ANALYSIS OF TRUNCATED RECURRENT EVENT DATA WITH APPLICATION TO WARRANTY CLAIMS

X. Joan Hu and J.F. Lawless

IIQP Research Report
RR-94-04

March 1994

Analysis of Truncated Recurrent Event Data with Application to Warranty Claims

By X. Joan Hu and J. F. Lawless†
University of Waterloo, Canada

SUMMARY

This paper investigates data on recurrent events that arise from sources such as warranty claims, where the observation period for a unit is unknown until it experiences at least one event (*e.g.*, warranty claim). This creates a type of truncation in the data. We consider nonparametric estimation of means and rates of the event occurrences for both discrete and continuous time cases with such "zero-truncated" data and examine the case where the population size and the distribution of observation times across units are at least approximately known. The behaviors of the proposed estimators are studied by simulation as well as through their asymptotic properties. We present an analysis of some car warranty data by applying the methodology developed in the paper.

Keywords: nonparametric estimation; truncation distribution; population size; recurrent events; warranty claims; zero-truncated data

1 Introduction

Field failure data provide important information about the reliability of manufactured products. Since followup of selected products is expensive, there is much interest in the utilization of information from sources such as failure surveillance reports or warranty claim records. However, there are inherent problems with such data; these include incorrect failure diagnosis or recording errors and delays in the reporting of events such as warranty claims to the manufacturer (Kalbfleisch, Lawless and Robinson 1991, Lawless and Kalbfleisch 1992). Another major difficulty is that the time origin for a unit is often unknown until it experiences at least one event (*e.g.*, failure or warranty claim), thus creating a type of truncation in the data (*e.g.*, Suzuki 1985 a,b, 1987; Kalbfleisch and Lawless 1988, Lawless and Kalbfleisch 1992).

†*Address for correspondence*: Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

This paper deals with data on recurrent events that arise from such sources although, as noted later, applications of the methodology are not restricted to warranty or field failure problems. The statistical aspects of the topic are as follows. An individual or unit experiences recurrent events over time $t > 0$; we let $N_i(t)$ denote the number of events over $(0, t]$ for unit i. Unit i is observed (*i.e.*, its times of event occurrence are recorded) over the time interval $(0, \tau_i]$; we call τ_i the observation or truncation time. We consider both discrete and continuous time. When t is discrete, we will assume it takes on values $t = 1, 2, 3, \ldots$, and the τ_i's likewise take on integer values. A set of M units, $i = 1, 2, \ldots, M$, is assumed to have independent and identically distributed event occurrence processes $\{N_i(t), t > 0\}$ and observation times τ_i, $i = 1, \ldots, M$, that are determined independently of the event processes. The statistical objective is to estimate the mean function $\Lambda(t) = E\{N_i(t)\}$, and the corresponding rate, defined as $\lambda(t) = \frac{d\Lambda(t)}{dt}$ when time is continuous and as $\lambda(t) = E\{N_i(t) - N_i(t-1)\}$ when time is discrete.

The novelty in the situations we consider is that unit i and τ_i are observed only if $N_i(\tau_i) > 0$, *i.e.*, provided that at least one event has occurred. Otherwise the τ_i value for the unit i is unknown, and we may even be unaware of its existence. The motivation for studying these situations came from attempts to utilize warranty data, which we now discuss.

Manufacturers collect information about failures or repairs that result in warranty claims. The repairs may be thought of as recurrent events, and it is of considerable interest to estimate the mean number of repairs $\Lambda(t)$ and the rate $\lambda(t)$ per unit from warranty records. "Time" may be elapsed calendar time since the unit was sold (*i.e.*, "age" of the unit) or some measure of usage such as accumulated mileage with automobiles. The problem is to estimate $\Lambda(t), t > 0$ from warranty data collected up to some point in time. For a unit sold before then the observation time τ_i is a function of when the unit was sold, the type of warranty plan, and possibly the usage history of the unit. In most situations the value of τ_i becomes known only when the unit has its first claim, thus falling into the scenario described in the preceding paragraph. Moreover, in some situations, for example when "time" is a measure of usage, the τ_i's may never be known exactly, but only approximately. We give a pair of specific examples.

Example 1. Many products, such as tools or small household appliances, have a fixed time warranty, say one year from the date of sale. In this case, if unit i is sold at calendar time x_i, X is the current calendar time, and t represents the age of the unit, then $\tau_i = \min(X - x_i, 1)$, with time measured in years. With products of this type, however, the manufacturer usually receives date of sale information for at most a small fraction of units. Thus, the date of sale and truncation time τ_i become known for most units only when they have a claim.

Example 2. For automobiles the manufacturer is informed of the date of sale for each car. However, since warranties usually have both calendar time and mileage limits the observation times τ_i are typically observed or estimated only when a claim is made. For example, if the warranty coverage is for two years or 20,000 miles and t represents age of the car in years, then $\tau_i = \min(X - x_i, 2, y_i)$, where x_i is the calendar time of sale, X is the current calendar time, and y_i is the age at which the car mileage reaches 20,000. Although x_i is known, y_i is not. It can, however, be estimated from the mileage at the time of the first claim; this is usually done under the generally reasonable assumption that in their first few years cars accumulate mileage approximately linearly with time. If t represented mileage instead of age, then we would have $\tau_i = \min(w_i(2), w_i(X - x_i), 20000)$, where $w_i(a)$ is the mileage on car i at age a. Once again, τ_i may be estimated from the mileage at the time of the first claim.

Estimation of rates of occurrence with incomplete data on observation time has been discussed by Suzuki (1987). Kalbfleisch and Lawless (1988), Lawless and Kalbfleisch (1992), Suzuki (1985 a,b), Suzuki and Epstein (1992), and Suzuki and Kasashima (1993) have considered similar problems for the estimation of failure time distributions. Our paper is different from previous work in two main respects: we consider nonparametric estimation of means and rates of occurrence, and we deal both with truncated data and the case where the distribution of observation times across units is at least approximately known. The methods developed are especially useful in the warranty data context described here, but also apply to other types of observational studies. For example, a sociological study that tracked repeated utilizations of a social service by a population of individuals might not know of an individual's existence until the first utilization occurred.

