Journal of the Korean Statistical Society 45 (2016) 250–259

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

Nonparametric comparison of recurrent event processes based on panel count data

Da Xu [a], Jianguo Sun [a,b,*], Dehui Wang [a]

[a] Institute of Mathematics, Jilin University, Changchun 130012, China
[b] Department of Statistics, University of Missouri, Columbia, MO 65211, USA

ARTICLE INFO

Article history:
Received 13 May 2015
Accepted 5 November 2015
Available online 28 November 2015

AMS 2000 subject classifications:
primary 62G10
secondary 62G20

Keywords:
Nonparametric test
Panel count data
Recurrent event process
Unequal observation scheme

ABSTRACT

Recurrent event processes arise in many fields including medical follow-up studies and reliability experiments and many procedures have been developed in the literature for their comparison nonparametrically (Cook and Lawless, 2007; Sun and Zhao, 2013). However, most of them are for either the complete data situation where one observes recurrent event data or the incomplete data situation where one observes panel count data with the same observation process. There also exist a couple of nonparametric comparison procedures for the panel count data situation that allow unequal observation processes, but apply only to limited situations. In this paper, we discuss the latter situation for both univariate and multivariate panel count data and propose a new type of nonparametric procedures that apply to more general situations. The proposed test statistics are shown to have asymptotic normal distributions, and an extensive simulation is conducted and suggests that they work well in practical situations. An application is also provided.

1. Introduction

This paper discusses nonparametric comparison of recurrent event processes that arise in many fields including medical follow-up studies and reliability experiments, and for the problem, many procedures have been developed in the literature for either recurrent event data or panel count data (Cook & Lawless, 2007; Sun & Zhao, 2013). By recurrent event data, we usually mean that all study subjects are observed continuously until certain times and thus one has complete data or information on the sample path or recurrent event process of interest. In contrast, panel count data are commonly used to refer to the situation where each study subject is observed only at discrete time points and as a result, only incomplete data on the sample path is available (Ai, You, & Zhou, 2013; Huang, Wang, & Zhang, 2006; Sun & Wei, 2000; Sun & Zhao, 2013). In particular, for the latter case, one only observes the number of the recurrent events that have occurred between observation times. Of course, the number of observations and observation times commonly vary from subject to subject. In the following, we will focus on the panel count data situation with unequal observation processes.

A well-known example of panel count data arose from the clinical trial of bladder tumor patients conducted by the Veterans Administration Cooperative Urological Research Group (Sun & Wei, 2000; Sun & Zhao, 2013; Wellner & Zhang, 2007). In this study, all patients had superficial bladder tumors when they entered the trial and these tumors were removed transurethrally. They were randomly allocated to one of three treatments: placebo, thiotepa and pyridoxine, and examined

* Corresponding author at: Institute of Mathematics, Jilin University, Changchun 130012, China.
E-mail address: sunj@missouri.edu (J. Sun).

http://dx.doi.org/10.1016/j.jkss.2015.11.001
1226-3192/

from time to time for the occurrences of new bladder tumors. In particular, at each follow-up visit, the number of the tumors that occurred since the previous visit was recorded and all of them were removed transurethrally. Also as pointed out by some authors (Li, Zhao, Sun, & Kim, 2014; Zhao & Sun, 2011), the patients in the thiotepa group seem to be observed much more frequently than those in the other groups. One main objective of the study is to assess the treatment effect on the tumor growth or the recurrence rate of bladder tumors.

For the treatment comparison based on panel count data, a number of procedures have been developed in the literature. For example, one of the early work on this was given by Thall and Lachin (1988), who suggested the use of some data grouping methods. Among the nonparametric procedures, Sun and Fang (2003) gave a method that involves the use of the isotonic regression estimator for the estimation of the common mean function and Park, Sun, and Zhao (2007) presented a class of two-sample tests also based on the isotonic regression estimator. Furthermore, Balakrishnan and Zhao (2011) and Zhang (2006) developed some multi-sample procedures by using the nonparametric maximum pseudo-likelihood approach, while Balakrishnan and Zhao (2009) proposed two classes of test statistics by using the nonparametric maximum likelihood approach. Note that all of the methods mentioned above along with most of the other available procedures require that observation times follow the same observation process, which clearly may not be true in practice. To address this, both Li et al. (2014) and Zhao and Sun (2011) proposed some nonparametric test procedures that allow different or unequal observation processes. As pointed out by Li et al. (2014), however, the procedure given in Zhao and Sun (2011) may not have power when the number of observations is small due to the estimation of the common mean function. Also as will be shown below, the method developed by Li et al. (2014) can fail too if the number of treatment groups is greater than two.

