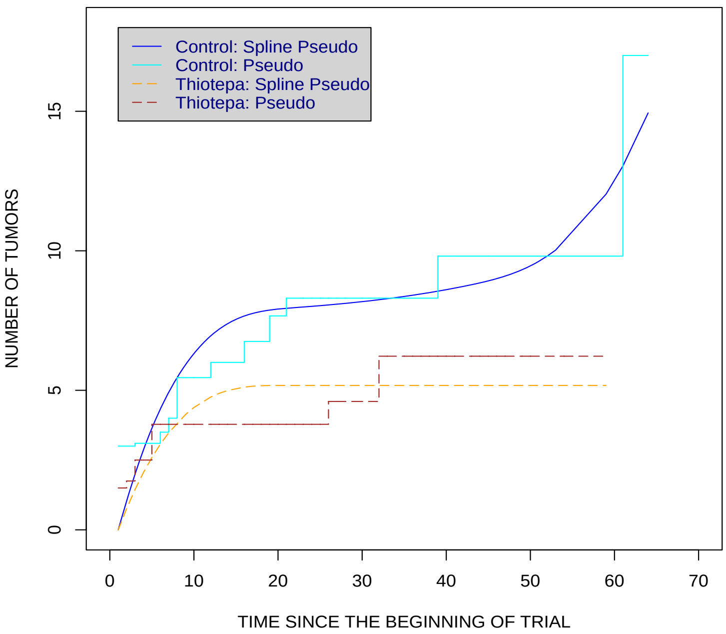

Figure 3. Nonparametric and spline pseudolikelihood estimations of the mean functions for control group and thiotepa group with the bladder tumour data.

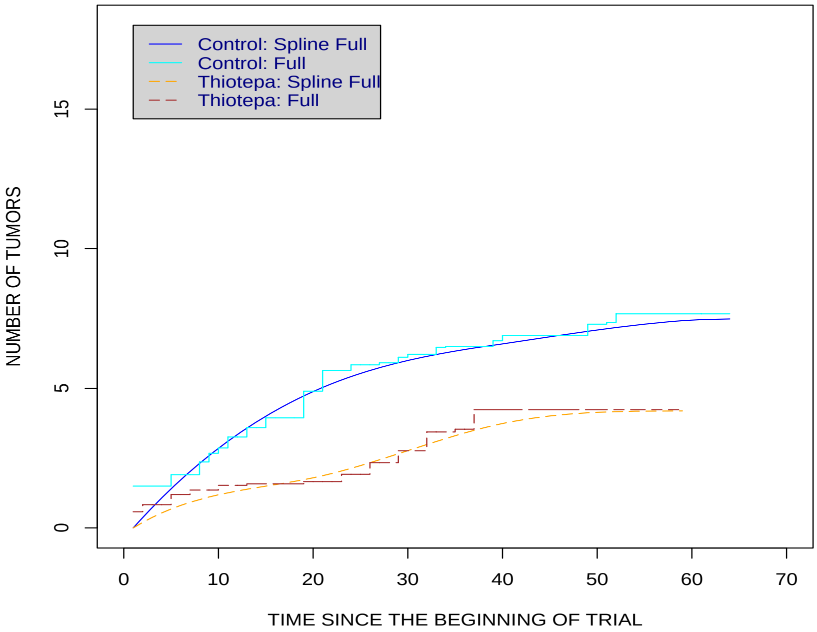

Figure 4. Nonparametric and spline likelihood estimations of the mean functions for control group and thiotepa group with the bladder tumour data.