It should be noted that this paper deals with marginal rates of event occurrence. It does not consider the estimation of event intensities, conditional on previous event history, except in the case of Poisson processes, where the intensity and rate of occurrence are the same.

The remainder of the paper is as follows. Section 2 discusses nonparametric estimation of mean and rate of occurrence functions when only "zero-truncated" data are available. Section 3 presents methods when the total number of units, and the distribution of observation times in the population are known; the increased information this provides is addressed. Section 2 and 3 are both based on Poisson models. Section 4 gives robust estimators for the rate of occurrence and mean functions. Simulation is used to study the behaviors of the estimators in Sections 2, 3 and 4 in a few situations. Section 5 considers the situation where data are aggregated across units. Section 6 illustrates the methodology on some car warranty data. Some remarks concerning extensions to the present work are made in Section 7.

2 Estimation from Zero-Truncated Data

We assume that $\{N_i(t) : t > 0\}$, $i = 1, \ldots, M$, are independent counting processes with common rate of occurrence function $\lambda(t)$ and mean function $\Lambda(t)$. The process i has an observation window $(0, \tau_i]$, where the τ_i's are determined independently of the event processes. In this section, we consider estimation of $\Lambda(t)$ when we are aware of only those processes with at least one event over $(0, \tau_i]$. The value of M is unknown in such a case. We assume for now that the counting processes are Poisson; a relaxation of this assumption is considered later.

Suppose for notational convenience that processes $i = 1, \ldots, m$ have at least one event and that for them the times of events $t_{ij}(j = 1, \ldots, n_i$, where $n_i = N_i(\tau_i))$ and observation times τ_i are observed. We then have the "zero-truncated" likelihood function

$$
\begin{aligned}
L_T &= \prod_{i=1}^{m} \Pr\{n_i, t_{ij}{}'s | n_i \geq 1, \tau_i\} \\
&= \prod_{i=1}^{m} \{\prod_{j=1}^{n_i} \lambda(t_{ij}) \frac{\exp[-\Lambda(\tau_i)]}{1 - \exp[-\Lambda(\tau_i)]}\}, \qquad (1)
\end{aligned}
$$

where we use " $\Pr\{.\}$" to represent either a probability or probability density, depending on whether the problem is in discrete or continuous time.

In later discussion we will assume that the τ_i's $(i = 1, \ldots, M)$ are independent and identically distributed (*i.i.d.*) with distribution function $G(.)$. This yields the likelihood function

$$
\begin{aligned}
L_{T_2} &= \prod_{i=1}^{m} \Pr\{n_i, t_{ij}{}'s, \tau_i | n_i \geq 1\} \\
&= L_T * \prod_{i=1}^{m} \frac{dG(\tau_i)[1 - \exp\{-\Lambda(\tau_i)\}]}{\int_0^\infty [1 - \exp\{-\Lambda(\tau)\}] dG(\tau)}. \qquad (2)
\end{aligned}
$$

If $G(.)$ is treated nonparametrically (*i.e.*, we do not assume a specific parametric form for it), L_{T_2} gives the same estimate of $\Lambda(t)$ as L_T; this is easily seen by defining $dF(\tau)$ as $[1 - \exp\{-\Lambda(\tau)\}]dG(\tau)$, and noting that the second term in (2) involves only $F(\tau)$, whereas the first involves only $\Lambda(t)$.

Estimation from (1) for parametric models $\lambda(t;\theta)$ poses no particular difficulties; we consider an example in Section 6. In the remainder of this section we develop nonparametric estimates. These are valuable for assessing the shape of $\Lambda(t)$ and checking parametric assumptions.

It is simplest to obtain nonparametric estimates of $\Lambda(t)$ by assuming that time is discrete; the estimates also apply to the continuous time case, as we discuss below. Thus, we suppose without loss of generality that t takes on values $1, 2, \ldots$, and let $n_i(t)$ be the number of events observed at time t for unit i. Letting $\tau = \max(\tau_1, \ldots, \tau_m)$, we can write $\log(L_T)$ as

$$l_T = \sum_{t=1}^{\tau} n_{\cdot}(t) \log \lambda(t) - \sum_{t=1}^{\tau} \sum_{i=1}^{m} \delta_i(t)\lambda(t) - \sum_{i=1}^{m} \log\{1 - \exp[-\Lambda(\tau_i)]\},$$

where $n_{\cdot}(t) = \sum_{i=1}^{m} \delta_i(t) n_i(t)$ and $\delta_i(t) = I(t \leq \tau_i)$ indicates whether $t \leq \tau_i$ is true or not. This gives

$$\frac{\partial l_T}{\partial \lambda(t)} = \frac{n_{\cdot}(t)}{\lambda(t)} - \sum_{i=1}^{m} \frac{\delta_i(t)}{1 - \exp[-\Lambda(\tau_i)]}, \quad t = 1, \ldots, \tau, \tag{3}$$

and

$$-\frac{\partial^2 l_T}{\partial \lambda(t) \partial \lambda(s)} = \frac{n_{\cdot}(t) I(s = t)}{\lambda(t)^2} - \sum_{i=1}^{m} \frac{\delta_i(t)\delta_i(s) \exp[-\Lambda(\tau_i)]}{\{1 - \exp[-\Lambda(\tau_i)]\}^2}. \tag{4}$$

Estimates $\hat{\lambda}_T(t)$, $t = 1, \ldots, \tau$, are obtained by solving the equations (3). This can in principle be done using Newton's method, but if τ is large it is simpler to use the iteration procedure

$$\tilde{\lambda}(t)^{(j+1)} = \frac{n_{\cdot}(t)}{\sum_{i=1}^{m} \frac{\delta_i(t)}{1 - \exp[-\tilde{\Lambda}^{(j)}(\tau_i)]}}, \tag{5}$$

where $\tilde{\lambda}(t)^{(j)}$ $(j = 1, 2, \ldots)$ is the j' th iterate towards $\hat{\lambda}_T(t)$, and we remember that $\Lambda(t) = \sum_{s=1}^{t} \lambda(s)$. We note that $\hat{\lambda}_T(t) = 0$ if $n_{\cdot}(t) = 0$, so (5) only has to be carried out for t values at which there is at least one event. Upon convergence (5) gives the maximum likelihood estimates $\hat{\lambda}_T(t)$ and corresponding mean function estimates $\hat{\Lambda}_T(t) = \sum_{s=1}^{t} \hat{\lambda}_T(s)$.

