Test (2011) 20: 1–42
DOI 10.1007/s11749-010-0223-1

INVITED PAPER

Nonparametric inference based on panel count data

Xingqiu Zhao · N. Balakrishnan · Jianguo Sun

Received: 14 September 2009 / Accepted: 18 November 2010 / Published online: 25 December 2010

Abstract Panel count data usually refer to data arising from studies on recurrent events in which the subjects under study are followed or observed only periodically rather than continuously. In such situations, an objective of interest is about the occurrence of some events that can occur multiple times or repeatedly and the studies resulting in this type of information are often referred to as event history studies. There are many fields such as medical studies, reliability experiments and social sciences wherein panel count data are encountered commonly. This article reviews basic concepts about panel count data, some common issues and questions of interest regarding them as well as the corresponding statistical procedures that are suitable for their analysis. In particular, we will discuss an estimation of the mean function of the underlying counting process characterizing the occurrence of the events, comparison of several processes and analysis of multiple state panel count data. Some discussion is also presented of situations involving dependent or informative observation processes.

Communicated by Domingo Morales.

This invited paper is discussed in the comments available at: doi:10.1007/s11749-010-0224-0, doi:10.1007/s11749-010-0225-z, doi:10.1007/s11749-010-0226-y, doi:10.1007/s11749-010-0227-x, doi:10.1007/s11749-010-0228-9, doi:10.1007/s11749-010-0229-8.

X. Zhao
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, Hong Kong
e-mail: xingqiu.zhao@polyu.edu.hk

N. Balakrishnan
Department of Mathematics & Statistics, McMaster University, Hamilton, Canada L8S 4K1
e-mail: bala@mcmaster.ca

J. Sun (✉)
Department of Statistics, University of Missouri, Columbia, MO 65211, USA
e-mail: sunj@missouri.edu

Keywords Bayesian estimation · Generalized least-squares · Mean function · Markov model · Nonparametric comparison · Nonparametric maximum likelihood · Nonparametric maximum pseudo-likelihood · Panel count data · Rate function

Mathematics Subject Classification (2000) Primary 62G05 · 62G20 · Secondary 62N02 · 62N03

1 Introduction

This article discusses statistical analysis of panel count data arising from recurrent event or event history studies regarding the occurrence of some recurrent events. The data from these studies are often referred to as event history data and they are encountered commonly in many areas such as medical studies, reliability experiments and social sciences (Nelson 2003; Vermunt 1997). The event history data can be generally classified into two types. One is from the studies that monitor subjects under study continuously and consequently provide information on the times of all occurrences of recurrent events. These data are usually referred to as recurrent event data (Cook and Lawless 2007). The other type is the so-called panel count data discussed here and they arise when subjects under study are examined or observed only at discrete time points and so provide only the numbers of occurrences of the events between subsequent observation times.

Examples of recurrent event data include data on occurrences of the hospitalizations of intravenous drug users (Wang et al. 2001), occurrences of the same infection such as recurrent pyogenic infections among inherited disorder patients (Lin et al. 2000), repeated occurrences of certain tumors, and warranty claims for an automobile (Kalbfleisch et al. 1991). These situations provide examples of panel count data if the continuous observation scheme is changed to a discrete observation scheme. The panel count data could occur for various reasons. For example, they may arise because the continuous observation may be too expensive or impossible, or when it is not practical to conduct continuous follow-ups of subjects under study.

To give a specific example of panel count data, consider the data discussed in Thall and Lachin (1988) and given in the data set IV of Appendix A in Sun (2006) among others. They arose from a clinical trial on the use of the natural bile chenodeoxycholic acid for the dissolution of cholesterol gallstones. The data include the observed information on the incidence of nausea from the first year follow-up on 111 patients with floating gallstones in high-dose and placebo groups. The original study involves 10-year follow-ups and consists of three groups, viz., placebo, low dose and high dose. Nausea is an unpleasant sensation vaguely referred to the epigastrium and abdomen, often culminating in vomiting, and it is very commonly associated with gallstone disease. During the study, the patients were scheduled to return for clinical observations at 1, 2, 3, 6, 9, and 12 months during the first year follow-up. At each visit, they were asked to report the total number of incidences of nausea among other symptoms that had occurred between successive visits. That is, the observed data include actual visit times and the numbers of incidences or occurrences of nausea between the visits. As expected, actual visit or observation times differ from patient to patient. Another example of panel count data, which has been discussed by many authors, can be found in the data set V of Appendix A in Sun (2006). It arose from a bladder cancer study consisting of patients with superficial bladder tumors and more discussion on it will be provided later on.

A special case of panel count data that often occurs in practice is when each subject is observed only once and such data are usually referred to as current status data (Sun and Kalbfleisch 1993). In this situation, only available information about the recurrent event of interest is the total number of the occurrences of the event up to the observation time. A common example of current status data arises in tumorgenicity experiments that concern the occurrence rate of certain tumors and in these experiments, it is often the case that only the number of tumors that have occurred before the death or sacrifice of the animal is known. Another area that frequently produces current status data is demographical studies (Diamond and McDonald 1991).

Many authors have discussed statistical analysis of recurrent event data (Cook and Lawless 1996; Lawless and Nadeau 1995; Lin et al. 2000; Pepe and Cai 1993; Wang and Chen 2000). For example, Andersen et al. (1993) have provided a comprehensive coverage of counting process approaches for the analysis of recurrent event data. More recently, Cook and Lawless (2007) have given a relatively complete and thorough review of the literature on recurrent event data wherein more references can be found.

Comparatively, sparse literature exists on the analysis of panel count data. Among the authors who considered this situation, Kalbfleisch and Lawless (1985) discussed the fitting of Markov model to panel count data and Sun and Kalbfleisch (1995) and Wellner and Zhang (2000) studied estimation of the mean function of the underlying counting process that yields panel count data. For treatment comparison based on panel count data, several approaches have been developed. For example, one of the early papers on this was given by Thall and Lachin (1988), who suggested the use of some data grouping method. Sun and Fang (2003) proposed a model-free approach assuming that the treatment indicators can be regarded as independent and identically distributed random variables, wherein the isotonic regression estimator is used for the mean function. Also Park et al. (2007) gave a class of nonparametric two-sample tests based on the isotonic regression estimator, while Zhang (2006) and Balakrishnan and Zhao (2010b) developed some multi-sample nonparametric procedures by using nonparametric maximum pseudo-likelihood approach. Instead of using the isotonic regression estimator or the nonparametric maximum pseudo-likelihood estimator, Balakrishnan and Zhao (2009) proposed two new classes of test statistics by using the nonparametric maximum likelihood estimator. They showed by means of Monte Carlo simulations that their tests are more powerful and robust than those based on the nonparametric maximum pseudo-likelihood approach. Note that the nonparametric maximum likelihood estimator is in general more efficient but more complicated both theoretically and computationally. The authors who considered regression analysis of panel count data include Sun and Wei (2000), Zhang (2002), and Wellner and Zhang (2007).

It is of interest and helpful to mention that in addition to the amount of relevant information available being different between recurrent event data and panel count data, yet another key difference is the observation process. In the case of the former, the observation process means the length of the whole follow-up, while in the case of the latter, it also includes a sequence of consecutive observation times. Most of the methods mentioned above or discussed in the literature assume the observation process to be independent of the underlying counting process generating recurrent event or panel count data. As will be discussed below, there are also some approaches that allow dependent observation processes.

To analyze recurrent event data, it is common and convenient to characterize the occurrences of recurrent events by counting processes and to model the intensity process of the counting process (Andersen et al. 1993). On the other hand, for the analysis of panel count data, it is usually more convenient to work directly on the mean function of the counting processes conditional on covariate processes due to the incomplete nature of the observed information. In this case, a natural and simple approach is to fit the panel count data to parametric Poisson processes or mixed parametric Poisson processes. For example, Hinde (1982) and Breslow (1984) discussed regression analysis of Poisson count data, while Thall (1988) suggested regression approaches for mixed Poisson processes. An alternative parametric approach is to treat the data as longitudinal count data and to use the generalized estimating equation approach (Diggle et al. 1994). In this article, we focus our attention on nonparametric approaches that regard observed data as realizations of some underlying counting processes.

In the following, first in Sect. 2, we briefly review some basic concepts on counting processes and some nonparametric estimation procedures for recurrent event data. Section 3 considers nonparametric estimation of the mean function of counting processes giving rise to panel count data by assuming that subjects under study come from a homogeneous population. For this specific problem, several inferential procedures are discussed. In Sect. 4, we discuss the treatment comparison problem based on panel count data. We formulate the problem through the comparison of the mean functions of different counting processes and discuss some nonparametric procedures. Section 5 presents some numerical results comparing nonparametric maximum likelihood-based and pseudo-likelihood-based treatment comparison procedures as well as two illustrative examples. In Sect. 6, the analysis of multiple state panel count data is discussed with the focus on the data arising from the Markov model.

Throughout Sects. 3 to 6, we will assume that the observation process is independent of the underlying counting process of interest. Of course, this may not hold in practice. Section 7 discusses some situations when the two processes may be related with each other and two specific ideas are described for modeling the relationship between the two processes. In Sect. 8, to make the article complete, we briefly consider the situation when there are covariates and describe a few inferential procedures with a focus on the marginal modeling approach for the underlying counting process. Section 9 indicates some ideas about the use of Bayesian approaches for the analysis of panel count data, and finally Sect. 10 contains some discussions and concluding remarks. Note that in practice, one could regard panel count data as a special type of longitudinal data and apply the methodology developed for general longitudinal data. However, a major drawback in this approach is that one would miss the special structure of panel count data. Moreover, some questions of interest in panel count data cannot be answered from the longitudinal data viewpoint.

2 Notation and review

In this section, we will first introduce some notation and review some basic concepts about counting processes that will be used throughout the article. Some nonparametric estimation procedures for recurrent event data will then be discussed.

2.1 Counting processes

Counting processes have been playing an essential role in the development of statistical models and inferential procedures for event history analysis. Some of the early and significant contributions to this were given by Aalen (1975, 1978) and Andersen and Borgan (1985). They and others established the connection between the counting process and event history analysis and showed how the theory of multivariate counting processes can provide a general framework and a useful tool for event history analysis. For detailed description and discussion of this, readers are referred to Andersen et al. (1993) in addition to the references mentioned above.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{T} = [0, \tau)$ be a continuous-time interval, where τ is a given terminal time, $0 < \tau \leq \infty$. A stochastic process X is a family of random variables $\{X(t) : t \in \mathcal{T}\}$. A filtration or history $(\mathcal{F}_t : t \in \mathcal{T})$ is an increasing right-continuous family of sub-σ-algebras of $\mathcal{F}$ such that $\mathcal{F}_t$ contains all the information generated by the stochastic process X on $[0, t]$. The process X is said to be adapted to the filtration if $X(t)$ is $\mathcal{F}_t$-measurable for every $t \in \mathcal{T}$. A process X is predictable with respect to $\mathcal{F}_t$ if $X(t)$ is known given the history $\mathcal{F}_{t-}$, where $\mathcal{F}_{t-}$ is generated by $(X(s), 0 \leq s < t)$.

A counting process is a stochastic process $\{N(t); t \geq 0\}$ with $N(0) = 0$ and $N(t) < \infty$ almost surely such that the path is right-continuous with probability one, piecewise constant, and has only jump discontinuities with jumps of size $+1$. To model the counting process, one usually employs its intensity process defined as

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{P\{N(t + \Delta t-) - N(t-) = 1 | \mathcal{F}_{t-}\}}{\Delta t}$$

and imposes some assumptions on its format. Given $\lambda(t)$, one can obtain the so-called cumulative intensity process $\Lambda(t) = \int_0^t \lambda(s)\, ds$ and could directly model $\Lambda(t)$ too. Sometimes instead of the intensity or cumulative intensity process, it may be more convenient to model the mean or rate function $r(t)$ of $N(t)$ defined as

$$E\big\{\Delta N(t)\big\} = r(t)\Delta t + o(\Delta t)$$

(Cook and Lawless 2007), where $\Delta N(t) = N(t + \Delta t-) - N(t-)$ representing the number of events in the short interval $[t, t + \Delta t)$. Given $r(t)$, the mean function can be calculated as $\int_0^t r(s)\, ds$. Note that it is easy to see that the mean or rate function cannot completely specify the counting process $N(t)$ and they are sometimes referred to as the marginal cumulative intensity or intensity function. One major advantage of dealing with the mean or rate function is that less assumptions are usually needed in modeling them compared to modeling the intensity process, which can thus lead to more robust inferential procedures.

Among the counting processes, the most commonly used one is perhaps the Poisson process $\{N(t); t \geq 0\}$ usually defined by

$$P\left\{N(t+dt) - N(t) = 1 | \mathcal{F}_{t-}\right\} = \lambda(t)\, dt + o(dt)$$

and

$$P\left\{N(t+dt) - N(t) \geq 2 | \mathcal{F}_{t-}\right\} = o(dt)$$

with $\lambda(t)$ being a left-continuous function. In other words, the Poisson process $N(t)$ has at most one jump over a small time interval and does not depend on its history. The Poisson process defined above is commonly referred to as a non-homogeneous Poisson process. If $\lambda(t)$ is constant, the process is usually called a homogeneous Poisson process. For a Poisson process $\{N(t); t \geq 0\}$, we have that, at each t, $N(t)$ follows the Poisson distribution with $E\{N(t)\} = \Lambda(t) = \int_0^t \lambda(s)\, ds$. That is, $\Lambda(t)$ is also the mean function of the Poisson process and in this situation, we have $r(t) = \lambda(t) = d\Lambda(t)/dt$.

2.2 Nonparametric inference for recurrent event data

Consider a study consisting of a single type of recurrent event and n independent subjects. Let $N_i(t)$ denote the number of occurrences of the event over the interval $[0, t]$ for subject i and assume that each subject is observed continuously up to time τ_i, denoting the time at the end of observation for subject i. That is, we have recurrent event data. Define the left-continuous function $Y_i(t) = I(t \leq \tau_i)$, indicating whether subject i is under observation at time t, $i = 1, \ldots, n$. First we will assume that all subjects come from a homogeneous population and the intensity process $\lambda_i(t)$ for $N_i(t)$ has the form $\lambda_i(t) = \alpha(t) Y_i(t)$, where $\alpha(t)$ is a non-negative deterministic function. To estimate $\Lambda(t) = \int_0^t \alpha(s)\, ds$, motivated by the fact that $N_i(t) - \int_0^t \alpha(s) Y_i(s)\, ds$ is a martingale, a commonly used estimate is given by the so-called Nelson–Aalen estimator,

$$\hat{\Lambda}(t) = \int_0^t \frac{J(s)\, dN.(s)}{Y.(s)}$$

(Andersen et al. 1993). In the above, $N.(t) = \sum_{i=1}^n N_i(t)$, $Y.(t) = \sum_{i=1}^n Y_i(t)$ and $J(t) = I(Y_.(t) > 0)$. It is easy to see that $N.(t)$ and $Y_.(t)$ denote the total number of occurrences of the event up to time t and the number of subjects still under observation at time t, respectively.

Let $t_1 < t_2 < \cdots$ denote the sequence of all distinct occurrence times; then the Nelson–Aalen estimator can be rewritten as

$$\hat{\Lambda}(t) = \sum_{j: t_j \leq t} \frac{\Delta N.(t_j)}{Y.(t_j)},$$

where $\Delta N.(t_j) = N_.(t_j) - N_.(t_{j-1})$.

Given $\hat{\Lambda}(t)$, it is obvious that one can estimate $\alpha(t)$ by

$$\hat{\alpha}(t) = \frac{\Delta N.(t)}{Y.(t)}$$

or more generally by a kernel estimate

$$\hat{\alpha}_K(t) = \frac{1}{b} \int_{t-b}^{t+b} K\left(\frac{t-s}{b}\right) d\hat{\Lambda}(s),$$

where $K(t)$ is a kernel function and b is a positive constant called the bandwidth.