In the following, we will propose a new class of nonparametric test procedures that allow unequal observation processes and apply to more general situations. The method will not involve the estimation of the common mean function and be constructed as the contrasts of the sample means of the integrated weighted responses from the underlying recurrent event processes. In Section 2, after introducing some notation and assumptions, we will first consider the situation of univariate panel count data and present a class of nonparametric test procedures. They will be generalized to the situation of multivariate panel count data in Section 3 and for both cases, the asymptotic normality of the test statistics will be established. Section 4 will present some results obtained from an extensive simulation study conducted for assessing the performance of the proposed approaches and in Section 5, we apply the proposed approach to the bladder tumor data discussed above. Section 6 concludes with some discussion and remarks.

2. Test procedures for univariate panel count data

Consider an event history study that involves n independent subjects from m different treatment groups and each subject may experience a single type of recurrent events. Suppose that there are n_l subjects in the lth group and let S_l denote the set of indices for the subjects belonging to group l, where $\sum_{l=1}^{m} n_l = n$, $l = 1, 2, \ldots, m$. Also let $N_i(t)$ be the underlying counting process representing the total number of the occurrences of the recurrent event of interest up to time t from subject i, and Z_i and C_i denote the group-indicating vector associated with the subject and the censoring or follow-up time on the subject, respectively. In the following, we will assume that each subject is observed only at the discrete time points $T_{i,1} < T_{i,2} < \cdots < T_{i,m_i}$, where m_i represents the total number of observation times on subject i. Then the observed data have the form $\{Z_i, C_i, m_i, T_{i,j}, N_i(T_{i,j}); j = 1, 2, \ldots, m_i, i = 1, 2, \ldots, n\}$. That is, one only has panel count data on the $N_i(t)$'s. Our goal will be to test the null hypothesis

$$H_0 : \mu_1(t) = \mu_2(t) = \cdots = \mu_m(t),$$

where $\mu_l(t) = E\{N_i(t)|Z_i\}$ for $i \in S_l$, the mean function of the underlying event history process for the subjects in group l.

Define $O_i(t) = O_i^*(C_i \wedge t)$, where $O_i^*(t) = \sum_{j=1}^{\infty} I(T_{i,j} \le t)$ and $a \wedge b$ denotes the minimum of a and b, $i = 1, 2, \ldots, n$. Note that $O_i^*(t)$ and $O_i(t)$ represent the underlying and observed observation processes, respectively. As discussed above, the subjects in different treatment groups may have different observation processes. That is, they may depend on the Z_i's. To characterize this, we will assume that $O_i^*(t)$ can be described by the following proportional rate model

$$E\{dO_i^*(t)|Z_i\} = \lambda_0(t)\ \exp(\gamma' Z_i)\, dt, \tag{1}$$

where $\lambda_0(t)$ is an unspecified continuous function and γ is a vector of unknown regression parameters. It is easy to see that under the model above, $\gamma = \mathbf{0}$ means that the observation processes are independent of the treatments or identical for all subjects. In the following, we will assume that $N_i(t)$ and $O_i^*(t)$ are independent of each other given Z_i and also that C_i is independent of $\{N_i(t), O_i^*(t), Z_i\}$, $i = 1, 2, \ldots, n$. Some comments on them will be given below.

To construct the test statistics, note that under model (1) and conditional on Z_i, for $i \in S_l$, one can easily show that

$$E\left\{\sum_{j=1}^{m_i} N_i(T_{i,j}) \middle| Z_i\right\} = E\left\{\int_0^\tau N_i(t) I(C_i \ge t) dO_i^*(t) \middle| Z_i\right\} = \exp(\gamma' Z_i) \int_0^\tau \mu_l(t) G(t) \lambda_0(t) dt,$$

where $G(t) = P(C_i \ge t)$ and τ denotes the longest follow-up time. It thus follows that we have

$$E\left\{\int_0^\tau \frac{N_i(t) dO_i(t)}{\exp(\gamma' Z_i)} \middle| Z_i\right\} = \int_0^\tau \mu_l(t) G(t) \lambda_0(t) dt. \tag{2}$$