The procedure above also applies to continuous-time processes. In that case we merely let τ be large enough so that all distinct event times and τ_i values can be associated with one of $1, 2, \ldots, \tau$. The values $\hat{\lambda}_T(t)$ are zero except when t is an event time and so do not give a particularly attractive estimate of the continuous rate function $\lambda(t)$, but the mean function estimate $\hat{\Lambda}_T(t)$ is attractive and is in fact a nonparametric maximum likelihood estimator (*m.l.e.*) with typical properties of *m.l.e.*'s, as we discuss below. (More appealing nonparametric estimates of $\lambda(t)$ may be obtained by smoothing, but this is beyond the scope of the paper.)

The asymptotic distribution of $\hat{\Lambda}_T(t)$ can be obtained for the discrete time case, where $t = 1, 2, \ldots$. It follows from standard maximum likelihood large sample theory that, for $t \leq \lim_{m\to\infty} \tau$, conditional on $\{N_i(\tau_i) > 0, \tau_i : i = 1, \ldots, m\}$, and provided that $\lambda(s) > 0$ $(s = 1, \ldots, t)$ and that $\sum_{i=1}^m \delta_i(t) \to \infty$, a.s., as $m \to \infty$,

$$\sqrt{m} \begin{pmatrix} \hat{\lambda}_T(1) - \lambda(1) \\ \vdots \\ \hat{\lambda}_T(t) - \lambda(t) \end{pmatrix} \xrightarrow{d} N(0, \Sigma_{1t}^{(T)}), \quad as \;\; m \to \infty, \tag{6}$$

where $\xrightarrow{d}$ stands for convergence in distribution and $\Sigma_{1t}^{(T)^{-1}} = V_{1t}^{(T)}$ has entries (see (4))

$$\begin{aligned} v_1^{(T)}(u,s) &= plim_{m\to\infty} \frac{1}{m}\{-\frac{\partial^2 l_T}{\partial\lambda(t)\partial\lambda(s)}\} \\ &= plim_{m\to\infty} \frac{1}{m}\sum_{i=1}^m \frac{\delta_i(s)\delta_i(u)}{(1-e^{-\Lambda(\tau_i)})}[\frac{I(u=s)}{\lambda(s)} - \frac{e^{-\Lambda(\tau_i)}}{(1-e^{-\Lambda(\tau_i)})}]. \end{aligned} \tag{7}$$

It also follows that for $\hat{\Lambda}_T(t) = \sum_{s=1}^t \hat{\lambda}_T(s)$, we have

$$\sqrt{m}(\hat{\Lambda}_T(t) - \Lambda(t)) \xrightarrow{d} N(0, \sigma_{1t}^{(T)^2}), \quad as \;\; m \to \infty, \tag{8}$$

where $\sigma_{1t}^{(T)^2} = (1, \ldots, 1)\Sigma_{1t}^{(T)^{-1}}(1, \ldots, 1)'$.

The variance $\sigma_{1t}^{(T)^2}$ in (8) may be estimated, and approximate confidence limits for $\Lambda(t)$ obtained, by replacing $\lambda(s)$'s and $\Lambda(\tau_i)$'s in (7) with the estimated values. If t is large, however, inversion of $V_{1t}^{(T)}$ is problematic and alternative ways of estimating $\sigma_{1t}^{(T)^2}$ or obtaining confidence intervals are preferable. This area deserves further study, but we leave this to another occasion and make only two remarks here. The first is that some form of resampling inference is a possibility. The other is that for the purpose of testing or interval estimation, it is simpler to employ parametric models; if strong assumptions are to be avoided then a piecewise linear model for $\Lambda(t)$ (*i.e.*, a piecewise constant model for $\lambda(t)$) with a moderate number of sections is attractive and easily handled. Section 6 provides an example.

In the continuous time case the consistency and the limiting asymptotic normality stated in (8), and the stronger result that, conditional on τ_i's $(i = 1, \ldots, m)$, $\sqrt{m}(\hat{\Lambda}_T(t) - \Lambda(t))$ converges to a mean-zero Gaussian process over the time interval $(0, \tau^*)$, if $\tau \to \tau^*$ *a.s.* as $m \to \infty$, can be established under suitable conditions. A rigorous proof of these results will be given elsewhere. A thorough study of variance estimation for $\hat{\Lambda}(t)$ in the continuous case is technically difficult. The use of resampling methods or parametric models, suggested in the preceding paragraph, is once again attractive.

3 Estimation with Known Population Size and Truncation Distribution

In some situations the number of units and the observation time distribution $G(.)$ may be more or less known. For Example 1 in Section 1, the manufacturer might have a reasonably accurate estimate of sales over time. Similarly, for Example 2 car manufacturers usually have estimates of the distribution of mileage accumulation rates for the population of cars. This allows $G(.)$ to be estimated, as we illustrate in Section 6. We now consider estimation of $\lambda(t)$ and $\Lambda(t)$ when $G(.)$ and M are known.