Now suppose that all subjects come from two different populations with n_1 and n_2 denoting the numbers of subjects from the first and second populations, respectively, where $n_1 + n_2 = n$. Also suppose that for the subjects in population l, we have $\lambda_i(t) = \alpha_l(t) Y_i(t)$ and one is interested in testing $H_0 : \alpha_1(t) = \alpha_2(t)$. That is, we have a two-sample comparison problem. To test H_0, let $\hat{\Lambda}_1(t)$ and $\hat{\Lambda}_2(t)$ denote the estimates of $\Lambda_1(t) = \int_0^t \alpha_1(s)\, ds$ and $\Lambda_2(t) = \int_0^t \alpha_2(s)\, ds$, respectively, defined as $\hat{\Lambda}(t)$ with subjects only from each individual population. Then a natural test statistic can be constructed as

$$\int_0^\tau W(t)\, d\left\{\hat{\Lambda}_1(t) - \hat{\Lambda}(t)\right\}$$

or

$$\int_0^\tau W(t)\, d\left\{\hat{\Lambda}_1(t) - \hat{\Lambda}_2(t)\right\},$$

where $W(t)$ is a bounded predictable weight process and τ denotes the longest follow-up time (Andersen et al. 1993).

3 Nonparametric estimation with panel count data

In this section, we discuss nonparametric estimation for panel count data. For this, our focus will be on the estimation of the mean function of the underlying counting process since it is impossible in general to deal with the intensity or cumulative intensity process. To introduce the notation, consider an event history study that involves n independent subjects from a homogeneous population and wherein each subject gives rise to a counting process $N_i(t)$, representing the total number of occurrences of the recurrent event of interest from subject i up to time t, $i = 1, \ldots, n$. In the following, to follow the literature, we will use $\Lambda(t)$ to denote the mean function of the $N_i(t)$'s. That is, $\Lambda(t) = E\{N_i(t)\}$, $i = 1, \ldots, n$. For subject i, suppose that $N_i(\cdot)$ is observed only at finite time points $T_{i,1} < \cdots < T_{i,K_i}$, where K_i denotes the number of observation times, $i = 1, \ldots, n$. Then the observed panel count data on the $N_i(t)$'s have the form

$$\left\{\left(K_i, T_{i,j}, N_i(T_{i,j})\right);\, j = 1, \ldots, K_i,\, i = 1, \ldots, n\right\}.$$

Note here that different subjects can have different numbers of observations and also different observation times.

For nonparametric estimation of $\Lambda(t)$, in the following, we will first describe a general and simple estimator of $\Lambda_0(t)$ commonly called the isotonic regression estimator. Two likelihood function-based estimation procedures will then be presented along with some other estimates proposed in the literature.

3.1 Isotonic regression estimator (IRE)

To present the IRE, we first consider a simple situation in which $T_{i,j} = s_j$ for all $i = 1, \ldots, n$ with $K_i = K$; that is, all subjects have the same observation time points and the same numbers of observations. This can be the case in a follow-up study with pre-specified observation time points and in which all subjects follow the pre-specified observation schedule. It is evident that for this situation, a natural estimate of $\Lambda(t)$ at the observation time s_l is given by the sample mean

$$\frac{\sum_{i=1}^n N_i(s_l)}{\sum_{i=1}^n I(s_l \le T_{i,K_i})} = \sum_{j=1}^{l} \frac{\sum_{i=1}^n I(s_j \le T_{i,K_i})[N_i(s_j) - N_i(s_{j-1})]}{\sum_{i=1}^n I(s_j \le T_{i,K_i})}.$$

Note that it is clear that one can estimate $\Lambda(t)$ only at the observation times s_l and for any time point between s_{l-1} and s_l, the mean function $\Lambda(t)$ can take on any value between the estimated values of $\Lambda(t)$ at s_{l-1} and s_l, where $s_0 = 0$. Assume that $\Lambda(t)$ is a step function with jumps only at the s_l's. Then, the sample mean estimate at these s_l's given above can be rewritten as

$$\int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n I(s \le T_{i,K_i})},$$

which is the Nelson–Aalen estimator described above for recurrent event data.

Now consider the general situation wherein subjects may have different observation times as well as different numbers of observations. In this case, it is clear that one can still estimate $\Lambda(t)$ by using the sample mean based on the observed values at each individual observation time. However, the resulting estimate may not be valid as it may not have the non-decreasing property that $\Lambda(t)$ possesses. To fix this, Sun and Kalbfleisch (1995) proposed the following IRE. Let $s_1 < \cdots < s_m$ denote the ordered and different time points of all observation time points $\{T_{i,j}\}$, and w_l and $\bar{N}_l$ be the number and mean value, respectively, of the observations made at s_l, $l = 1, \ldots, m$. Then, the isotonic regression estimator, denoted by $\hat{\Lambda}_I = (\hat{\Lambda}_{I,1}, \ldots, \hat{\Lambda}_{I,m})$, of $\Lambda(t)$ at the s_l's is defined as $\Lambda_I = (\Lambda_1, \ldots, \Lambda_m) = (\Lambda(s_1), \ldots, \Lambda(s_m))$ that minimizes the weighted sum of squares

$$\sum_{l=1}^{m} w_l (\bar{N}_l - \Lambda_l)^2 \tag{1}$$

subject to the order restriction $\Lambda_1 \le \cdots \le \Lambda_m$. It can be easily seen that if $\bar{N}_1 \le \cdots \le \bar{N}_m$, then we have $\hat{\Lambda}_{I,l} = \bar{N}_l$, $l = 1, \ldots, m$. Also for the simple situation discussed above, the IRE reduces to the sample mean estimator presented earlier.

By definition, the IRE defined above is the isotonic regression of $\{\bar{N}_1, \ldots, \bar{N}_m\}$ with weights $\{w_1, \ldots, w_m\}$ (Robertson et al. 1988). Thus, $\hat{\Lambda}_I(s_\ell)$ has a closed-form expression given by

$$\hat{\Lambda}_I(s_\ell) = \max_{r \le \ell} \min_{s \ge \ell} \frac{\sum_{v=r}^{s} w_v \bar{N}_v}{\sum_{v=r}^{s} w_v} = \min_{s \ge \ell} \max_{r \le \ell} \frac{\sum_{v=r}^{s} w_v \bar{N}_v}{\sum_{v=r}^{s} w_v}, \quad \ell = 1, \ldots, m$$

using the max-min formula for the isotonic regression (Robertson et al. 1988). Also it can be shown that for current status data, the IRE is actually the nonparametric maximum likelihood estimate of $\Lambda(t)$ if the $N_i(t)$'s are Poisson processes (Sun and Kalbfleisch 1995).

3.2 Nonparametric likelihood-based estimation

It is clear that the IRE does not depend on any distributional assumption. Sometimes it may be reasonable to assume that the $N_i(t)$'s are non-homogeneous Poisson processes and in this case, it is apparent that one may want to use the likelihood-based estimates. More specifically, let $\Lambda(t)$ be defined as before and assume that $\{(K_i; T_{i,1}, \ldots, T_{i,K_i})\}$ are independent of N_i. By ignoring the dependency of the events within a subject, one can construct a pseudo-log-likelihood function for $\Lambda(t)$ as

$$l_n^{\mathrm{ps}}(\Lambda) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \big\{ N_i(T_{i,j}) \log\big(\Lambda(T_{i,j})\big) - \Lambda(T_{i,j}) \big\}, \tag{2}$$

omitting the parts independent of Λ.

Let $s_1 < \cdots < s_m$ denote the ordered distinct observation time points in the set of all observation time points $\{T_{i,j}, j = 1, \ldots, K_i, i = 1, \ldots, n\}$. Then, a nonparametric maximum pseudo-likelihood estimator (NPMPLE) of Λ can be defined as a non-decreasing step function with possible jumps only at the s_l's that maximizes $l_n^{\mathrm{ps}}(\Lambda)$. Note that the pseudo-log-likelihood function can be rewritten as

$$l_n^{\mathrm{ps}}(\Lambda) = \sum_{l=1}^{m} w_l (\bar{N}_l \log \Lambda_l - \Lambda_l).$$

It can be shown that the maximization of the above pseudo-log-likelihood function is equivalent to the minimization of (1) (Robertson et al. 1988; Wellner and Zhang 2000). That is, the NPMPLE is the same as the IRE and hence can be computed by using the max-min formula for the IRE presented in the preceding subsection.

Instead of using the pseudo-log-likelihood function l_n^{ps}, to estimate $\Lambda(t)$, one can also apply the full log-likelihood function of the mean function Λ given by

$$l_n(\Lambda) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \big(N_i(T_{i,j}) - N_i(T_{i,j-1}) \big) \log\big(\Lambda(T_{i,j}) - \Lambda(T_{i,j-1}) \big) - \sum_{i=1}^{n} \Lambda(T_{i,K_i}),$$

after omitting the parts independent of Λ, where $T_{i,0}=0$ and $N_i(0)=0$. Let $s_1<\cdots<s_m$ be defined as above. Also let $w_l=\sum_{i=1}^n I(T_{i,K_i}=s_l)$ for $l=1,\ldots,m$ as before, and

$$\tilde{N}_{l,l'}=\sum_{i=1}^{n}\sum_{j=1}^{K_i}\big(N_i(T_{i,j})-N_i(T_{i,j-1})\big)I(T_{i,j}=s_l,T_{i,j-1}=s_{l'}),$$

for $0\le l'<l\le m$. Then the log-likelihood function $l_n(\Lambda)$ can be rewritten as

$$l_n(\Lambda)=\sum_{l'=0}^{m-1}\sum_{l=l'+1}^{m}\tilde{N}_{l,l'}\log\big[\Lambda(s_l)-\Lambda(s_{l'})\big]-\sum_{l=1}^{m}w_l\Lambda(s_l). \tag{3}$$

As the NPMPLE, we can define the nonparametric maximum likelihood estimate (NPMLE) of Λ to be the non-decreasing, non-negative step function with possible jumps only at the s_l's that maximizes $l_n(\Lambda)$.

It can be easily seen that the maximization of the log-likelihood function $l_n(\Lambda)$ over $\Lambda(t)$ or the m-dimensional parameter vector $\Lambda=(\Lambda_1,\ldots,\Lambda_m)$ with $\Lambda_1\le\cdots\le\Lambda_m$ does not have a closed-form solution. Both Wellner and Zhang (2000) and Hu et al. (2009a) studied this maximization problem. The former gave a modified iterative convex minorant (MICM) algorithm, while the latter proposed a simpler and faster alternative algorithm. The former also investigated the asymptotic properties of both NPMPLE and NPMLE. In particular, they derived the asymptotic distribution of the NPMPLE, which can be used to construct confidence bands of the estimate. Sun and Kalbfleisch (1995) also provided some discussion of the construction of the confidence bands for the IRE.

In the case of current status data, it is easy to see that the pseudo-log-likelihood function and the log-likelihood function given in (2) and (3) become the same and have the form

$$l_n^{\text{ps}}(\Lambda)=l_n(\Lambda)=\sum_{i=1}^{n}\big\{N_i(T_{i,1})\log\big(\Lambda(T_{i,1})\big)-\Lambda(T_{i,1})\big\}.$$

That is, the NPMPLE and NPMLE are identical. For the determination of the two estimates and also the IRE, one can use the algorithms described above and the simplification occurs only when all subjects have the same observation times. In this case, we have $m=1$ and the value of all three estimates at $s_1=T_{i,1}$ is given by the sample mean of the $N_i(T_{i,1})$'s.

In comparing the IRE and NPMLE of $\Lambda(t)$, it is easy to see that the latter could be more efficient than the former. Wellner and Zhang (2000) studied this by simulation and suggested that this is true for both non-homogeneous Poisson processes and some other counting processes. A disadvantage of the latter is that its determination is much more involved in terms of computation and requires much more computing time than that of the IRE. In general, the IRE provides a general idea about the shape of the mean function $\Lambda(t)$, especially for the case in which the number of observations for each subject is small. The NPMLE should be used, for example, if the non-homogeneous Poisson assumption seems reasonable.

We remark that the focus of this paper is on the event that can occur repeatedly. A related case is that the event can occur only once and the corresponding literature is commonly referred to as failure time data analysis. In this case, $N(t)$ is a one-jump counting process and the type of data discussed here is usually called interval-censored data (Sun 2006). Although the estimates of the mean function discussed above can be applied to the interval-censored data in theory, there exist some estimation procedures specifically developed for interval-censored failure time data that one may prefer as the objectives of interest between the two fields often differ.

3.3 Some other estimation procedures

For estimation of the mean function $\Lambda(t)$, in addition to the approaches discussed above, several other procedures are available. Among these, one procedure, which is similar to the IRE, is to minimize the generalized least-squares function

$$\sum_{i=1}^{n}\sum_{j=1}^{K_i}\sum_{l=1}^{K_i} W(T_{i,j}, T_{i,l})\big\{N_i(T_{i,j}) - \Lambda(T_{i,j})\big\}\big\{N_i(T_{i,l}) - \Lambda(T_{i,l})\big\}$$

subject to the non-decreasing property of $\Lambda(t)$, where $(W(T_{i,j}, T_{i,l}))$ is a known $K_i \times K_i$ symmetric weight matrix. It is obvious that this can give a class of estimates depending on the choice of the weight matrix. It can be easily shown that by using the identity weight matrix, the procedure yields the IRE. In other words, the estimates defined above can be seen as generalizations of the IRE. Hu et al. (2009b) discussed this approach and suggested some other choices for the weight matrix. Furthermore, they showed using a simulation study that the generalized least-squares estimates defined above can have its efficiency close to that of the NPMLE for Poisson processes and better efficiency than the NPMLE for non-Poisson processes.

Note that both the IRE and the NPMLE of $\Lambda(t)$ discussed above are nonparametric in the sense that they make no assumption about the dependence of $\{N_i(T_{i,j}), j = 1, \ldots, K_i\}$. Sometimes, it may be reasonable to impose some structures on the dependence for estimation of $\Lambda(t)$. Zhang and Jamshidian (2003) considered this and assumed that, given a latent variable Z_i, $E[N_i(t)|Z_i] = Z_i\Lambda(t)$ and $\{N_i(T_{i,j}), j = 1, \ldots, K_i\}$ are independent, $i = 1, \ldots, n$. Furthermore, by assuming that the Z_i's follow a gamma distribution, they developed an EM-algorithm for the estimation of $\Lambda(t)$. As the NPMLE, this estimate could be more efficient than the IRE, but its determination is also much more involved numerically than that of the IRE. Also it may be difficult to verify in practice the assumed latent variable model.

Other authors who considered the estimation problem discussed here include Lu et al. (2007) and Hu and Lagakos (2007). The former studied both pseudo-likelihood and likelihood-based approaches when the mean function $\Lambda(t)$ can be approximated by the monotone cubic I-splines, while the latter investigated the problem for a general response process that includes the counting process as a special case. Note that all approaches described so far focus on estimating the mean function $\Lambda(t)$, for which one has to take into account the monotone property of $\Lambda(t)$. Sometimes as mentioned before, one may want to estimate the rate function or could estimate the mean function by first estimating the rate function. Thall and Lachin (1988) considered this and proposed to estimate the rate function by the empirical estimate and then to estimate $\Lambda(t)$ by the integral of the obtained rate function estimate.

4 Nonparametric comparison with panel count data

It is well known that treatment comparison is one of the common objectives in data analysis and in this section we will consider this problem in the context of panel count data with respect to mean functions. Consider an event history study giving only panel count data, and let the $N_i(t)$'s, K_i and $T_{i,j}$'s be defined as in the last section. Also, let $\Lambda_i(t) = E\{N_i(t)\}$, $i = 1, \dots, n$, and suppose now our goal is to test the null hypothesis $H_0 : \Lambda_1(t) = \cdots = \Lambda_n(t)$. That is, all subjects in the study have identical mean functions. In the following, we will first discuss the two-sample situation. That is, all subjects in the study come from two different treatment groups. We will subsequently discuss the situation when there are k different treatment groups.