3.1 Nonparametric maximum likelihood

Parametric models are readily fitted by maximizing the likelihood (9) below. We once again focus on nonparametric estimation of $\Lambda(t)$. The data consist of $n_i = N_i(\tau_i)$, the t_{ij}'s ($j = 1, \ldots, n_i$), and τ_i if $n_i \geq 1$, plus the knowledge of the number of units with $n_i = 0$. As in Section 2 we consider the discrete time case and assume for convenience that units $i = 1, \ldots, m$ are those with $n_i \geq 1$. The likelihood function is then

$$L_{TK} = \prod_{i=1}^{m}\{\prod_{j=1}^{n_i} \lambda(t_{ij}) \exp[-\Lambda(\tau_i)]dG(\tau_i)\} * \{\int_0^\infty \exp[-\Lambda(\tau)]dG(\tau)\}^{M-m}. \tag{9}$$

Defining $\delta_i(t) = I(t \leq \tau_i)$ as in Section 2, we can write $\log(L_{TK})$ as

$$\begin{aligned} l_{TK} &= \sum_{t=1}^{\tau} n_.(t)\log(\lambda(t)) - \sum_{t=1}^{\tau}\sum_{i=1}^{m} \delta_i(t)\lambda(t) \\ &+ (M-m)\log(\int_0^\infty \exp[-\Lambda(\tau)]dG(\tau)) + \sum_{i=1}^{m}\log(dG(\tau_i)). \end{aligned}$$

For the discrete time case, $dG(s) = \Pr(\tau_i = s)$, denoted by $g(s)$. We then have

$$\frac{\partial l_{TK}}{\partial \lambda(t)} = \frac{n_.(t)}{\lambda(t)} - \sum_{i=1}^{m} \delta_i(t) - (M-m)\frac{A(t)}{A(1)}, \quad \text{and} \tag{10}$$

$$-\frac{\partial^2 l_{TK}}{\partial \lambda(t)\partial \lambda(s)} = \frac{n_.(t)}{\lambda(t)^2} I(s=t) - (M-m)\{\frac{A(\max(s,t))A(1) - A(s)A(t)}{A(1)^2}\}, \tag{11}$$

when $t, s = 1, \ldots, \tau^*$, where $A(t) = \int_t^\infty \exp[-\Lambda(s)]dG(s) = \sum_{s=t}^{\tau^*} \exp[-\Lambda(s)]g(s)$, and $\tau^* = \sup\{s : g(s) > 0\}$ with $\bar{G}(s) = \int_s^\infty dG(u) = \sum_{u=s}^{\tau^*} g(u)$. Note that $A(1) = \Pr(n_i = 0)$ here.

A convenient algorithm for obtaining the $\hat{\lambda}_{TK}(t)$'s is

$$\tilde{\lambda}(t)^{(j+1)} = \frac{n_.(t)}{\sum_{i=1}^m \delta_i(t) + (M-m)\frac{\tilde{A}^{(j)}(t)}{\tilde{A}^{(j)}(1)}}, \tag{12}$$

where $\tilde{A}^{(j)}(t) = \sum_{s=t}^{\tau^*} \exp[-\tilde{\Lambda}^{(j)}(s)]g(s)$. Note that for $t > \tau = \max(\tau_1, \ldots, \tau_m)$, $\tilde{\lambda}(t)^{(j)} = 0, j = 1, \ldots$, thus, $\hat{\lambda}_{TK}(t) = 0$. And, for $t \le \tau$, $\tilde{A}^{(j)}(t)$ in (12) is

$$\sum_{s=t}^{\tau-1} \exp[-\tilde{\Lambda}^{(j)}(s)]g(s) + \exp[-\tilde{\Lambda}^{(j)}(\tau)]\bar{G}(\tau).$$

Provided that m and $\sum_{i=1}^m \delta_i(t) \to \infty$, a.s., and thus, $\tau \to \tau^*$ a.s., as $M \to \infty$, and assuming $\lambda(u) > 0 (s = 1, \ldots, \tau^*)$, for $t = 1, \ldots, \tau^*$,

$$\sqrt{M} \begin{pmatrix} \hat{\lambda}_{TK}(1) - \lambda(1) \\ \vdots \\ \hat{\lambda}_{TK}(t) - \lambda(t) \end{pmatrix} \xrightarrow{d} N(0, \Sigma_t^{(TK)}), \quad \text{as } M \to \infty, \tag{13}$$

where $V_t^{(TK)} = \Sigma_t^{(TK)^{-1}}$ has (u, s) entry (see (11)

$$\begin{aligned} v^{(TK)}(u,s) &= plim_{M\to\infty} \frac{1}{M}\{-\frac{\partial^2 l_{TK}}{\partial\lambda(t)\partial\lambda(s)}\} \\ &= \frac{I(u=s)\bar{G}(u)}{\lambda(u)} - \frac{A(\max(s,u))A(1) - A(u)A(s)}{A(1)}, \end{aligned} \tag{14}$$

where $u, s = 1, \ldots, t$.

In addition, for $\hat{\Lambda}_{TK}(t) = \sum_{s=1}^t \hat{\lambda}_{TK}(s)$, we have

$$\sqrt{M}(\hat{\Lambda}_{TK}(t) - \Lambda(t)) \xrightarrow{d} N(0, \sigma_t^{(TK)^2}), \tag{15}$$

where $\sigma_t^{(TK)^2} = (1, \ldots, 1)\Sigma_t^{(TK)}(1, \ldots, 1)'$.

The same remarks apply to estimation of $\sigma_t^{(TK)^2}$ as in the preceding section, and the same options are available.

For the continuous time case, the consistency and asymptotic normality of the estimator $\hat{\Lambda}_{TK}(t)$ will be proved elsewhere, along with the stronger result that $\sqrt{M}(\hat{\Lambda}_{TK}(t) - \Lambda(t))$ converges to a mean-zero Gaussian process over $(0, \tau^*)$ under suitable conditions.