4.1 Two-sample comparison

In this subsection, we will discuss several procedures for comparing two populations in terms of the mean functions of the underlying counting processes. For this, suppose that all subjects come from two different populations or groups and define Z_i to be the binary (0 or 1) group indicator for subject i, $i = 1, \dots, n$. Let $\hat{\Lambda}_I(t)$ denote the IRE of the $\Lambda_i(t)$'s under the null hypothesis H_0. To test H_0, motivated by the t-test statistic, Sun and Kalbfleisch (1993) and Sun and Fang (2003) proposed to apply the test statistic

$$U_{\mathrm{SF}} = \sum_{i=1}^{n} Z_i \sum_{j=1}^{K_i} \left\{ N_i(T_{i,j}) - \hat{\Lambda}_I(T_{i,j}) \right\}, \tag{4}$$

representing the summation of the differences between the observed numbers of the event of interest and the estimated numbers of the event over the group with $Z_i = 1$. It is easy to see that if all subjects have only one observation at the same time point t_0 ($K_i = 1$, $T_{i,1} = t_0$), we have

$$U_{\mathrm{SF}} = \sum_{i:Z_i=1} N_i(t_0) - \frac{\sum_{i=1}^{n} Z_i}{n} \sum_{i=1}^{n} N_i(t_0).$$

Let $\hat{\Lambda}_I^{(u)}(t)$, $\{s_l^{(u)}\}$ and $\{w_l^{(u)}\}$ be defined as $\hat{\Lambda}_I(t)$, $\{s_l\}$ and $\{w_l\}$ in the definition of the IRE, but based only on subjects with $Z_i = u$, $u = 0, 1$. Then, the test statistic U_{SF} in (4) can be expressed as

$$U_{\mathrm{SF}} = \int w^{(1)}(t) \left[\hat{\Lambda}_I^{(1)}(t) - \hat{\Lambda}_I(t) \right] d\bar{N}^{(1)}(t), \tag{5}$$

where $w^{(1)}(t)$ is a step function with jumps only at the $s_l^{(1)}$'s, $w^{(1)}(s_l^{(1)}) = w_l^{(1)}$ and $\bar{N}^{(1)}(t) = \sum_l I(t \geq s_l^{(1)})$. That is, U_{SF} is the integrated weighted difference between an individual group estimator $\hat{\Lambda}_I^{(1)}(t)$ and the overall estimator $\hat{\Lambda}_I(t)$. Sun and Fang (2003) showed that under some regularity conditions and H_0, one can approximate the distribution of $n^{-1/2}U_{\mathrm{SF}}$ by the normal distribution with mean zero and variance

$$\hat{\sigma}_{\mathrm{SF}}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(Z_i - \bar{Z})\sum_{j=1}^{K_i}\left\{N_i(T_{i,j}) - \hat{\Lambda}_I(T_{i,j})\right\}\right]^2,$$

where $\bar{Z} = \sum_{i=1}^{n} Z_i/n$. Hence one can test the hypothesis H_0 using the statistic $T_{\mathrm{SF}} = U_{\mathrm{SF}}/(n^{1/2}\hat{\sigma}_{\mathrm{SF}})$ based on the standard normal distribution.

To apply the test procedure described above, one requirement is that the treatment or group indicators Z_i's can be regarded as independent and identically distributed random variables. Of course, this may not hold in practice. To relax this, Park et al. (2007) presented an alternative class of nonparametric tests also based on the IRE. Let Λ_1 and Λ_2 be the common mean functions of two groups, respectively. To test the hypothesis $H_0 : \Lambda_1(t) = \Lambda_2(t)$, let $\hat{\Lambda}_1^{\mathrm{ps}}$ and $\hat{\Lambda}_2^{\mathrm{ps}}$ be the IREs or NPMPLEs of Λ_1 and Λ_2 based on the data from the subjects in each individual group, respectively. Motivated by the idea commonly used in survival studies (Pepe and Fleming 1989), Park et al. (2007) suggested to use the statistic

$$U_{\mathrm{PSZ}} = \sqrt{\frac{n_1 n_2}{n}}\int_0^{\tau} \xi(t)\left\{\hat{\Lambda}_1^{\mathrm{ps}}(t) - \hat{\Lambda}_2^{\mathrm{ps}}(t)\right\} dG_n(t). \tag{6}$$

Here n_1 and n_2 are the numbers of subjects in the two groups with $n_1 + n_2 = n$, τ denotes the largest observation time, $\xi(t)$ is a bounded weight process, and

$$G_n(t) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i} I(T_{i,j} \le t).$$

The statistic U_{PSZ} is the integrated weighted difference between $\hat{\Lambda}_1^{\mathrm{ps}}$ and $\hat{\Lambda}_2^{\mathrm{ps}}$ and is sensitive especially to stochastically ordered mean functions. Statistics similar to U_{PSZ} are commonly used in survival analysis. For two-sample survival comparison with right-censored data, for example, Pepe and Fleming (1989) proposed some test statistics that have the same form as U_{PSZ} with the estimated survival functions in the place of $\hat{\Lambda}_1^{\mathrm{ps}}$ and $\hat{\Lambda}_2^{\mathrm{ps}}$. Petroni and Wolfe (1994) and Zhang et al. (2001) used similar statistics for the same comparison problem based on interval-censored data. Also it can be easily shown that the test statistic U_{PSZ} can be rewritten as

$$U_{\mathrm{PSZ}} = \sqrt{\frac{n_1 n_2}{n^3}}\sum_{i=1}^{n}\sum_{j=1}^{K_i} \xi(T_{i,j})\left\{\hat{\Lambda}_1^{\mathrm{ps}}(T_{i,j}) - \hat{\Lambda}_2^{\mathrm{ps}}(T_{i,j})\right\}.$$

That is, U_{PSZ} is a Wilcoxon-type statistic. Similar procedures are often used in the analysis of repeated measurement data (Davis and Wei 1988).

For the selection of the weight process $\xi(t)$ in U_{PSZ}, a simple and natural choice is $\xi_1(t) = 1$. Another natural choice is $\xi_2(t) = Y(t) = \sum_{i=1}^{n} I(t \le T_{i,K_i})/n$, in which case, the weights are proportional to the number of subjects under observation. One could also use

$$\xi_3(t) = \frac{Y_1(t)Y_2(t)}{Y(t)},$$

where $Y_1(t)$ and $Y_2(t)$ are defined as $Y(t)$ with the summation being over subjects only in each of the two groups, respectively. Weight processes similar to ξ_3 are commonly used when recurrent event data are observed (Andersen et al. 1993). Under some regularity conditions and H_0, Park et al. (2007) showed that U_{PSZ} has an asymptotic normal distribution with mean zero and variance that can be consistently estimated by

$$\hat{\sigma}^2_{\mathrm{PSZ}} = \frac{n_2}{n}\tilde{\sigma}_1^2 + \frac{n_1}{n}\tilde{\sigma}_2^2,$$

where

$$\tilde{\sigma}_l^2 = \frac{1}{n_l}\sum_{i\in S_l}\left[\sum_{j=1}^{K_i}\xi(T_{i,j})\left\{N_i(T_{i,j}) - \hat{\Lambda}_l^{\mathrm{ps}}(T_{i,j})\right\}\right]^2,$$

with S_l denoting the set of indices for subjects in group l, $l = 1, 2$. Thus, the test of the hypothesis H_0 can be carried out using the statistic $T_{\mathrm{PSZ}} = U_{\mathrm{PSZ}}/\hat{\sigma}_{\mathrm{PSZ}}$ based on the standard normal distribution.

As mentioned above, the statistic U_{PSZ} represents the integrated weighted, general difference between $\hat{\Lambda}_1^{\mathrm{ps}}$ and $\hat{\Lambda}_2^{\mathrm{ps}}$. In practice, one may be more interested in different types of the difference between $\Lambda_1(t)$ and $\Lambda_2(t)$ such as the absolute difference. In this case, one can construct the alternative test statistic

$$\int_0^\tau \xi(t)\left\{\hat{\Lambda}_1^{\mathrm{ps}}(t) - \hat{\Lambda}_2^{\mathrm{ps}}(t)\right\}^2 dG_n(t)$$

or

$$\int_0^\tau \xi(t)\left|\hat{\Lambda}_1^{\mathrm{ps}}(t) - \hat{\Lambda}_2^{\mathrm{ps}}(t)\right| dG_n(t)$$

for testing H_0. Of course, one then needs to derive the asymptotic null distributions for these statistics.

Note that in the test procedures discussed above, the IRE or NPMPLE was employed for the estimation of the mean functions. In practice, one can develop similar test procedures by replacing the IRE with the NPMLE. Among others, Balakrishnan and Zhao (2010a) considered this approach and proposed the following test statistic:

$$U_{\mathrm{BZ}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i\left[\sum_{j=1}^{K_i-1}\hat{\Lambda}(T_{i,j})\left\{\frac{\Delta N_i(T_{i,j+1})}{\Delta\hat{\Lambda}(T_{i,j+1})} - \frac{\Delta N_i(T_{i,j})}{\Delta\hat{\Lambda}(T_{i,j})}\right\}\right.$$
$$\left. + \hat{\Lambda}(T_{i,K_i})\left\{1 - \frac{\Delta N_i(T_{i,K_i})}{\Delta\hat{\Lambda}(T_{i,K_i})}\right\}\right].$$

In the above, $\Delta\Lambda(T_{i,j}) = \Lambda(T_{i,j}) - \Lambda(T_{i,j-1})$, $\Delta N(T_{i,j}) = N(T_{i,j}) - N(T_{i,j-1})$ and $\hat{\Lambda}(t)$ denotes the NPMLE of the common mean function $\Lambda(t)$ under the hypothesis H_0. Under some regularity conditions and H_0, they showed that U_{BZ} has an asymptotic normal distribution with mean zero and variance

$$\sigma_{\mathrm{BZ}}^2 = \mathrm{E}\Bigg[\big(Z - E(Z)\big)\Bigg\{\sum_{j=1}^{K-1} \Lambda_0(T_{1,j})\left(\frac{\Delta N(T_{1,j+1})}{\Delta\Lambda_0(T_{1,j+1})} - \frac{\Delta N(T_{1,j})}{\Delta\Lambda_0(T_{1,j})}\right) + \Lambda_0(T_{1,K})\left(1 - \frac{\Delta N(T_{1,K})}{\Delta\Lambda_0(T_{1,K})}\right)\Bigg\}\Bigg]^2.$$

Also they gave a consistent estimate of σ_{BZ}^2 as

$$\hat{\sigma}_{\mathrm{BZ}}^2 = \frac{1}{n}\sum_{i=1}^{n}\Bigg[(Z_i - \bar{Z})\Bigg\{\sum_{j=1}^{K_i-1} \hat{\Lambda}(T_{i,j})\left(\frac{\Delta N_i(T_{i,j+1})}{\Delta\hat{\Lambda}(T_{i,j+1})} - \frac{\Delta N_i(T_{i,j})}{\Delta\hat{\Lambda}(T_{i,j})}\right) + \hat{\Lambda}(T_{i,K_i})\left(1 - \frac{\Delta N_i(T_{i,K_i})}{\Delta\hat{\Lambda}(T_{i,K_i})}\right)\Bigg\}\Bigg]^2,$$

where $\bar{Z} = \sum_{i=1}^{n} Z_i/n$. Hence one can perform the testing of the hypothesis H_0 by using the statistic $T_{\mathrm{BZ}} = U_{\mathrm{BZ}}/\hat{\sigma}_{\mathrm{BZ}}$ based on the standard normal distribution.

4.2 k-sample comparison

Now we consider the more general treatment comparison problem in which the n study subjects come from k different populations or treatment groups. In the following, we will let $\Lambda_l(t)$ to denote the common mean function for the subjects in the lth group and n_l the number of subjects in the group, $l = 1, \ldots, k$, where $n_1 + \cdots + n_k = n$. Suppose that one is interested in testing the hypothesis $H_0 : \Lambda_1(t) = \cdots = \Lambda_k(t)$. For the problem, we will first discuss several procedures that can be regarded as generalization of the approaches considered in the previous section for the two-sample comparison and were developed based on the IRE or NPMPLE. A few other methods based on the NPMLE will then be presented.

4.2.1 NPMPLE-based tests procedures

Let $\hat{\Lambda}_l^{\mathrm{ps}}(t)$ denote the IRE or NPMPLE of the mean function $\Lambda_l(t)$ based on the subjects only in the lth group. To test H_0, it is natural to generalize the test statistic defined in (6) as

$$\mathbf{U}_Z = (U_{1,2}, U_{1,3}, \ldots, U_{1,k})^T,$$

where

$$U_{1,l} = \left(\frac{n_1 n_l}{n}\right)^{1/2} \int_0^{\tau} \eta(t)\big\{\hat{\Lambda}_1^{\mathrm{ps}}(t) - \hat{\Lambda}_l^{\mathrm{ps}}(t)\big\}\, dG_n(t)$$

with $G_n(t) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{K_i} I(T_{i,j} \le t)$ as before and $\eta(t)$ being a bounded weight process.

The statistic $\mathbf{U}_Z$ compares each mean function $\Lambda_l(t)$ $(l > 1)$ to the same mean function $\Lambda_1(t)$. For the selection of the weight function $\eta(t)$, Zhang (2006) suggested to use $\eta(t) = 1$, $\eta(t) = n^{-1} \sum_{i=1}^{n} I(t \le T_{i,K_i})$ or $\eta(t) = \sum_{i=1}^{n} I(t > T_{i,K_i})$. Also he showed that under some regularity conditions and the hypothesis H_0, one can approximate the distribution of $\mathbf{U}_Z$ by the multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix

$$\hat{\Sigma}_Z = \tilde{\sigma}_Z^2 \begin{bmatrix} \frac{n_1+n_2}{n} & (\frac{n_2 n_3}{n^2})^{1/2} & \cdots & (\frac{n_2 n_k}{n^2})^{1/2} \\ (\frac{n_2 n_3}{n^2})^{1/2} & \frac{n_1+n_3}{n} & (\frac{n_3 n_4}{n^2})^{1/2} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ (\frac{n_2 n_k}{n^2})^{1/2} & (\frac{n_3 n_k}{n^2})^{1/2} & \cdots & \frac{n_1+n_k}{n} \end{bmatrix},$$

where

$$\tilde{\sigma}_Z^2 = \frac{1}{n} \sum_{i=1}^{n} \left[\sum_{j=1}^{K_i} \eta(T_{i,j}) \left\{ N_i(T_{i,j}) - \hat{\Lambda}^{\mathrm{ps}}(T_{i,j}) \right\} \right]^2,$$

with $\hat{\Lambda}^{\mathrm{ps}}$ being the NPMPLE (IRE) of the common mean function under H_0 based on all samples combined together. Thus the test of the null hypothesis H_0 can be carried out by using $T_Z = \mathbf{U}_Z^T \hat{\Sigma}_Z \mathbf{U}_Z$ which has asymptotically a central χ^2-distribution with $(k-1)$ degrees of freedom.

For the test statistic given above, one requirement is that the weight function is the same for each component. To relax this, following Zhang (2006), Balakrishnan and Zhao (2010b) proposed a class of test statistics $\mathbf{U}_{\mathrm{BZ}}^{\mathrm{ps}} = (U_2^{\mathrm{ps}}, \ldots, U_k^{\mathrm{ps}})^T$ with

$$U_l^{\mathrm{ps}} = \sqrt{n} \int_0^{\tau} W_l(t) \left\{ \hat{\Lambda}_1^{\mathrm{ps}}(t) - \hat{\Lambda}_l^{\mathrm{ps}}(t) \right\} dG_n(t), \quad l = 2, \ldots, k. \tag{7}$$

Furthermore, they established its asymptotic normality under H_0. Note that both test statistics $\mathbf{U}_Z$ and $\mathbf{U}_{\mathrm{BZ}}^{\mathrm{ps}}$ used group 1 or an individual group as the baseline group for comparison. Instead, one can compare each individual group to the overall average group, which is commonly used in, for example, failure time data analysis (Pepe and Fleming 1989; Cook et al. 1996; Zhang et al. 2001). More specifically, for testing the hypothesis H_0, one can apply the statistic $\mathbf{V}_{\mathrm{BZ}}^{\mathrm{ps}} = (V_1^{\mathrm{ps}}, \ldots, V_k^{\mathrm{ps}})^T$, where

$$V_l^{\mathrm{ps}} = \sqrt{n} \int_0^{\tau} W_l(t) \left\{ \hat{\Lambda}^{\mathrm{ps}}(t) - \hat{\Lambda}_l^{\mathrm{ps}}(t) \right\} dG_n(t), \quad l = 1, \ldots, k. \tag{8}$$

Balakrishnan and Zhao (2010b) studied this statistic and derived its asymptotic normality under some regularity conditions and H_0.