In Example 2 of Section 1, if t represents mileage, it would be better to allow the τ_i's to come from different distributions. In particular, we could use $\tau_i \sim G_{a_i}(.)$ with $a_i = X - x_i$, since τ_i depends on the time since the car was sold. Suzuki and Kasashima (1992), in this context, give a nonparametric estimate for the distribution of mileage to the first failure or claim. If we assume $\tau_i \sim G_{a_i}(.)$ with $a_i \in \{1, \ldots, K\}$ and $G_{a_i}(.)'s$ known, the likelihood corresponding to (9) is

$$L_{TK} = \prod_{i=1}^{m} \{\prod_{j=1}^{n_i} \lambda(t_{ij}) \exp[-\Lambda(\tau_i)] dG_{a_i}(\tau_i)\} \times \prod_{k=1}^{K} \{\int_0^{\infty} \exp[-\Lambda(\tau)] dG_k(\tau)\}^{m_k}, \qquad (16)$$

where m_k is the size of $\{i: \ a_i = k, \text{ and } N_i(\tau_i) = 0\}$. We can obtain the *m.l.e.* of $\Lambda(t)$ for both discrete and continuous time cases by a straightforward extension of the methods in this section.

3.2 Information added by knowing M and $G(.)$

To compare the truncated data likelihood (1) and the more informative likelihood (9), we could compute the expected information matrix for each case, noting that in both cases we would need to take expectation with respect to the distribution of the τ_i's as well as the Poisson distribution of events. We have instead opted for simulation in order to compare the nonparametric estimates from the two likelihoods. This allows us to examine and compare the estimators in a finite sample, not asymptotic, context and avoids the inversion of large information matrices for obtaining asymptotic variances.

We do two simulations, all with events occurring according to a homogeneous Poisson process. Our objective is to check and compare the behaviors of the estimators $\hat{\Lambda}_T(t)$ and $\hat{\Lambda}_{TK}(t)$. We also compare them with the estimator $\hat{\Lambda}_c(t)$, which would be obtained if we knew every τ_i, $(i = 1, \ldots, M)$, and not just those for which $n_i \geq 1$. In this case, $\hat{\Lambda}_c(t)$ is the most efficient nonparametric estimator possible. The simulations are as follows.

1. We used $M = 150$, and generated τ_i, $i = 1, \ldots, 150$, from the uniform distribution on $(0, 300]$, and 150 corresponding time homogeneous Poisson processes $N_i(t)$, $i = 1, \ldots, 150$, with rate $\lambda = 0.01$. Here the expected value of m is around 102. Based on the generated data, we obtained estimates $\hat{\Lambda}_T(t)$, $\hat{\Lambda}_{TK}(t)$, and $\hat{\Lambda}_c(t)$, which are respectively the nonparametric *m.l.e.*'s of $\Lambda(t)$ based on L_T, L_{TK} and L_c, where L_c is the censored data likelihood

$$\begin{aligned} &\textstyle\prod_{\delta_i=1} \{\prod_{j=1}^{n_i} \lambda(t_{ij}) \exp[-\Lambda(\tau_i)]\} dG(\tau_i) \times \prod_{\delta_i=0} \exp[-\Lambda(\tau_i)] dG(\tau_i) \\ \propto \ &\textstyle\prod_{\delta_i=1} \{\prod_{j=1}^{n_i} \lambda(t_{ij}) \exp[-\Lambda(\tau_i)]\} \times \prod_{\delta_i=0} \exp[-\Lambda(\tau_i)]. \qquad (17) \end{aligned}$$

Note that $\hat{\Lambda}_c(t) = \sum_{s=1}^{t} \hat{\lambda}_c(s) = \sum_{s=1}^{t} \frac{n_{.}(t)}{\sum_{i=1}^{M} \delta_i(t)}$ is the Nelson-Aalen estimator in this situation (see Andersen and Borgan 1985). We repeated the simulation $n = 100$ times, and give the sample means of the estimates in Fig.1 (a1), and the corresponding sample mean square errors in Fig.1 (a2).

2. We took $\lambda = 0.001$ and kept the other aspects of the above simulation unchanged. Now, the expected value of m is around 21. Figs 1 (b1) and (b2) present the sample means and the sample mean square errors of the three estimators of $\Lambda(t)$.

From Fig.1, we see that all of the estimators for $\Lambda(t)$ are essentially unbiased, except for some slight positive bias in $\Lambda_T(t)$ when m is small (Fig.1 (b1)). The estimators based on L_T and L_{TK} have similar mean square errors in simulation 1, where truncation is fairly light ($\lambda = 0.01, M = 150, E(m) \doteq 102$), and are about as efficient as $\hat{\Lambda}_c(t)$ based on (17). In simulation 2, where truncation is much heavier ($\lambda = 0.001, M = 150, E(m) \doteq 21$), the information added by knowing M and $G(.)$ plays an important role: L_T is relatively much less informative, but $\hat{\Lambda}_{TK}(t)$ is still about as efficient as $\hat{\Lambda}_c(t)$. Obviously the heavier truncation is, the more important knowledge of M and $\bar{G}(.)$ is.

Notice that $\hat{\Lambda}_T(t)$ is defined over $(0, \tau]$ with $\tau = \max(\tau_1, \ldots, \tau_m)$, and $\hat{\Lambda}_c(t)$ over $(0, \tau_*]$ with $\tau_* = \max(\tau_1, \ldots, \tau_M)$ while $\hat{\Lambda}_{TK}(t)$ is over $(0, \tau^*]$ with $\tau^* = \sup\{s : \bar{G}(s) > 0\}$. We know $\tau \leq \tau_* \leq \tau^*$ generally. For our simulations, the sample means, as well as the sample mean square errors, of $\hat{\Lambda}_T(t)$, $\hat{\Lambda}_c(t)$, and $\hat{\Lambda}_{TK}(t)$ are, respectively, over $(0, \min_k(\tau^{(k)})]$, $(0, \min_k(\tau_*^{(k)})]$ and $(0, \tau^*]$ with $\tau^* = 300$, where $\tau^{(k)}$ and $\tau_*^{(k)}$ are the observed values of τ and τ_* at the k^{th} simulation, respectively. This explains why the curves in the figures for $\hat{\Lambda}_T(t)$ are the shortest, and for $\hat{\Lambda}_c(t)$ are the second shortest.