With respect to the selection of the weight process $W_l(t)$, a simple and natural choice is $W_l^{(1)}(t) = 1, l = 1, \ldots, k$. Another natural choice is $W_l^{(2)}(t) = Y(t) = \sum_{i=1}^{n} I(t \le T_{i,k_i})/n, l = 1, \ldots, k$, and in this case the weights are proportional to the number of subjects under observation. Also one can use

$$W_l^{(3)}(t) = g\big(Y_l(t), Y(t)\big),$$

where g is a fixed function and $Y_l(t)$ is defined as $Y(t)$ with the summation being only over subjects in the lth group.

4.2.2 NPMLE-based test procedures

As discussed for the two-sample comparison problem, to test H_0, one can easily construct some test statistics similar to those described in the previous subsection by simply replacing the NPMPLE with the NPMLE. In this subsection, we will introduce several other available test procedures based on the NPMLE. To this end, in this subsection, we will let $\hat{\Lambda}_l(t)$ to denote the NPMLE of the mean function $\Lambda_l(t)$ based on the subjects only in group l and $\hat{\Lambda}(t)$ the NPMLE of the common mean function based on the data from all n subjects. Also for subject i, define $\mathbf{Z}_i$ to be the k-dimensional vector of treatment or group indicators whose lth element is equal to one if the subject is from group l and zero otherwise.

To test the hypothesis $H_0 : \Lambda_1(t) = \cdots = \Lambda_k(t)$, similar to the test statistic U_{BZ} for the two-sample problem, one could use the test statistic

$$\mathbf{U} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{Z}_i \Bigg[\sum_{j=1}^{K_i - 1} \hat{\Lambda}(T_{i,j}) \left\{ \frac{\Delta N_i(T_{i,j+1})}{\Delta \hat{\Lambda}(T_{i,j+1})} - \frac{\Delta N_i(T_{i,j})}{\Delta \hat{\Lambda}(T_{i,j})} \right\} + \hat{\Lambda}(T_{i,K_i}) \left\{ 1 - \frac{\Delta N_i(T_{i,K_i})}{\Delta \hat{\Lambda}(T_{i,K_i})} \right\} \Bigg].$$

Note that the only difference between U_{BZ} and $\mathbf{U}$ defined above is that $\mathbf{Z}_i$ is a vector here, while Z_i is scaler in U_{BZ}. As with U_{BZ}, one can prove that under some regularity conditions and H_0, $\mathbf{U}$ has an asymptotic normal distribution with mean zero (Balakrishnan and Zhao 2010a). It should be noted that the asymptotic covariance matrix has only the rank of $(k-1)$ as the summation of all components of $\mathbf{U}$ is equal to 0.

In addition to the test statistic $\mathbf{U}$ described above, for testing H_0, Balakrishnan and Zhao (2009) proposed to use the test statistics $\mathbf{U}_{BZ1} = (U_2, \ldots, U_k)^T$ with

$$\begin{aligned} U_l = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Bigg[\sum_{j=1}^{K_i - 1} & W_l(T_{i,j}) \hat{\Lambda}(T_{i,j}) \\ & \times \left\{ \left(\frac{\Delta \hat{\Lambda}_1(T_{i,j+1})}{\Delta \hat{\Lambda}(T_{i,j+1})} - \frac{\Delta \hat{\Lambda}_1(T_{i,j})}{\Delta \hat{\Lambda}(T_{i,j})} \right) \right. \\ & \quad \left. - \left(\frac{\Delta \hat{\Lambda}_l(T_{i,j+1})}{\Delta \hat{\Lambda}(T_{i,j+1})} - \frac{\Delta \hat{\Lambda}_l(T_{i,j})}{\Delta \hat{\Lambda}(T_{i,j})} \right) \right\} \\ & \quad + W_l(T_{i,K_i}) \hat{\Lambda}(T_{i,K_i}) \end{aligned}$$

$$\times\left\{\left(1-\frac{\Delta\hat{\Lambda}_1(T_{i,K_i})}{\Delta\hat{\Lambda}(T_{i,K_i})}\right)-\left(1-\frac{\Delta\hat{\Lambda}_l(T_{i,K_i})}{\Delta\hat{\Lambda}(T_{i,K_i})}\right)\right\}\Bigg] \tag{9}$$

for $l = 2, \dots, k$. Here as before, the $W_l(t)$'s are some bounded weight processes. Furthermore, they showed that under some regularity conditions and H_0, $\mathbf{U}_{\mathrm{BZ1}}$ has an asymptotic normal distributions with mean vector $\mathbf{0}$ and covariance matrixes $\mathbf{\Sigma}_{\mathrm{BZ1}}$ that can be consistently estimated by

$$\hat{\mathbf{\Sigma}}_{\mathrm{BZ1}} = \mathbf{H}\,\mathbf{diag}\big(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_k^2\big)\mathbf{H}'.$$

Here

$$\mathbf{H} = \begin{pmatrix} -\sqrt{\frac{n}{n_1}} & \sqrt{\frac{n}{n_2}} & 0 & \cdots & 0 \\ -\sqrt{\frac{n}{n_1}} & 0 & \sqrt{\frac{n}{n_3}} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -\sqrt{\frac{n}{n_1}} & 0 & 0 & \cdots & \sqrt{\frac{n}{n_k}} \end{pmatrix}$$

and

$$\begin{aligned}\hat{\sigma}_l^2 = \frac{1}{n}\sum_{i=1}^{n}\Bigg[&\sum_{j=1}^{K_i-1} W_l(T_{i,j})\hat{\Lambda}(T_{i,j}) \\ &\times\left\{\frac{\Delta N_i(T_{i,j+1})}{\Delta\hat{\Lambda}(T_{i,j+1})}-\frac{\Delta N_i(T_{i,j})}{\Delta\hat{\Lambda}(T_{i,j})}\right\} \\ &+ W_l(T_{i,K_i})\hat{\Lambda}(T_{i,K_i})\left\{1-\frac{\Delta N_i(T_{i,K_i})}{\Delta\hat{\Lambda}(T_{i,K_i})}\right\}\Bigg]^2.\end{aligned}$$

Thus one can perform the testing of the hypothesis H_0 based on $T_{\mathrm{BZ1}} = \mathbf{U}_{\mathrm{BZ1}}^T\hat{\mathbf{\Sigma}}_{\mathrm{BZ1}}^{-1}\mathbf{U}_{\mathrm{BZ1}}$ with the central χ^2-distribution with $(k-1)$ degrees of freedom.

Similar to $\mathbf{U}_{\mathrm{BZ1}}$, one can consider the statistics $\mathbf{V}_{\mathrm{BZ}} = (V_1, \dots, V_k)^T$, where

$$\begin{aligned}V_l = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg[&\sum_{j=1}^{K_i-1} W_l(T_{i,j})\hat{\Lambda}(T_{i,j}) \\ &\times\left\{\frac{\Delta\hat{\Lambda}_l(T_{i,j+1})}{\Delta\hat{\Lambda}(T_{i,j+1})}-\frac{\Delta\hat{\Lambda}_l(T_{i,j})}{\Delta\hat{\Lambda}(T_{i,j})}\right\} \\ &+ W_l(T_{i,K_i})\hat{\Lambda}(T_{i,K_i})\left\{1-\frac{\Delta\hat{\Lambda}_l(T_{i,K_i})}{\Delta\hat{\Lambda}(T_{i,K_i})}\right\}\Bigg]\end{aligned} \tag{10}$$

for $l = 1, \dots, k$ (Balakrishnan and Zhao 2009). The statistic V_l represents the integrated weighted difference between the rates of increase of $\hat{\Lambda}$ and $\hat{\Lambda}_l$ over the observation period and U_l has a similar interpretation. For the selection of the weight process $W_l(t)$, in addition to the choices discussed above, one could also use

$$Y_l(t), \frac{Y_l(t)}{Y(t)}, \frac{Y_1(t)Y_l(t)}{Y(t)},$$

or

$$1 - Y_l(t), \frac{1 - Y_l(t)}{1 - Y(t)}, \frac{(1 - Y_1(t))(1 - Y_l(t))}{1 - Y(t)}.$$

It is easy to see that all test procedures described above base the comparison on the estimated mean functions and they all are nonparametric or distribution-free. In some situations, it may be reasonable to specify some models for the underlying counting processes and in this case, the comparison can be performed based on regression techniques that will be briefly discussed below. As another alternative, for large data sets, one could also apply the grouping comparison procedure proposed by Thall and Lachin (1988). This procedure first partitions the entire study period into several fixed, consecutive intervals and transforms the observed numbers of events on each subject over each interval into vectors of variables. The comparison is then conducted by using the procedure given in Wei and Lachin (1984) for multivariate non-negative-valued random vectors. A shortcoming of this procedure is that the test result could depend on the selection of the intervals.

5 Some numerical results and illustrative examples

In this section, we will present some numerical results to compare and illustrate the inferential procedures discussed in the previous sections. First some simulation results will be given with the focus on the comparison of the NPMPLE-based and NPMLE-based test procedures and the inferential procedures will then be illustrated by using two sets of real panel count data. As commented before, the major advantage of NPMPLE-based test procedures is that they can be easily implemented, while the NPMLE-based test procedures are expected to be more efficient. Thus one question of interest is how much efficiency or power that one may lose by using the former ones in addition to the assessment of the overall performance of each procedure.

5.1 Some simulation results

To evaluate the test procedures discussed above, following Balakrishnan and Zhao (2009), we will focus on the two-sample comparison problem and consider the four test statistics T_{SF}, T_{PSZ}, T_Z and T_{BZ} given in the previous section. Note that for $k = 2$, we have $T_Z = T_{\text{PSZ}}$ and $T_{\text{BZ}} = T_{\text{BZ1}}$. The test procedures based on T_{SF}, T_{PSZ} and T_Z make use of the IRE or NPMPLE, while the one based on T_{BZ} relies on the NPMLE. Balakrishnan and Zhao (2009) investigated the performance of these procedures for the study with the same numbers of subjects in both samples or the balanced design. Here we will consider the situation where the sample sizes differ between the two samples.

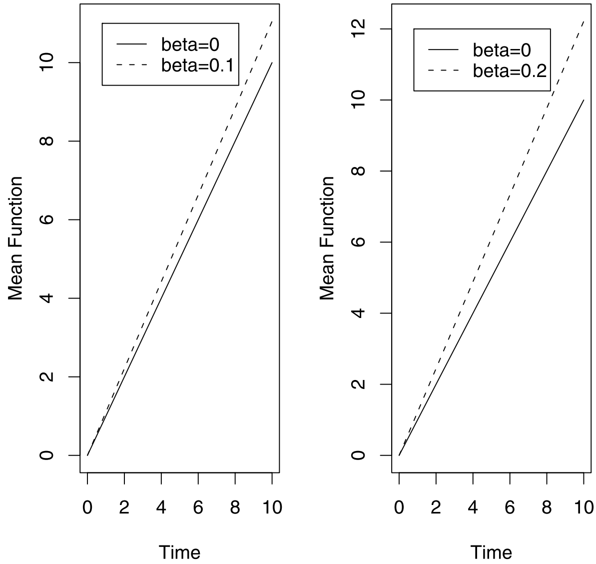

Fig. 1 True mean functions for Case 1 with $\nu = 1$ and $\beta = 0.1, 0.2$

To generate panel count data, let $\{\nu_i, i = 1, \ldots, n\}$ be independent and identically distributed random variables and assume that, given ν_i, $N_i(t)$ is a Poisson process with the mean function $\Lambda_i(t|\nu_i) = E(N_i(t)|\nu_i)$. We first generate the number of observation times K_i from the uniform distribution $U\{1, \ldots, 10\}$ and then given K_i, we generate the observation times $T_{i,j}$'s as the order statistics of K_i random variables from $U\{1, \ldots, 10\}$ for simplicity. Also let S_l denote the set of indices for subjects in group l as before, $l = 1, 2$. For the two mean functions, we will consider the following two cases:

Case 1. $\Lambda_i(t|\nu_i) = \nu_i t$ for $i \in S_1$, $\Lambda_i(t|\nu_i) = \nu_i t \exp(\beta)$ for $i \in S_2$.

Case 2. $\Lambda_i(t|\nu_i) = \nu_i t$ for $i \in S_1$, $\Lambda_i(t|\nu_i) = \nu_i \sqrt{\beta t}$ for $i \in S_2$.

To give an idea about the shapes of the mean functions, Figs. 1 and 2 display them with $\nu_i = 1$ and different values of β. It can be seen that in Case 1, the two mean functions do not overlap, while they cross over in Case 2.

For the random variables ν_i's, two situations will be studied. One is that $\nu_i = 1$ for all i and the other is to assume that $\nu_i \sim \text{Gamma}(2, 1/2)$. For the former, the $N_i(t)$'s are Poisson processes and for the latter, the $N_i(t)$'s are mixed Poisson processes. To calculate the test statistics, we consider the following four weight processes:

$$W^{(1)}(t) = 1, \qquad W^{(2)}(t) = Y(t) = \frac{1}{n} \sum_{i=1}^{n} I(t \le T_{i,K_i}),$$

$$W^{(3)}(t) = \frac{Y_1(t) Y_2(t)}{Y(t)}, \quad \text{and} \quad W^{(4)}(t) = 1 - Y(t) = \frac{1}{n} \sum_{i=1}^{n} I(t > T_{i,K_i}).$$

Fig. 2 True mean functions for Case 2 with $\nu = 1$ and $\beta = 3, 5$

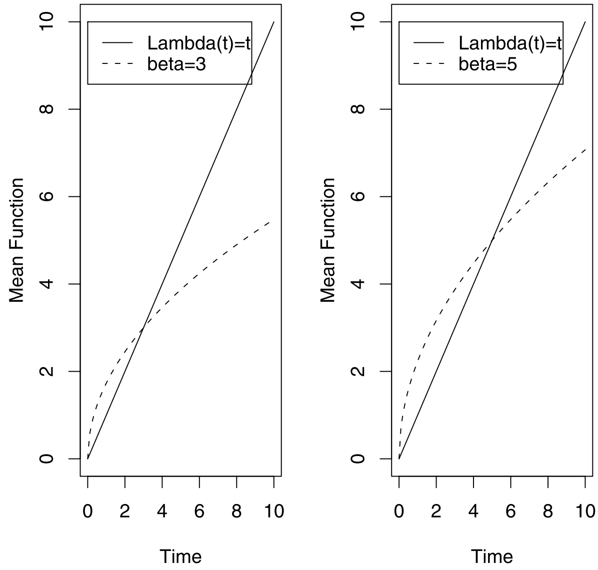

Table 1 Estimated size and power for Poisson processes in Case 1

β	T_{BZ}				T_{PSZ} & T_Z				T_{SF}
	$W^{(1)}$	$W^{(2)}$	$W^{(3)}$	$W^{(4)}$	$W^{(1)}$	$W^{(2)}$	$W^{(3)}$	$W^{(4)}$	
$n_1 = 40, n_2 = 60$									
0.0	0.049	0.048	0.047	0.047	0.048	0.050	0.049	0.046	0.045
0.1	0.348	0.233	0.233	0.162	0.248	0.236	0.235	0.289	0.252
0.2	0.850	0.659	0.652	0.428	0.716	0.688	0.688	0.762	0.718
0.3	0.997	0.954	0.951	0.762	0.969	0.959	0.960	0.985	0.970
$n_1 = 80, n_2 = 120$									
0.0	0.053	0.051	0.052	0.052	0.057	0.054	0.054	0.046	0.055
0.1	0.558	0.386	0.377	0.225	0.413	0.388	0.388	0.453	0.417
0.2	0.994	0.933	0.934	0.725	0.962	0.947	0.946	0.966	0.961
0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

The results given below are based on 1000 replications with the total sample size $n = 100$ or 200 and the sample size $n_1 = 40$ or 80 for the first sample.