We remark that means and mean square errors for a fourth estimator (denoted SM in the figures) which is introduced in Section 4, are also shown in Fig.1.

4 Robust Estimation

4.1 A robust estimator

We consider another estimator which can be used if M and $G(.)$ are known. Since

$$n_{.}(t) - [\sum_{i=1}^{M} \delta_i(t)]\lambda(t) = 0 \tag{18}$$

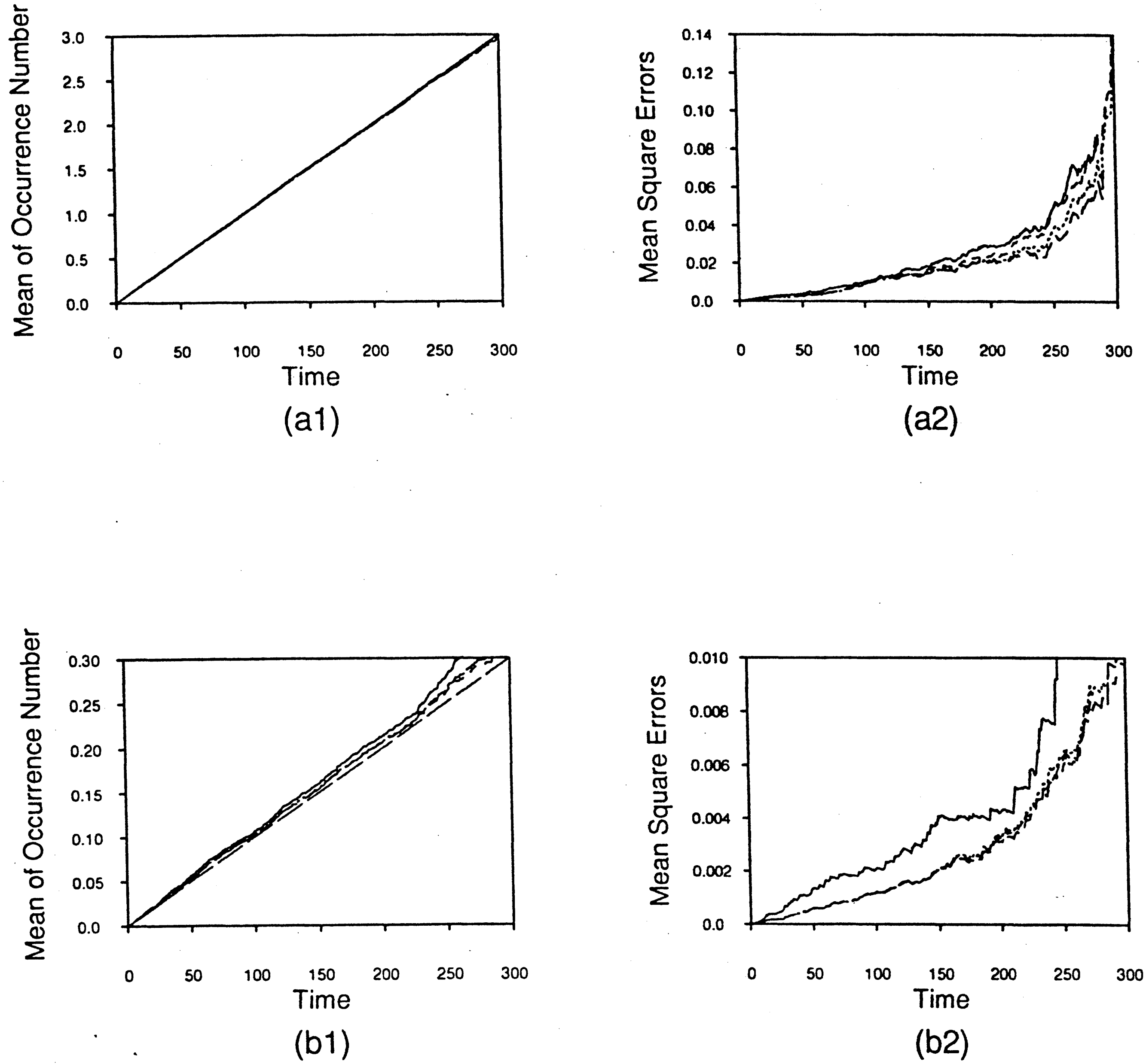

Fig.1. Simulation with $M = 150$, $G(.) = U(0,300)$, and $(N(t), t > 0)$ being a time homogeneous Poisson process, where $E\{N(t)\} = \lambda t$: (a1) plot of sample means of the estimates for the simulation with $\lambda = 0.01$ ($E\{m\} \doteq 102$) $n = 100$ times repeated; (a2) plot of the sample mean square errors of the estimates corresponding to (a1); (b1) plot of sample means of the estimates for the simulation with $\lambda = 0.001$ ($E\{m\} \doteq 21$) $n = 100$ times repeated; (b2) plot of the sample mean square errors of the estimates corresponding to (b1). (———, $\hat{\Lambda}_T(t)$; .... , $\hat{\Lambda}_{TK}(t)$; – – – –, $\hat{\Lambda}_{SM}(t)$; ————, $\hat{\Lambda}_c(t)$; —— ——, $\Lambda(t)$)

is a unbiased estimating equation (*i.e.*, the left side has expectation 0) for each $t = 1, 2, \ldots$. provided observation times are independent of the counting processes, we see

$$n_{\cdot}(t) - M\bar{G}(t)\lambda(t) = 0 \tag{19}$$

is also unbiased by noting $E\{\delta_i(t)\} = \bar{G}(t) = \Pr(\tau_i \geq t)$, $t = 1, \ldots$. Solving (19), we get

$$\hat{\lambda}_{SM}(t) = \frac{n_{\cdot}(t)}{M\bar{G}(t)} \tag{20}$$

and $\hat{\Lambda}_{SM}(t) = \sum_{s=1}^{t} \hat{\lambda}_{SM}(s)$, for $t = 1, \ldots, \tau^*$. Note that (19) is valid even if the counting process is not Poisson but $E\{n_i(t)\} = \lambda(t)$, and that $\hat{\lambda}_{SM}(t)$ is in fact unbiased.