Tables 1 and 2 give the estimated sizes and powers of the test procedures based on T_{SF}, T_{PSZ}, T_Z and T_{BZ} at significance level $\alpha = 0.05$ under Case 1 for the mean functions with different β values. The results in Table 1 correspond to the situation where the $N_i(t)$'s are Poisson processes and the results in Table 2 are for the mixed Poisson processes $N_i(t)$'s. In both tables, the first part is for the situation with the total sample size of 100 and the second part is for the situation with the total sample size of 200. It can be seen that all four procedures perform reasonably well and that the performance does not seem to depend on the weight process. In particular, as

Table 2 Estimated size and power for mixed Poisson processes in Case 1

β	T_{BZ}				T_{PSZ} & T_Z				T_{SF}
	$W^{(1)}$	$W^{(2)}$	$W^{(3)}$	$W^{(4)}$	$W^{(1)}$	$W^{(2)}$	$W^{(3)}$	$W^{(4)}$	
$n_1 = 40, n_2 = 60$									
0.0	0.050	0.057	0.053	0.050	0.049	0.047	0.048	0.048	0.045
0.1	0.109	0.090	0.090	0.107	0.069	0.069	0.068	0.080	0.090
0.2	0.246	0.213	0.213	0.174	0.154	0.153	0.155	0.160	0.191
0.3	0.447	0.411	0.414	0.334	0.309	0.302	0.302	0.297	0.357
$n_1 = 80, n_2 = 120$									
0.0	0.052	0.051	0.050	0.052	0.044	0.041	0.041	0.046	0.042
0.1	0.140	0.138	0.138	0.124	0.097	0.094	0.095	0.093	0.112
0.2	0.377	0.358	0.355	0.280	0.282	0.282	0.283	0.282	0.320
0.3	0.735	0.687	0.691	0.578	0.620	0.616	0.615	0.605	0.643

Table 3 Estimated power for Poisson processes in Case 2

β	T_{BZ}				T_{PSZ} & T_Z				T_{SF}
	$W^{(1)}$	$W^{(2)}$	$W^{(3)}$	$W^{(4)}$	$W^{(1)}$	$W^{(2)}$	$W^{(3)}$	$W^{(4)}$	
$n_1 = 40, n_2 = 60$									
3	1.000	0.677	0.634	1.000	0.953	0.894	0.893	1.000	0.952
5	0.974	0.077	0.085	1.000	0.182	0.098	0.099	0.882	0.169
8	0.116	0.722	0.725	0.994	0.419	0.560	0.558	0.070	0.429
$n_1 = 80, n_2 = 120$									
3	1.000	0.899	0.876	1.000	1.000	0.994	0.994	1.000	1.000
5	1.000	0.074	0.087	1.000	0.305	0.135	0.131	0.987	0.297
8	0.200	0.944	0.946	1.000	0.678	0.827	0.826	0.068	0.697

expected, the performance becomes better when the sample size increases and the NPMLE-based procedures show better power than the NPMPLE-based procedures for all situations considered here except for $W^{(j)}$ $(j = 2, 3, 4)$ in Case 1 with $\nu_i = 1$. Also the power decreases from the Poisson processes to the mixed Poisson processes, that is, when more variability exists.

The results obtained under Case 2 for the mean functions are presented in Tables 3 and 4 with all other set-up being the same as in Tables 1 and 2. Note that here we only presented the results on the estimated power. It is clear that when the underlying mean functions cross, the performance or selection of a test procedure is much more complicated. One difference between this and the situation considered in Tables 1 and 2 is that the NPMPLE-based procedures could have better power in some situations although the NPMLE-based procedures still perform better in most cases. Another difference is that the power can heavily depend on the choice of the weight process and the specific shapes of the mean functions. This will make the selection of the weight process and thus the test procedure more difficult since it may not be possi-

Table 4 Estimated power for mixed Poisson processes in Case 2

β	T_{BZ}				T_{PSZ} & T_Z				T_{SF}
	$W^{(1)}$	$W^{(2)}$	$W^{(3)}$	$W^{(4)}$	$W^{(1)}$	$W^{(2)}$	$W^{(3)}$	$W^{(4)}$	
$n_1 = 40, n_2 = 60$									
3	0.849	0.323	0.310	0.989	0.462	0.394	0.394	0.718	0.402
5	0.411	0.065	0.066	0.945	0.079	0.059	0.058	0.298	0.061
8	0.059	0.265	0.272	0.766	0.115	0.140	0.141	0.047	0.138
$n_1 = 80, n_2 = 120$									
3	0.993	0.491	0.476	1.000	0.708	0.604	0.601	0.957	0.668
5	0.669	0.063	0.064	0.999	0.104	0.062	0.061	0.472	0.084
8	0.083	0.471	0.484	0.968	0.179	0.237	0.237	0.061	0.205

ble to know the shapes of the true mean functions in general. Of course, for a given problem, one can simply try different weight processes but the interpretation of the results may not be easy in some situations as will be seen in the examples discussed in the next subsection. More comments on this are made below. Overall, the results here gave similar conclusions to those obtained in Balakrishnan and Zhao (2009).

5.2 Two illustrative examples

In this subsection, we discuss the analysis of two sets of panel count data to illustrate the nonparametric estimation approaches and the test procedures based on the statistics T_{SF}, T_{PSZ}, T_Z and T_{BZ}. The first example concerns the data arising from a floating gallstones study and then the second example is based on a bladder tumor study.

5.2.1 A floating gallstones study

Thall and Lachin (1988) described a follow-up study on the patients with floating gallstones. The data are given in Table 1 of Thall and Lachin (1988) and consist of the first year follow-up of the patients in two study groups, placebo (48) and high-dose chenodiol (65), from the National Cooperative Gallstone Study. Note that the original study consists of 916 patients who were randomized to placebo, low-dose, or high-dose groups and followed for up to two years. The observed data include the successive visit times in study weeks and the associated counts of episodes of nausea for patients in different treatment groups. During the study, patients were scheduled to return for clinical visits at 1, 2, 3, 6, 9, and 12 months. In reality, most of the patients visited about six times within the first year, but actual visit times differed from patient to patient. Some patients had only one visit and some had as many as nine visits. The problem of interest here is to estimate the occurrence rates of the nausea and compare the two treatment groups in terms of the incidence rates of nausea.

For the analysis, we will assume that all patients in the placebo group share the mean function $\Lambda_1(t)$, while the patients in the high-dose chenodiol group have the mean function $\Lambda_2(t)$. Figure 3 presents both NPMPLE and NPMLE of the two mean

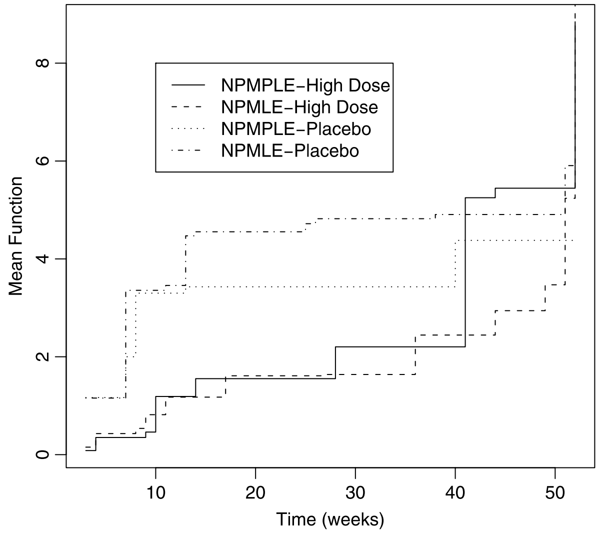

Fig. 3 The estimated mean functions for the floating gallstone study

Table 5 The p-values for testing no treatment effect for the floating gallstone data

Weight	T_{BZ} $W^{(1)}$	$W^{(2)}$	$W^{(3)}$	$W^{(4)}$	T_{PSZ} & T_{Z} $W^{(1)}$	$W^{(2)}$	$W^{(3)}$	$W^{(4)}$	T_{SF}
p-value	0.861	$\ll$0.001	$\ll$0.001	$\ll$0.001	0.454	0.417	0.413	0.891	0.143

functions and indicates that the occurrence rates of nausea seem to be different for the patients in the two groups in the middle of the year. Table 5 gives the test results obtained by applying the test procedures based on T_{SF}, T_{PSZ}, T_{Z} and T_{BZ} by using the same weight processes as those used in the simulation study for testing the null hypothesis $H_0 : \Lambda_1(t) = \Lambda_2(t)$. It can be seen that except the procedures based on T_{BZ} with weight processes $W^{(2)}$, $W^{(3)}$ and $W^{(4)}$, all other procedures suggest that there are no significant differences between the incidence rates of nausea of the patients in the two groups. Some general comments on this will be given below.

5.3 A bladder tumor study

Now we consider the panel count data that arose from the bladder tumor study conducted by the Veterans Administration Cooperative Urological Research Group (Byar et al. 1977) and given in Andrews and Herzberg (1985). The study is a randomized clinical trial and consists of the patients with superficial bladder tumors when they entered the trial. At the enrollment, they were assigned randomly to one of three treatments: placebo, thiotepa and pyridoxine. At subsequent follow-up visits, any tumors noticed were removed and treatment was continued. The study included 116 patients, of which there were 47 in placebo group, 38 in thiotepa group and 31 in pyridoxine. One of the objectives of the study is to estimate the tumor occurrence rates and to determine the treatment effects on the frequency of tumor recurrence.

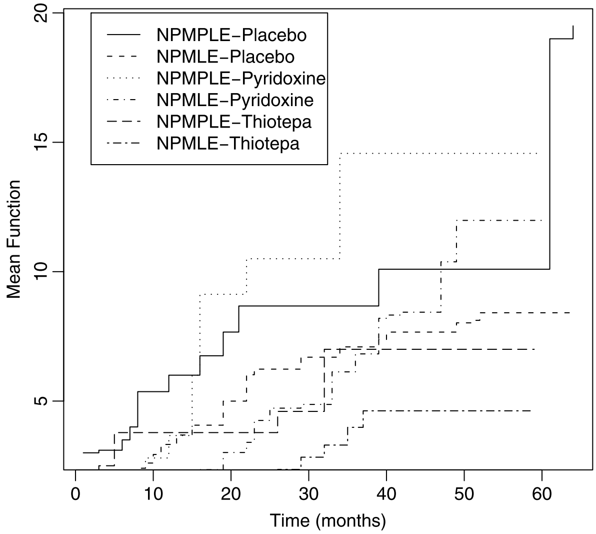

Fig. 4 The estimated mean functions for the bladder tumor study

Table 6 The p-values for testing no treatment effects for the bladder tumor data

	T_{BZ}			T_Z		
Weight	$W^{(1)}$	$W^{(2)}$	$W^{(4)}$	$W^{(1)}$	$W^{(2)}$	$W^{(4)}$
p-value	0.195	$<10^{-8}$	$<10^{-8}$	0.0851	0.1445	0.0840

Let $\Lambda_1(t)$, $\Lambda_2(t)$ and $\Lambda_3(t)$ be the mean functions corresponding to the three treatment groups: placebo, thiotepa and pyridoxine, respectively. First we calculated the NPMPLE and NPMLE of the three mean functions and these are displayed in Fig. 4. It can be seen that the mean functions are quite different from each other. To test $H_0: \ \Lambda_1(t) = \Lambda_2(t) = \Lambda_3(t)$, we applied the test procedures based on T_Z and T_{BZ} and the obtained p-values are given in Table 6 with the three weight processes considered in the simulation study. As in the floating gallstones example, the procedures based on T_{BZ} with weight processes $W^{(2)}$ and $W^{(4)}$ indicate that the tumor occurrence rates were quite different for the patients in the three treatment groups, while all other procedures suggest that there are no significant differences.

6 Estimation with multiple state panel count data

In the previous sections, the focus has been on panel count data concerning occurrence rates of certain recurrent events such as infections or hospitalizations. In practice, a different type of panel count data may occur concerning how often a subject stays in certain status and/or move from one status to another status and involving observations on study subjects only at discrete time points (Bartholomew 1983; Kalbfleisch and Lawless 1985; Singer and Spilerman 1976a, 1976b; Wasserman 1980). Such an example can be found in, for instance, Kalbfleisch and Lawless (1985) who discussed a survey study of public school students on their smoking status. In the study, the students were surveyed a few times and at each time, the information on their smoking status, including never smoked, currently a smoker, and has smoked but quit, was collected. Some of the questions of interest were how long the students stayed at each of the three states and what were the probabilities of moving from one state to another state. Gentleman et al. (1994) gave a similar example arising from an AIDS study. In this section, we discuss the analysis of these panel count data with the focus being on data arising from continuous-time finite state Markov models.

Consider a longitudinal study involving n independent subjects. Suppose that each subject in the study moves among k states, denoted by $1, \ldots, k$, following a continuous-time Markov chain. Let $X_r(t)$ be the state occupied at time t for subject r, $r = 1, \ldots, n$. For $0 \le s \le t$, let $P(s,t)$ be the common $k \times k$ transition probability matrix with entries

$$p_{ij}(s,t) = P\left\{X_r(t) = j | X_r(s) = i\right\},$$

for $i, j = 1, \ldots, k$. Define the transition intensities by

$$q_{ij}(t) = \lim_{\Delta t \to 0} \frac{p_{ij}(t, t + \Delta t)}{\Delta t}, \quad i \neq j$$

and

$$q_{ii}(t) = -\sum_{j \neq i} q_{ij}(t), \quad i = 1, \ldots, k.$$

Let $Q(t)$ be the $k \times k$ transition intensity matrix with entries $q_{ij}(t)$. It is well known that the process $X_r(t)$ can be specified by $Q(t)$. If $Q(t) = Q = (q_{ij})$ is independent of t, the Markov process is said to be time-homogeneous. In this case, the process is stationary and we have

$$P(t) = P(s, s + t) = P(0, t).$$

Also it is known (see Cox and Miller 1965) that

$$P(t) = e^{Qt} = \sum_{\tau=0}^{\infty} \frac{Q^\tau t^\tau}{\tau!}.$$

Suppose that $\{X_r(t)\}$ is time-homogeneous and $q_{ij}(\theta)$ depends on b functionally independent parameters $\theta_1, \ldots, \theta_b$, with $\theta = (\theta_1, \ldots, \theta_b)$ for each $i, j = 1, \ldots, k$. Assume that each study subject is observed at distinct times $t_0, t_1, \ldots, t_m$. Then conditional on the distribution of subjects at t_0, the likelihood function for θ is

$$L(\theta) = \prod_{l=1}^{m} \prod_{i,j=1}^{k} \left\{p_{ij}(w_l; \theta)\right\}^{n_{ijl}}, \tag{11}$$

where $w_l = t_l - t_{l-1}$, n_{ijl} denotes the number of subjects in state i at t_{l-1} and j at t_l, $l = 1, \ldots, m$. This yields the log-likelihood as

$$\log L(\theta) = \sum_{l=1}^{m} \sum_{i,j=1}^{k} n_{ijl} \log p_{ij}(w_l; \theta). \tag{12}$$

The maximum likelihood estimate, say $\hat{\theta}$, of θ can be obtained by maximizing (12). There exist various algorithms for finding $\hat{\theta}$. One approach is to utilize a numerical algorithm that requires no derivatives of $\log L(\theta)$; see Wasserman (1980). Kalbfleisch and Lawless (1985) presented a more efficient quasi-Newton procedure that uses first derivatives of $\log L(\theta)$. Here, it was assumed that all subjects are observed at the same times. For the general situation where the inspection times may be different over subjects, one can refer to Gentleman et al. (1994).

In the above discussion, it has been assumed that $X(t)$ is time-homogeneous. Of course, this may not be true in practice. To relax this, one can consider the Markov process $X(t)$ with the intensity matrix

$$Q(t) = Q_0 g(t; \lambda),$$

where Q_0 is a fixed intensity matrix with unknown entries (q_{ij}) and $g(t; \lambda)$ is a known function with an unknown parameter λ. For given λ, let $s = \int_0^t g(u; \lambda)\, du$ and define $Y(s) = X(t)$. Then, the process $\{Y(s) : 0 < s < \infty\}$ is a homogeneous Markov process with intensity matrix Q_0. Thus, for any given λ, we replace t_l by $s_l = \int_0^{t_l} g(u; \lambda)\, du$ and w_l by $w_l^* = s_l - s_{l-1}$. The parameters of Q_0 can be estimated by using the above method for given λ. In addition, the maximized log-likelihood can be obtained for that λ. By varying λ, this additional parameter can be estimated by observing the effect on the maximized log-likelihood. For modeling non-homogeneous Markov processes, another possibility is to use a piecewise transition intensity matrix (Kalbfleisch and Lawless 1985).

Also sometimes covariates may exist and one may be interested in the estimation of covariate effects. For this, suppose there exists a vector of p covariates $\mathbf{Z} = (Z_1, \ldots, Z_p)^T$ with $Z_1 = 1$. For given $\mathbf{Z}$, assume that the state process $X(t)$ is a homogeneous Markov chain with transition intensity matrix

$$Q(\mathbf{Z}) = \big(q_{ij}(\mathbf{Z})\big),$$

where

$$q_{ij}(\mathbf{Z}) = \exp\big(\beta_{ij}^T \mathbf{Z}\big), \quad i \neq j$$

and

$$q_{ii}(\mathbf{Z}) = -\sum_{j \neq i} q_{ij}(\mathbf{Z}).$$

Here $\beta_{ij} = (\beta_{1ij}, \ldots, \beta_{pij})^T$ is a vector of p regression parameters relating the instantaneous rate of transitions from state i to state j to the covariate $\mathbf{Z}$. Let r be the number of distinct covariate vectors in the sample and let these distinct covariates be denoted by $\mathbf{Z}_h = (Z_{1h}, \dots, Z_{ph})$ with $Z_{1h} = 1$, and let

$$Q_h = Q(\mathbf{Z}_h) = \big(q_{ij}(\mathbf{Z}_h)\big), \quad h = 1, \dots, r.$$

Let $n_{ijl}^{(h)}$ be the number of subjects with covariate values $\mathbf{Z}_h$ that are in state i at time t_{l-1} and in state j at time t_l. Then the log-likelihood function has the form

$$\log L(\beta) = \sum_{h=1}^{r} \sum_{l=1}^{m} \sum_{i,j=1}^{k} n_{ijl}^{(h)} \log p_{ij}(w_l; \mathbf{Z}_h), \tag{13}$$

where

$$P_h(t) = \exp(Q_h t) = \big(p_{ij}(t; \mathbf{Z}_h)\big)$$

and β denotes the vector of all parameters β_{ij} $(i \neq j)$ together. It is clear that the maximum likelihood estimate of β can be obtained by maximizing (13).

In practice, the goodness-of-fit of a model is often of interest. For the present situation, it can be assessed by comparing the observed transition frequencies n_{ijl}'s with expected frequencies $e_{ijl} = n_{i,l} \hat{p}_{ij}(w_l)$, where $n_{i,l} = \sum_{j=1}^{k} n_{ijl}$. A likelihood ratio, or asymptotically equivalent Pearson chi-squared statistic to test the fit of the Markov model is readily obtained by methods similar to those used in Markov chains (see, e.g., Anderson and Goodman 1957; Gentleman et al. 1994). If none of the $p_{ij}(w_l)$'s is allowed to be zero, the likelihood ratio statistic is

$$\Lambda = 2 \sum_{l=1}^{m} \sum_{i,j=1}^{k} n_{ijl} \log(n_{ijl}/e_{ijl}),$$

which is asymptotically (m fixed, $n \to \infty$) a chi-squared variate with $mk(k-1) - b$ degrees of freedom. The related Pearson statistic is

$$\chi^2 = \sum_{l=1}^{m} \sum_{i,j=1}^{k} (n_{ijl} - e_{ijl})^2 / e_{ijl}.$$

Kalbfleisch and Lawless (1985) illustrated these methods through the public student smoking data discussed above.

We remark that only time-homogeneous Markov models were discussed in this section. This is because there exists little literature on nonparametric estimation about heterogeneous Markov models under the framework considered here. Obvious nonparametric estimators for $p_{ij}(w_l)$ are defined through $n_{ijl}/n_{i,l}$ (e.g. Kalbfleisch and Lawless 1985, p. 868), but these do not incorporate the Markov information nor the multi-state model structure (i.e. the allowed transitions), and hence they are not optimal. Besides, they do not serve for the purpose of estimating the transition intensities along time. For the situation where the event of interest can occur only once or failure time data, some discussion on the nonparametric estimation can be found in, for example, Commenges (2002) and Meira-Machado et al. (2009). Also it is worth noting that the methodology discussed in the referred papers can be applied to situations when the recurrent event under study can occur only up to a finite number of times, say, g. In this case, $N(t)$ takes values of 0, 1, ..., g and one could define a g-state progressive model for the analysis.

7 Nonparametric estimation with dependent observation processes

As mentioned before, one main difference between recurrent event data and panel count data is that the latter involves an observation process that does not exist in the former. So far we have assumed that the observation process is independent of the underlying counting process of interest and thus one can make inference about the counting process conditional on the observation process as done in the previous sections. In practice, however, the observation process may depend on or contain information about the counting process. That is, the two processes may be related. For example, this can be the case if the recurrent event of interest is some type of markers for a disease under study and the subjects or the markers can be observed or measured only during their repeated hospitalization.

In this section, we will briefly discuss two latent variable or frailty modeling approaches for the analysis of panel count data when the underlying counting process of interest and the observation process may be related. One models the rate function, while the other models the mean function.

7.1 Rate function-based frailty model

Let Z be a non-negative latent variable with $E(Z) = 1$. Suppose that given Z, the event process $N(t)$ is a non-homogeneous Poisson process with the rate function

$$\lambda(t|Z) = Z\lambda_0(t), \quad t \in [0, \tau],$$

where $\lambda_0(t)$ is an unspecified function. Note that here for convenience, we still use $\lambda(t)$ for the rate function. Furthermore, it is assumed that given Z, the event process $N(t)$ is independent of the number of observations K and the observation times $\{T_1, \dots, T_K\}$. That is, the event process $N(t)$ and the observation process including K and $\{T_1, \dots, T_K\}$ are correlated only through the frailty variable Z. We will assume that the distribution of the frailty variable and the probability of the observation process given the frailty can be arbitrary.

Define

$$\Lambda_0(t) = \int_0^t \lambda_0(t)\, du.$$

Then the mean function of the event process is given by

$$E\big(N(t)\big) = E(Z)\Lambda_0(t) = \Lambda_0(t).$$

Consider an event history study involving n subjects. For each subject, let $N_i(t)$ be the underlying counting process that is observed only at distinct observation times

$\{T_{i,1}, \ldots, T_{i,K_i}\}$, where K_i is the number of observations, $i = 1, \ldots, n$. Assume that $\{N_i, K_i, T_{i,j}, j = 1, \ldots, K_i\}$ $(i = 1, \ldots, n)$ is a random sample of $\{N, K, T_j, j = 1, \ldots, K\}$ of size n. Let $N_{i,j} = N_i(T_{i,j})$. Suppose our goal here is to estimate $\Lambda_0(t)$ nonparametrically based on the observed data $\{K_i, T_{i,j}, N_{i,j}, j = 1, \ldots, K_i, i = 1, \ldots, n\}$.

Let $Y_i = T_{i,K_i}, m_i = N_i(Y_i)$ and $m_{ij} = N_{i,j} - N_{i,j-1}$. Define $F(t) = \Lambda_0(t)/\Lambda_0(\tau)$, $0 \le t \le \tau$. Then the conditional likelihood function, given Z_i, K_i, m_i and $\{T_{i,j}, j = 1, \ldots, K_i\}$, has the form

$$L \propto \prod_{i=1}^{n} \prod_{j=1}^{K_i} \left\{ \frac{\Lambda_0(T_{i,j}) - \Lambda_0(T_{i,j-1})}{\Lambda_0(Y_i)} \right\}^{m_{ij}} = \prod_{i=1}^{n} \prod_{j=1}^{K_i} \left\{ \frac{F(T_{i,j}) - F(T_{i,j-1})}{F(Y_i)} \right\}^{m_{ij}}. \tag{14}$$

Huang et al. (2006) considered this and suggested the nonparametric maximum likelihood estimator of F, denoted by $\hat{F}_n$, to be the non-decreasing, non-negative step function with possible jumps only at observation times $\{T_{i,j}, j = 1, \ldots, K_i, i = 1, \ldots, n\}$ that maximizes L in (14). Also they pointed out that $\hat{F}_n(t)$ can be computed by the self-consistency algorithm proposed by Turnbull (1976) or the EM-algorithm. In the EM-algorithm, both the E-step and the M-step have simple closed-form solutions.

To estimate $\Lambda_0(t)$, note that $\Lambda_0(t) = F(t)\Lambda_0(\tau)$ and thus we only need to estimate $\Lambda_0(\tau)$. For this, Huang et al. (2006) proposed to use

$$\hat{\Lambda}_n(\tau) = n^{-1} \sum_{i=1}^{n} \frac{m_i}{\hat{F}_n(Y_i)}$$

and hence a natural estimate of $\Lambda_0(t)$ is given by $\hat{\Lambda}_n(t) = \hat{F}_n(t)\hat{\Lambda}_n(\tau)$.

Let

$$\mathcal{F} = \left\{ F : [0, \tau] \to [0, M] \mid F \text{ is non-decreasing} \right\}$$

and

$$\mu(t) = E\left\{ \sum_{j=1}^{K} I(T_{i,j} \le t) \right\}.$$

Define the $L_2(\mu)$ metric d on $\mathcal{F}$ by

$$d^2(F_1, F_2) = \int \left| F_1(t) - F_2(t) \right|^2 d\mu(t), \quad F_1, F_2 \in \mathcal{F}.$$

Huang et al. (2006) showed that under some regularity conditions,

$$n^{\frac{1}{2}} d(\hat{\Lambda}_n, \Lambda_0) = O_p(1).$$

That is, the estimate $\hat{\Lambda}_n(t)$ is consistent. However, the asymptotic distribution of the estimate has not been established yet.

7.2 Mean function-based frailty model

Again let Z be a non-negative latent variable with $E(Z)=1$ and suppose that the event process $N(t)$ is a non-homogeneous Poisson process. Also as in the previous subsection, we assume that the event process and the observation process are correlated only through the latent variable Z. However, instead of modeling the rate function, we assume that given Z, the mean function of $N(t)$ has the form

$$\Lambda(t|Z)=Z\Lambda_0(t), \quad t\in[0,\tau],$$

where $\Lambda_0(t)$ is an unspecified, unknown function as before. This gives

$$E\big(N(t)\big)=E(Z)\Lambda_0(t)=\Lambda_0(t).$$

To estimate $\Lambda_0(t)$, assume that Z follows Gamma$(\alpha, 1/\alpha)$ and the distributions of K_i's and $T_{i,j}$'s are unrelated to the parameters Λ_0 and α. Then one can construct a pseudo-likelihood function based on the observations $\{K_i, T_{i,j}, N_{i,j}, j=1,\dots,K_i, i=1,\dots,n\}$ as

$$L_n(\alpha,\Lambda)=\prod_{i=1}^{n}\prod_{j=1}^{K_i}\frac{\Gamma(N_{i,j}+\alpha^{-1})}{\Gamma(\alpha^{-1})N_{i,j}!}\frac{\{\alpha\Lambda(T_{i,j})\}^{N_{i,j}}}{\{1+\alpha\Lambda(T_{i,j})\}^{N_{i,j}+\alpha^{-1}}}. \tag{15}$$

Zhang and Jamshidian (2003) considered this pseudo-likelihood function and defined the nonparametric maximum pseudo-likelihood estimator of the mean function Λ_0 to be the non-decreasing, non-negative step function with possible jumps only at observation times $\{T_{i,j}, j=1,\dots,K_i, i=1,\dots,n\}$ that maximizes $L_n(\alpha,\Lambda)$ in (15). Also they presented an EM-algorithm for computing the estimator. However, its asymptotic properties are still unknown.

A lot of research is still needed for the analysis of panel count data with dependent observation process. Here we have only described two simple models and it is obvious that many other models could be considered. Also one may investigate the treatment comparison problem discussed above and develop appropriate inferential procedures when there are covariates of interest, as considered below.

8 Nonparametric analysis with covariates

As mentioned before, the focus of this article is on nonparametric inference based on panel count data. However, to make it complete, we will briefly consider the situation where there are some covariates and discuss several available estimation procedures for covariate effects with the focus being on the marginal modeling approach. As in the preceding sections, we assume that there are n independent subjects and each gives rise to a counting process. Also we assume that for each subject, there exists a p-dimensional vector of covariates denoted by X_i, assumed to be time-independent. In the following, we will first discuss the Poisson process-based approaches that assume that the counting processes of interest are non-homogeneous Poisson processes. They will be followed by some estimating equation-based approaches that do not rely on the Poisson process assumption and some procedures that can be applied when there exists a dependent observation process.

8.1 The Poisson process-based procedures

Let the $N_i(t)$'s, $T_{i,j}$'s, $N_{i,j}$'s, and s_l's be as defined before. Define $\Lambda(t|X_i) = E\{N_i(t)|X_i\}$, the mean function given covariates X_i. In this subsection, we will assume that the $N_i(t)$'s are non-homogeneous Poisson processes and the mean function $\Lambda(t|X)$ has the form

$$\Lambda(t|X) = \Lambda_0(t)\exp\bigl(X'\beta\bigr). \tag{16}$$

Here $\Lambda_0(t)$ is an unknown baseline mean function, the mean function for subjects with $X = 0$, and β is a p-dimensional vector of regression parameters. As in Sect. 3.2, by ignoring the dependence of $\{N_i(T_{i,j}), j = 1, \ldots, K_i\}$ for each i, one can easily derive a pseudo-log-likelihood function given by

$$l_p(\Lambda_0, \beta) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}\bigl\{N_{i,j}\log\Lambda_0(T_{i,j}) + N_{i,j}X_i'\beta - \Lambda_0(T_{i,j})\exp\bigl(X_i'\beta\bigr)\bigr\}.$$

To estimate $\Lambda_0(t)$ and β, it is natural to maximize the pseudo-log-likelihood function $l_p(\Lambda_0, \beta)$. For this, let the w_l's and $\bar{N}_l$'s be as defined in Sect. 3 and define

$$\bar{a}_l(\beta) = \frac{1}{w_l}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\exp\bigl(X_i'\beta\bigr)I(T_{i,j} = s_l)$$

and

$$\bar{b}_l(\beta) = \frac{1}{w_l}\sum_{i=1}^{n}\sum_{j=1}^{K_i}N_{i,j}X_i'\beta I(T_{i,j} = s_l)$$

for given β, $l = 1, \ldots, m$. Then the pseudo-log-likelihood function $l_p(\Lambda_0, \beta)$ can be rewritten as

$$l_p(\Lambda_0, \beta) = \sum_{l=1}^{m} w_l\bigl\{\bar{N}_l\log\Lambda_0(s_l) - \bar{a}_l(\beta)\Lambda_0(s_l) + \bar{b}_l(\beta)\bigr\}.$$

As with the nonparametric situation, only the values of $\Lambda_0(t)$ at the s_l's can be estimated. Let $\hat{\Lambda}_0(t)$ and $\hat{\beta}$ denote the estimators of $\Lambda_0(t)$ and β that maximize l_p with $\hat{\Lambda}_0(t)$ being a non-decreasing step function with possible jumps only at the s_l's. Then the determination of $\hat{\Lambda}_0(t)$ and $\hat{\beta}$ is equivalent to maximizing $l_p(\Lambda_0, \beta) = l_p(\Lambda, \beta)$ over the $(m + p)$ unknown parameters $\Lambda = (\Lambda_1, \ldots, \Lambda_m)$ and β with $\Lambda_1 \leq \cdots \leq \Lambda_m$, where $\Lambda_l = \Lambda_0(s_l)$, $l = 1, \ldots, m$. To maximize $l_p(\Lambda, \beta)$, Zhang (2002) proposed a two-step iterative algorithm that maximizes l_p over Λ and β alternatively. Note that for fixed β, the maximization of l_p over Λ is equivalent to maximizing

$$\sum_{l=1}^{m} w_l\bar{a}_l(\beta)\biggl(\frac{\bar{N}_l}{\bar{a}_l(\beta)}\log\Lambda_l - \Lambda_l\biggr).$$

That is, the $\hat{\Lambda}_0(s_l)$'s are the isotonic regression estimator of $\{\bar{N}_1/\bar{a}_1(\beta), \ldots, \bar{N}_m/\bar{a}_m(\beta)\}$ with weights $\{w_1\bar{a}_1(\beta), \ldots, w_m\bar{a}_m(\beta)\}$. Thus for given β, we have

$$\hat{\Lambda}_\beta(s_l) = \max_{r\le l}\min_{s\ge l} \frac{\sum_{v=r}^{s} w_v \bar{N}_v}{\sum_{v=r}^{s} w_v \bar{a}_v(\beta)} = \min_{s\ge l}\max_{r\le l} \frac{\sum_{v=r}^{s} w_v \bar{N}_v}{\sum_{v=r}^{s} w_v \bar{a}_v(\beta)}$$

by the max-min formula of the isotonic regression estimate.