Under mild conditions, we can prove (see Appendix A) that $\sqrt{M}(\hat{\Lambda}_{SM}(t) - \Lambda(t)) \xrightarrow{d} N(0, \sigma_t^{(SM)^2})$, $t = 1, \ldots, \tau^*$, with

$$\sigma_t^{(SM)^2} = (1, \ldots, 1)C_t^{-1}D_tC_t^{-1}(1, \ldots, 1)', \tag{21}$$

where $C_t = diag\{\bar{G}(1), \ldots, \bar{G}(t)\}$, and D_t is a $t \times t$ matrix with entries $\lim_{M\to\infty} \frac{1}{M}\text{Cov}\{n_{\cdot}(s), n_{\cdot}(u)\}, s = 1, \ldots, t; u = 1, \ldots, t$. If different individuals have common covariance structure, $D_t = (v(s,u)^*)_{t\times t}$ with $v(s,u)^* = \text{Cov}\{\delta_i(s)n_i(s), \delta_i(u)n_i(u)\}$, which can be estimated consistently by

$$\hat{v}(s,u)^* = \frac{1}{M}\sum_{i=1}^{M}\{\delta_i(s)n_i(s) - \bar{G}(s)\hat{\lambda}_{SM}(s)\}\{\delta_i(u)n_i(u) - \bar{G}(u)\hat{\lambda}_{SM}(u)\}.$$

Note that this variance estimator is suitable regardless of whether the event processes are Poisson.

From (20), we have

$$\text{Var}\{\hat{\Lambda}_{SM}(t)\} = \sum_{i=1}^{M}\sum_{s=1}^{t}\sum_{u=1}^{t}\frac{\text{Cov}(\delta_i(s)n_i(s), \delta_i(u)n_i(u))}{M^2\bar{G}(s)\bar{G}(u)}$$

and if different individuals have common covariance structure,$\text{Var}\{\hat{\Lambda}_{SM}(t)\}$ is consistently estimated by using $\hat{v}(s,u)^*$ to give

$$\hat{V}_{SM}(t) = \sum_{i=1}^{M}\{\sum_{s=1}^{t}\frac{\delta_i(s)n_i(s) - \bar{G}(s)\hat{\lambda}_{SM}(s)}{M\bar{G}(s)}\}^2. \tag{22}$$

$\hat{\Lambda}_{SM}(t)$ and $\hat{V}_{SM}(t)$ can be written in integral form:

$$\hat{\Lambda}_{SM}(t) = \int_0^t \frac{dN_{\cdot}(s)}{M\bar{G}(s)}, \quad \hat{V}_{SM}(t) = \sum_{i=1}^{M}\{\int_0^t \frac{1}{M\bar{G}(s)}[\delta_i(s)dN_i(s) - \frac{dN_{\cdot}(s)}{M}]\}^2,$$

with $dN_.(s) = n_.(s)ds$. These expressions define valid nonparametric estimators in the case of continuous time processes.

In Fig.1, we also show the sample means and mean square errors of $\hat{\Lambda}_{SM}(t)$ based on the data generated in the simulation study of Section 3. We observe that $\hat{\Lambda}_{SM}(t)$ is also essentially unbiased and about as efficient as $\hat{\Lambda}_{TK}(t)$ for the situations considered, and is defined over $(0, \tau^*]$, where $\tau^* = 300$. In fact, from the formula of $\hat{\lambda}_{SM}(t)$, (20), we note that $\hat{\Lambda}_{SM}(t)$ is very close to $\hat{\Lambda}_c(t)$ when M is large enough.

To check the robustness of the various estimators when the recurrent events do not come from identical Poisson processes, we consider a mixed Poisson model (see Lawless 1987). Under the model, each unit i has an associated random variable α_i such that events for it occur according to a Poisson process with rate function $\alpha_i\lambda(t)$. The α_i's have a gamma distribution with mean 1, so that $E\{N_i(t)\} = \Lambda(t)$ still holds. We did two simulations as follows.

1. τ_i, $i = 1, \ldots, 150$, are generated as earlier from $U(0, 300)$; α_i, $i = 1, \ldots, 150$, are from a gamma distribution with mean one and variance one. The 150 counting processes $N_i(t)$, $i = 1, \ldots, 150$, are then independent homogeneous Poisson processes with rate function $\lambda_i(t) = \alpha_i\lambda(t) = 0.01\alpha_i$. In Figs 2 (a1)-(a2), we present the sample means and sample mean square errors of $\hat{\Lambda}_T(t)$, $\hat{\Lambda}_{TK}(t)$, $\hat{\Lambda}_{SM}(t)$, and $\hat{\Lambda}_c(t)$ obtained by repeating the simulation $n = 100$ times.

2. We generated data in the same way as for simulation 1 except that α_i, $i = 1, \ldots, 150$ were from a gamma distribution with mean one and variance 0.2. Figs 2 (c1)-(c2), respectively, give the sample means and sample mean square errors of the estimates for $n = 100$ simulations.

The above two simulations were each repeated with $\lambda(t) = 0.001$. The corresponding results are shown in Figs 2 (b1)-(b2), and Figs 2 (d1)-(d2).

We remark that the mean and variance of $N_i(t)$ under the mixed Poisson model are $\Lambda(t)$ and $\Lambda(t)[1 + \text{Var}(\alpha_i)\Lambda(t)]$, respectively. The larger either $\text{Var}(\alpha_i)$ or $\Lambda(t)$ is, the more overdispersed the $N_i(t)$'s are relative to the Poisson distribution.