For given $\Lambda_0(t)$ or Λ, one can simply use the Newton–Raphson algorithm for the estimation of β. It can be easily shown that the pseudo-log-likelihood function l_p is a concave function of β for given $\Lambda_0(t)$ and its value increases after each iteration (Zhang 2002). For the convergence criterion for the two-step algorithm given above, one can compare the relative absolute change of either the log-likelihood function l_p between two successive estimators of $\Lambda_0(t)$ and β or the difference between the two successive estimators. For the variance estimation of the resulting estimates, one could apply the simple bootstrap procedure.

For estimation of $\Lambda_0(t)$ and β, instead of using the pseudo-log-likelihood function l_p, one may consider to maximize the following full log-likelihood function:

$$l(\Lambda_0, \beta) = \sum_{l'=0}^{m-1}\sum_{l=l'+1}^{m} \tilde{n}_{l,l'} \log\big[\Lambda_0(s_l) - \Lambda_0(s_{l'})\big] - \sum_{l=1}^{m} b_l(\beta)\Lambda_0(s_l) + \sum_{i=1}^{n} N_{i,K_i} X_i'\beta,$$

where $b_l(\beta) = \sum_{i=1}^{n} I(T_{i,K_i} = s_l)\exp(X_i'\beta)$ and

$$\tilde{n}_{l,l'} = \sum_{i=1}^{n}\sum_{j=1}^{K_i} (N_{i,j} - N_{i,j-1}) I(T_{i,j} = s_l, T_{i,j-1} = s_{l'})$$

for $0 \le l' < l \le m$. It is obvious that $l(\Lambda_0, \beta)$ could yield more efficient estimates than $l_p(\Lambda_0, \beta)$ (Wellner et al. 2004). Wellner and Zhang (2007) studied the asymptotic properties of the maximum pseudo-likelihood estimator $(\hat{\beta}_n^p, \hat{\Lambda}_n^p)$ and the maximum likelihood estimator $(\hat{\beta}_n, \hat{\Lambda}_n)$. They established strong consistency, derived the rate of convergence of both estimators in some L_2-metrics related to the observation scheme, and also proved the asymptotic normality of both $\hat{\beta}_n^p$ and $\hat{\beta}_n$ under some mild conditions, but the asymptotic distributions of $\hat{\Lambda}_n^p$ and $\hat{\Lambda}_n$ are unknown.

8.2 The estimating equation-based procedures

In the previous subsection, the underlying counting processes giving rise to panel count data were assumed to be non-homogeneous Poisson processes. Clearly this may not be true in reality. In this subsection, we present some inferential procedures that do not require the Poisson assumption.

Let the $N_i(t)$'s, $T_{i,j}$'s, X_i's and $\Lambda(t|X)$ be as defined before and suppose that the mean function $\Lambda(t|X)$ satisfies the model (16). For subject i, suppose that there exists a random variable C_i representing the follow-up time on the subject. In some cases, we may have $C_i = T_{i,K_i}$ and sometimes this may not be the case. Define $O_i(t) = \sum_{j=1}^{K_i} I(T_{i,j} \le t)$, a counting process representing the total number of observations on subject i up to time t, $i = 1, \ldots, n$. In the following, we will assume that $N_i(t)$, $O_i(t)$ and C_i are independent of each other given X_i. Also we will assume that the goal is to make inference about regression parameters β.

To estimate β, for each subject, define a new process

$$\tilde{N}_i(t) = \int_0^t N_i(s)\, dO_i(s),$$

which has possible jumps only at the observation time points $T_{i,j}$'s with respective jump sizes $N_i(T_{i,j})$. It can be easily seen that as for the $O_i(t)$'s, we have recurrent event or complete data for the $\tilde{N}_i(t)$'s rather than panel count data. Also it can be easily shown that

$$E\left\{d\tilde{N}_i(t)|O_i(s), X_i\right\} = \Lambda_0(t) \exp\left(X_i'\beta\right) dO_i(t).$$

Let $o_i(t) = O_i(t) - O_i(t^-)$ and define

$$S^{(j)}(\beta; t) = \frac{\sum_{i=1}^n I(C_i \ge t) X_i^{\otimes j} \exp(X_i'\beta) o_i(t)}{\sum_{i=1}^n o_i(t)}$$

for t with $\sum_{i=1}^n o_i(t) > 0$ and $j = 0, 1, 2$, where $a^{\otimes j} = 1, a, aa'$, for $j = 0, 1, 2$. For the estimation of β, motivated by the partial score function with respect to the Cox type of models, Hu et al. (2003) proposed to use the estimating function

$$U_n(\beta; \tilde{N}, W) = \sum_{i=1}^n \int_0^\tau W(t) I(C_i \ge t)\left\{X_i - \bar{X}(t; \beta)\right\} d\tilde{N}_i(t),$$

where $W(\cdot)$ is a weight function as before and $\bar{X}(t; \beta) = S^{(1)}(\beta; t)/S^{(0)}(\beta; t)$. If all subjects have only one observation at time, say, t_0, then the estimating function given above reduces to

$$U_n(\beta; \tilde{N}, 1) = \sum_{i=1}^n X_i N_i(t_0) - \left\{\sum_{i=1}^n \int_0^{t_0} \frac{1}{\sum_{j=1}^n \exp(X_j'\beta)} dN_i(t)\right\} \times \left\{\sum_{i=1}^n X_i \exp\left(X_i'\beta\right)\right\}$$

with $W(t) = 1$.

Let $\hat{\beta}_n$ denote the solution to $U_n(\beta; \tilde{N}, 1) = 0$. Hu et al. (2003) showed that $\hat{\beta}_n$ is a consistent estimate of β and one can approximate the distribution $\sqrt{n}(\hat{\beta}_n - \beta_0)$ by the multivariate normal distribution with mean vector $\mathbf{0}$ and the covariance matrix $\hat{\mathbf{\Sigma}}_n = \hat{A}(\beta)^{-1} \hat{B}(\beta) \hat{A}(\beta)^{-1}$ with β replaced by $\hat{\beta}_n$, where β_0 denotes the true value of β,

$$\hat{A}(\beta) = -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} I(C_i \geq t)\left[\frac{S^{(2)}(\beta;t)}{S^{(0)}(\beta;t)} - \bar{X}(t;\beta)^{\otimes 2}\right] d\tilde{N}_i(t)$$

and

$$\hat{B}(\beta) = \frac{1}{n}\left[\sum_{i=1}^{n}\int_0^{\tau}\left\{X_i - \bar{X}(t;\beta)\right\} d\hat{M}_i(t;\beta)\right]^{\otimes 2}$$

with

$$\hat{M}_i(t;\beta) = \int_0^{t} I(C_i \geq s)\left\{N_i(s) - \hat{\Lambda}_0(s;\beta)\exp\left(X_i'\beta\right)\right\} dO_i(s)$$

and

$$\hat{\Lambda}_0(t;\beta) = \frac{\sum_{i=1}^{n} I(C_i \geq t) N_i(t) o_i(t)}{\sum_{i=1}^{n} I(C_i \geq t)\exp(X_i'\beta) o_i(t)}.$$

In the procedure given above, it was assumed that the covariates X_i have no effect on the observation process $O_i(t)$. In practice, this may not be true and for this, as with model (16) for $N_i(t)$, one natural way is to assume that

$$E\left\{O_i(t)|X_i\right\} = \tilde{\Lambda}_0(t)\exp\left(X_i'\gamma\right) \tag{17}$$

for the effect of X_i on $O_i(t)$, where $\tilde{\Lambda}_0(t)$ is an unknown baseline mean function and γ denotes the vector of regression parameters as $\Lambda_0(t)$ and β, respectively. Among others, Hu et al. (2003) considered the model (17) together with model (16) and generalized the estimation procedure described above. As the procedure given above, the generalized estimation procedure also does not involve the baseline mean functions $\Lambda_0(t)$ as well as $\tilde{\Lambda}_0(t)$. Sun and Wei (2000) investigated a more general situation where $N_i(t)$, $O_i(t)$ and C_i may depend on each other, but are independent given X_i. In particular, they used models (16) and (17) and the proportional hazards model for the effect of covariates on the follow-up time C_i.

In comparing the Poisson-based and estimating equation-based estimation procedures, it is clear that the former could be more efficient than the latter if the Poisson process assumption is valid. Of course, in practice, it may be difficult to check or verify this assumption without prior information. On the other hand, the former could be much more complicated than the latter partly because of the involvement or the need of estimation of the baseline mean functions. Another advantage of the estimating equation-based procedures is that they give a closed-form estimation of the variance.

Other authors who have investigated the problem discussed here include Cheng and Wei (2000), Lawless and Zhan (1998), Staniswalls et al. (1997), and Sun and Matthews (1997). In particular, Cheng and Wei (2000) developed an inferential approach similar to the estimation procedure based on $U_n(\beta; \tilde{N}, W)$. Lawless and Zhan (1998) and Staniswalls et al. (1997) gave some approaches that base the inference on the modeling of the rate function of the underlying counting process instead of the mean function.

8.3 Estimation with dependent observation processes

For all the procedures discussed so far in this section, it was assumed that observation times $T_{i,j}$'s or the counting process $O_i(t)$ that characterizes them is independent of the underlying counting process $N_i(t)$ governing the observed panel count data either completely or conditional on covariates. In practice, however, this may not be true. For example, the observation times could be hospitalization times and this could be the case in an observational study concerning the occurrence rate of certain symptoms related to a disease under study which can be observed or known only when the patients are in the hospital due to the disease. It is apparent that the patients may come to the hospital simply because of these symptoms and thus the observation times are related to the occurrence process. A more specific example is given by the bladder cancer panel count data discussed before, and for the data several authors noticed that some patients in the study had more visits than others, suggesting that the occurrence of bladder tumors and the visit may be related (Sun and Wei 2000).

In the case of dependent observation processes, as discussed in Sect. 7, a common approach is to employ some latent variable models to describe the relationship between the two processes involved. For the situation considered here, as in Sect. 7, let Z be a non-negative latent variable with $E(Z) = 1$ and suppose that given X and Z, the event process $N(t)$ is a non-homogeneous Poisson process with the intensity function

$$\lambda(t|X, Z) = Z\lambda_0(t)\exp\big(\beta' X\big), \quad t \in [0, \tau], \tag{18}$$

where $\lambda_0(t)$ is an unspecified function and β is a p-dimensional vector of regression parameters. Given X and Z, the event process $N(t)$ is independent of the number of observations K and the observation times $\{T_1, \ldots, T_K\}$. Here, the event process $N(t)$ and the observation times $\{T_1, \ldots, T_K\}$ can be correlated through the frailty variable Z, and the distribution of the frailty variable and the conditional distribution of the observation times given the frailty can be arbitrary and are left unspecified. Note that under model (18) the conditional likelihood function, given $\{Z_i, X_i, K_i, T_{ij}, j = 1, \ldots, K_i, i = 1, \ldots, n\}$, has the same expression given by (14). Thus, $F(t) = \Lambda_0(t)/\Lambda_0(\tau)$ can be estimated by $\hat{F}_n(t)$ given in Sect. 7.1. To estimate $\Lambda_0(\tau)$ and β, Huang et al. (2006) proposed the use of the estimating equation

$$n^{-1}\sum_{i=1}^{n} W_i X_{1i}\big\{m_i\hat{F}(Y_i) - \exp\big(X_{1i}'\theta\big)\big\} = 0, \tag{19}$$

where $X_{1i}' = (1, X_i')$, $Y_i = T_{i,K_i}$, $m_i = N_i(Y_i)$, $\theta' = (\gamma, \beta')$, $\gamma = \log\Lambda(\tau)$, and W_i is a weight function. Let $\hat{\theta}_n' = (\hat{\gamma}_n, \hat{\beta}_n')$ denote the solution to (19). Then $\Lambda_0(t)$ can be estimated by $\hat{\Lambda}_0(t) = \hat{F}_n(t)e^{\hat{\gamma}_n}$. Under some regularity conditions, Huang et al. (2006) showed that $\hat{\beta}_n - \beta \to 0$ almost surely as $n \to \infty$ and $d(\hat{\Lambda}_0 1_{[0,t]}, \Lambda_0 1_{[0,t]}) \to 0$ almost surely for all $t \in [0, \tau]$ as $n \to \infty$. Here the metric d is as defined in Sect. 7.1. The asymptotic distributions of $\hat{\beta}_n$ and $\hat{\Lambda}_0(t)$ have not been derived yet.

For the problem discussed here, Sun et al. (2007) also proposed a joint model of the event process $N(t)$ and the observation process $O(t)$ through a shared latent variable and developed a two-step inferential procedure for the estimation of regression parameters. Sometimes all of the event processes, the observation times and the censoring times may be related with each other and one may be interested in modeling them together. Among others, He et al. (2009) studied this case and developed an inferential procedure for the analysis of such panel count data by using two latent variables to characterize the relationship.

9 Bayesian analysis

In addition to the literature discussed above, there also exists some limited research on Bayesian approaches for the analysis of panel count data. Of course, for nonparametric inference in this case, one usually needs to put some restrictions on the nonparametric component of the model such that the mean function can be approximated by some kernel functions (Ishwaran and James 2004). To be more specific, suppose we observe panel count data $\{K_i, T_{i,j}, N_{i,j}, j = 1, \ldots, K_i, i = 1, \ldots, n\}$ and the mean function of the event process $N_i(t)$ can be modeled by

$$\Lambda(t|\mu) = \int_{\mathcal{S}} \int_0^t k_0(s, \nu)\, ds \mu(d\nu),$$

where k_0 is a pre-specified kernel and μ is a finite measure over a measurable space $(\mathcal{S}, A)$. Define F by $F(A|\nu) = \int_A k_0(s, \nu)\, ds$ for each Borel-measurable set A. Let $A_{ij} = (T_{i,j-1}, T_{i,j}]$ and $A_i = (0, T_{i,K_i}]$. Then the likelihood function is

$$L(\mu) = \exp\left\{-\sum_{i=1}^{n} \int_{\mathcal{S}} \int_0^{\infty} Y_i(t) F(dt|\nu)\mu(d\nu)\right\}$$
$$\times \prod_{i=1}^{n} \prod_{j=1}^{K_i} \prod_{l=1}^{m_{ij}} \int_{\mathcal{S}} F(A_{ij}|\nu_{ijl})\mu(d\nu_{ijl}),$$

where $Y_i(t) = I(t \in A_i)$ and $m_{ij} = N_{i,j} - N_{i,j-1}$.

Let ν be the vector of missing values ν_{ijl}. We assume a prior on (ν, μ) with the joint product measure

$$\prod_{i,j,l} \mu(d\nu_{ijl})\mathcal{G}(d\mu|\alpha, \beta),$$

where $\mathcal{G}(\cdot|\alpha, \beta)$ denotes a weighted gamma process law with shape parameter α (a finite measure over $\mathcal{S}$) and scale parameter β (a positive integrable function over $\mathcal{S}$). That is, for each Borel-measurable set $A \in \mathcal{A}$, the random measure μ, defined by

$$\mu(A) = \int_A \beta(s)\gamma_\alpha(ds),$$

is said to have a $\mathcal{G}(\cdot|\alpha, \beta)$ law, where γ_α is a gamma process over $\mathcal{S}$ with shape measure α. We call γ_α a gamma process with shape parameter α if $\gamma_\alpha(A)$ is a gamma($\alpha(A)$) random variable with mean $\alpha(A)$ and variance $\alpha(A)$. Thus, by Theorem 3 of James (2003), for any integrable function $g(\nu, \mu)$, the posterior for the likelihood is given by

$$\int g(\nu, \mu)\pi(d\nu, d\mu|\mathbf{X})$$
$$= \int\int g(\nu, \mu)\mathcal{G}\Big(d\mu|\alpha + \sum_{i,j,l} \delta_{\nu_{ijl}}, \beta^*\Big)\pi(d\nu|\mathbf{X}),$$

where

$$\pi(d\nu|\mathbf{X}) \propto m_0(d\nu) \prod_{i=1}^{n}\prod_{j=1}^{K_i}\prod_{l=1}^{m_{ij}} \beta^*(\nu_{ijl})F(A_{ij}|\nu_{ijl}),$$

$$\beta^*(\nu) = \frac{\beta(\nu)}{1 + \beta(\nu)\sum_{i=1}^{n} F(A_i|\nu)},$$

and

$$m_0(d\nu) = \int \prod_{i=1}^{n}\prod_{j=1}^{K_i}\prod_{l=1}^{m_{ij}} P(\nu_{ijl})\mathcal{P}(dP|\alpha)$$

is the Pólya urn density for a Dirichlet process $\mathcal{P}(\cdot|\alpha)$ (Ferguson 1973, 1974).

The posterior law for the function $g(\nu, \mu)$ can be approximated by the Pólya urn Gibbs sampling and the Blocked Gibbs sampling. For panel count data in general, the Blocked Gibbs sampler is preferred and more details on this issue can be found in Ishwaran and James (2004).

10 Discussion and concluding remarks

The analysis of panel count data is still a relatively new and growing field and there are still many open problems. In the preceding sections, our focus has been on the mean function of underlying counting processes generating panel count data. As mentioned before, given the structure of panel count data and the amount of observed information, it is much more convenient to deal with the mean function rather than the intensity process or rate function. On the other hand, sometimes one may want to directly model the intensity process or rate function (Ishwaran and James 2004; Lawless and Zhan 1998; Staniswalls et al. 1997). For this purpose, however, one usually has to make certain assumptions about the shape of the intensity process or rate function in order to perform nonparametric inference (Sun and Rai 2001; Sun and Matthews 1997).

Yet another feature of the methods described here is that they were developed mainly for panel count data in which observation and censoring times differ from subject to subject. For the situation wherein observation times or intervals are the same for all subjects, the data can be regarded as multivariate data and any method that accommodates multivariate positive integer-valued response variables can then be used for the analysis. This holds even though subjects may miss some intermediate observations and/or drop out of the study early. In this case, the resulting data can be seen as multivariate data with missing values. Also, as mentioned before, one can treat panel count data as a special case of longitudinal data and apply the methods developed for longitudinal data. However, these methods may not be able to take into account the special structure of panel count data and thus would be less efficient.

Note that for nonparametric inference on panel count data with covariates discussed in Sects. 6 and 8, only linear effects were considered. Of course, nonlinear covariate effects may exist in practice and in this case, one could use linear or some flexible covariate effects to approximate them. For example, one can replace a linear predictor $\beta' Z$ by $\sum_{j=1}^{p} g_j(Z_j)$, where Z is a p-dimensional covariate vector and g_j is a smooth function of Z_j and could be estimated by smoothing methods such as splines.

In preceding sections, all discussions have been on univariate situations. That is, there exists only one type of recurrent event. In reality, there may exist several related types of recurrent events that are of interest and for which only panel count data are available. Among others, Chen et al. (2005) discussed the analysis of multivariate panel count data using a marginal mixed Poisson process approach by assuming that the baseline intensity function is piecewise constant. He et al. (2008) studied the same problem and proposed some estimating equation-based approaches. One limitation for the approaches discussed above is that it has been assumed that all study subjects come from a single population. Sometimes the subjects may arise from a mixture of G different populations characterized by $N_i(t) = \sum_{g=1}^{G} z_{i,g} C_{i,g}(t)$, where $z_{i,g}$ indicates if subject i belongs to the subpopulation or cluster g and $C_{i,g}(t)$ is a non-homogeneous Poisson process. Among others, Nielsen and Dean (2008) investigated this type of situation and provided an example of such a panel count data arising from an experiment for testing the difference of pheromones in disrupting the mating pattern of the cherry bark tortrix moth.

For future research directions, one area that definitely needs more research is the treatment comparison based on panel count data. Although a few procedures are available, the two examples in Sect. 5 suggested that they may give different conclusions. Thus more work is needed for the selection of a particular procedure given a practical problem or for the development of new, more powerful procedures. Clearly an ideal situation is to develop an approach that provides an automatic choice, which would be difficult. A more practical idea is to combine an appropriate procedure with several weight functions or several procedures and develop some method that gives a minimum p-value like the Bonferroni procedure.

Another direction that has been briefly discussed here is the situation when the event history process and the observation process may be related. It is easy to see that all methods described for this have some limitations. For example, the approaches given in Sect. 7 were developed based on the Poisson process assumption and although they could still be used for general situations, their performance is not completely known. A related situation is that there exists a terminal event such as death that affects or is related to the recurrent event of interest. Ghosh and

Lin (2000) discussed this for recurrent event data and the problem is much harder for panel count data. For a recurrent event, instead of the occurrence rate of the event, sometimes one may be interested in the gap time of the event, the time between successive occurrences of the event (Sun et al. 2006; Zhao and Sun 2006; Du 2009). For the analysis of gap times, although some approaches have been proposed for the case of recurrent event data, there exists little research for panel count data situations. The same is true for the model or variable selection based on panel count data (Tong et al. 2009) and for the regression model that involves time-dependent covariates or time-varying covariate effects (Sun et al. 2009).

Acknowledgements We wish to thank the editor and the two referees for their many very helpful comments and suggestions, which greatly improved the paper. The research is supported in part by a grant from the Research Grant Council of Hong Kong and a grant from the Natural Sciences and Engineering Research Council of Canada (N. Balakrishnan).

References

Aalen OO (1975) Statistical inference for a family of counting presses. PhD Thesis. University of California, Berkeley

Aalen OO (1978) Nonparametric inference for a family of counting processes. Ann Stat 6:701–726

Andersen PK, Borgan O (1985) Counting process models for life history data: a review. Scand J Stat 12:97–158

Andersen PK, Borgan O, Gill RD, Keiding N (1993) Statistical models based on counting processes. Springer, New York

Anderson TW, Goodman LA (1957) Statistical inference about Markov chains. Ann Math Stat 28:89–110

Andrews DF, Herzberg AM (1985) Data: A collection of problems from many fields for the student and research worker. Springer, New York

Balakrishnan N, Zhao X (2009) New multi-sample nonparametric tests for panel count data. Ann Stat 37:1112–1149

Balakrishnan N, Zhao X (2010a) A nonparametric test for the equality of counting processes with panel count data. Comput Stat Data Anal 54:135–142

Balakrishnan N, Zhao X (2010b) A class of multi-sample nonparametric tests for panel count data. Ann Inst Stat Math, im press

Bartholomew DJ (1983) Some recent developments in social statistics. Int Stat Rev 51:1–9

Breslow NE (1984) Extra-Poisson variation in log-linear models. Appl Stat 33:38–44

Byar DP, 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

Chen BE, Cook RJ, Lawless JF, Zhan M (2005) Statistical methods for multivariate interval-censored recurrent events. Stat Med 24:671–691

Cheng SC, Wei LJ (2000) Inferences for a semiparametric model with panel data. Biometrika 87:89–97

Commenges D (2002) Inference for multi-state models from interval-censored data. Stat Methods Med Res 11:167–182

Cook RJ, Lawless JF (1996) Interim monitoring of longitudinal comparative studies with recurrent event responses. Biometrics 52:1311–1323

Cook RJ, Lawless JF (2007) The statistical analysis of recurrent events. Springer, New York

Cook RJ, Lawless JF, Nadeau C (1996) Robust tests for treatment comparisons based on recurrent event response. Biometrics 52:557–571

Cox DR, Miller HD (1965) The theory of stochastic processes. Methuen, London

Davis CS, Wei LJ (1988) Nonparametric methods for analyzing incomplete nondecreasing repeated measurements. Biometrics 44:1005–1018

Diamond ID, McDonald JW (1991) The analysis of current status data. In: Trussel J, Hankinson R, Tilton J (eds) Demographic applications of event history analysis. Oxford University Press, Oxford

Diggle PJ, Liang KY, Zeger SL (1994) The analysis of longitudinal data. Oxford University Press, New York

Du P (2009) Nonparametric modeling of the gap time in recurrent data. Lifetime Data Anal 15:256–277

Ferguson TS (1973) A Bayesian analysis of some nonparametric problems. Ann Stat 1:209–230

Ferguson TS (1974) Prior distributions on spaces of probability measures. Ann Stat 2:615–629

Gentleman RC, Lawless JF, Lindsey JC (1994) Multi-state Markov models for analysing incomplete diseases history data with illustrations for HIV disease. Stat Med 13:805–821

Ghosh D, Lin DY (2000) Nonparametric analysis of recurrent events and death. Biometrics 56:554–562

He X, Tong X, Sun J, Cook RJ (2008) Regression analysis of multivariate panel count data. Biostatistics 9:234–248

He X, Tong X, Sun J (2009) Semiparametric analysis of panel count data with correlated observation and follow-up times. Lifetime Data Anal 15:177–196

Hinde J (1982) Compound Poisson regression models. In: Gilchrist R (ed) GLIM 82: Proceedings of the international conference in generalized linear models. Springer, Berlin, pp 109–121

Hu XJ, Lagakos SW (2007) Nonparametric estimation of the mean function of a stochastic process with missing observations. Lifetime Data Anal 13:51–73

Hu XJ, Sun J, Wei LJ (2003) Regression parameter estimation from panel counts. Scand J Stat 30:25–43

Hu XJ, Lagakos SW, Lockhart RA (2009a) Marginal analysis of panel counts through estimating functions. Biometrika 96:445–456

Hu XJ, Lagakos SW, Lockhart RA (2009b) Generalized least squares estimation of the mean function of a counting process based on panel counts. Stat Sin 19:561–580

Huang C-Y, Wang M-C, Zhang Y (2006) Analysing panel count data with informative observation times. Biometrika 93:763–775

Ishwaran H, James LF (2004) Computational methods for multiplicative intensity models using weighted gamma processes: proportional hazards, marked point processes, and panel count data. J Am Stat Assoc 99:175–190

James LF (2003) Bayesian calculus for gamma processes with applications to semiparametric intensity models. Sankhya, Ser A 65:196–223

Kalbfleisch JD, Lawless JF (1985) The analysis of panel data under a Markov assumption. J Am Stat Assoc 80:863–871

Kalbfleisch JD, Lawless JF Robinson JA (1991) Methods for the analysis and prediction of warranty claims. Technometrics 33:273–285

Lawless JF, Nadeau JC (1995) Some simple robust methods for the analysis of recurrent events. Technometrics 37:158–168

Lawless JF, Zhan M (1998) Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. Can J Stat 26:549–565

Lin DY, Wei LJ, Yang I, Ying Z (2000) Semiparametric regression for the mean and rate functions of recurrent events. J R Stat Soc B 62:711–730

Lu M, Zhang Y, Huang J (2007) Estimation of the mean function with panel count data using monotone polynomial splines. Biometrika 94:705–718

Meira-Machado L, de Uña-Álvarez J, Cadarso-Suarez C, Andersen PK (2009) Multi-state models for the analysis of time-to-event data. Stat Methods Med Res 18:195–222

Nelson WB (2003) Recurrent events data analysis for product repairs, disease recurrences, and other applications. ASA-SIAM series on statistics and applied probability. SIAM, Philadelphia

Nielsen JD, Dean CB (2008) Clustered mixed nonhomogeneous Poisson process spline models for the analysis of recurrent event panel data. Biometrics 64:751–761

Park D-H, Sun J, Zhao X (2007) A class of two-sample nonparametric tests for panel count data. Commun Stat Theory Methods 36:1611–1625

Pepe MS, Cai J (1993) Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. J Am Stat Assoc 88:811–820

Pepe MS, Fleming TR (1989) Weighted Kaplan–Meier statistics: a class of distance tests for censored survival data. Biometrics 45:497–507

Petroni GR, Wolfe RA (1994) A two sample test for stochastic ordering with interval-censored data. Biometrics 50:77–87

Robertson T, Wright FT, Dykstra R (1988) Order restricted statistical inference. Wiley, New York

Singer B, Spilerman S (1976a) The representation of social processes by Markov models. Am J Sociol 82:1–54

Singer B, Spilerman S (1976b) Some methodological issues in the analysis of longitudinal surveys. Ann Econ Sociol Meas 5:447–474

Staniswalls JG, Thall PF, Salch J (1997) Semiparametric regression analysis for recurrent event interval counts. Biometrics 53:1334–1353

Sun J (2006) The statistical analysis of interval-censored failure time data. Springer, New York

Sun J, Fang HB (2003) A nonparametric test for panel count data. Biometrika 90:199–208

Sun J, Kalbfleisch JD (1993) The analysis of current status data on point processes. J Am Stat Assoc 88:1449–1454

Sun J, Kalbfleisch JD (1995) Estimation of the mean function of point processes based on panel count data. Stat Sin 5:279–290

Sun J, Matthews DE (1997) A random-effect regression model for medical follow-up studies. Can J Stat 25:101–111

Sun J, Rai SN (2001) Nonparametric tests for the comparison of point processes based on incomplete data. Scand J Stat 28:725–732

Sun J, Wei LJ (2000) Regression analysis of panel count data with covariate-dependent observation and censoring times. J R Stat Soc B 62:293–302

Sun J, Tong X, He X (2007) Regression analysis of panel count data with dependent observation times. Biometrics 63:1053–1059

Sun L, Park D, Sun J (2006) The additive hazards model for recurrent gap times. Stat Sin 16:919–932

Sun L, Zhu L, Sun J (2009) Regression analysis of multivariate recurrent event data with time-varying covariate effects. J Multivar Anal 100:2214–2223

Thall PF (1988) Mixed Poisson likelihood regression models for longitudinal interval count data. Biometrics 44:197–209

Thall PF, Lachin JM (1988) Analysis of recurrent events: nonparametric methods for random-interval count data. J Am Stat Assoc 83:339–347

Tong X, He X, Sun L, Sun J (2009) Variable selection for panel count data via nonconcave penalized estimating function. Scand J Stat 36:620–635

Turnbull BW (1976) The empirical distribution with arbitrarily grouped censored and truncated data. J R Stat Soc B 38:290–295

Vermunt JK (1997) Log-linear models for event histories. Sage Publications, Newbury Park

Wang MC, Chen YQ (2000) Nonparametric and semiparametric trend analysis of stratified recurrence time data. Biometrics 56:789–794

Wang MC, Qin J, Chiang CT (2001) Analyzing recurrent event data with informative censoring. J Am Stat Assoc 96:1057–1065

Wasserman S (1980) Analyzing social networks as stochastic processes. J Am Stat Assoc 75:280–294

Wei LJ, Lachin JM (1984) Two-sample asymptotically distribution-free tests for incomplete multivariate observations. J Am Stat Assoc 79:653–661

Wellner JA, Zhang Y (2000) Two estimators of the mean of a counting process with panel count data. Ann Stat 28:779–814

Wellner JA, Zhang Y (2007) Two likelihood-based semiparametric estimation methods for panel count data with covariates. Ann Stat 35:2106–2142

Wellner JA, Zhang Y, Liu H (2004) A semiparametric regression model for panel count data: when do pseudo-likelihood estimators become badly inefficient? In: Proceedings of the second Seattle symposium in biostatistics. Springer, New York, pp 143–174

Zhang Y (2002) A semiparametric pseudolikelihood estimation method for panel count data. Biometrika 89:39–48

Zhang Y (2006) Nonparametric k-sample tests with panel count data. Biometrika 93:777–790

Zhang Y, Jamshidian M (2003) The gamma-frailty Poisson model for the nonparametric estimation of panel count data. Biometrics 59:1099–1106

Zhang Y, Liu W, Zhan Y (2001) A nonparametric two-sample test of the failure function with interval censoring, case 2. Biometrika 88:677–686

Zhao Q, Sun J (2006) Semiparametric and nonparametric analysis of recurrent events with observation gaps. Comput Stat Data Anal 51:1924–1933