In the figures for sample means of the estimates, we see corroboration of the fact that $\hat{\Lambda}_{SM}(t)$ and $\hat{\Lambda}_c(t)$ are unbiased, but that $\hat{\Lambda}_T(t)$ is biased. The bias depends on how overdispersed $N_i(t)$ is, and can be substantial when overdispersion is large. Figs 2 (a2), (b2), (c2), and (d2) show a similar efficiency for $\hat{\Lambda}_{SM}(t)$ and $\hat{\Lambda}_c(t)$. The behavior of $\hat{\Lambda}_{TK}(t)$, which is

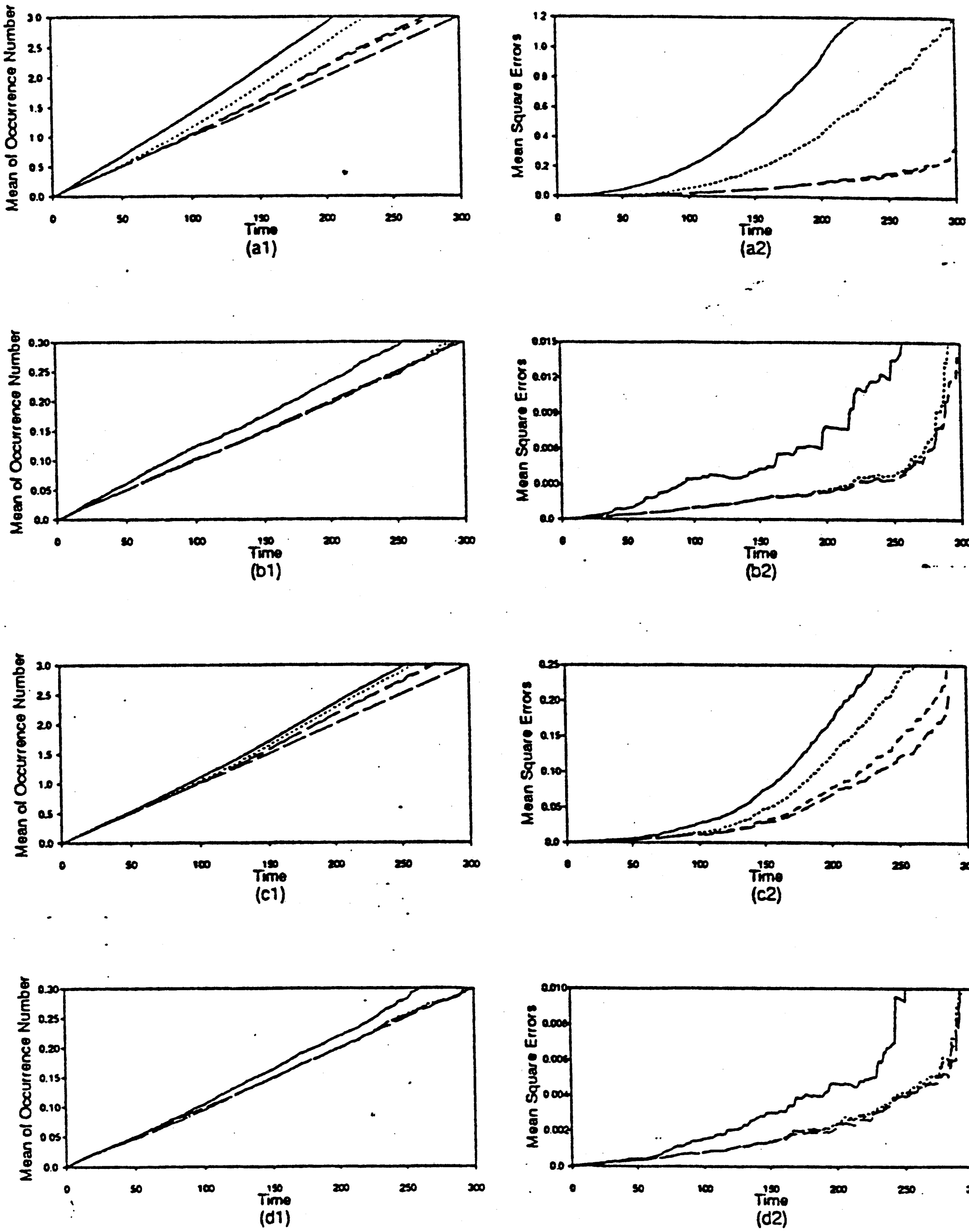

Fig.2. Simulation with $M = 150$, $G(.) = U(0,300)$, and $(N(t), t > 0)$ being a mixed Poisson process, where $E\{N(t)|\alpha\} = \lambda t\alpha$, $\alpha \sim \Gamma(a, \sigma^2)$, and $E\{\alpha\} = a, V\{\alpha\} = \sigma^2$: (a1) plot of sample means of the estimates for the simulation with $\lambda = 0.01, a = \sigma = 1$ ($E\{m\} \doteq 80$) $n = 100$ times repeated; (a2) plot of the sample mean square errors of the estimates corresponding to (a1); (b1) and (b2) plot of the results given by the simulation corresponding to (a1) with $\lambda = 0.001$ instead ($E\{m\} \doteq 19$); (c1) and (c2) plot of the results given by the simulation corresponding to (a1) with $\sigma^2 = 0.2$ instead ($E\{m\} \doteq 97$); (d1) and (d2) plot of the results given by the simulation corresponding to (c1) with $\lambda = 0.001$ instead ($E\{m\} \doteq 20$). (———— ,$\hat{\Lambda}_T(t)$; .... ,$\hat{\Lambda}_{TK}(t)$; − − − − , $\hat{\Lambda}_{SM}(t)$; ———,$\hat{\Lambda}_c(t)$; —— ——, $\Lambda(t)$)

based on the Poisson assumption, is apparent. When truncation is heavy (giving small m) it is more or less unbiased and is comparable to $\hat{\Lambda}_{SM}(t)$ and $\hat{\Lambda}_c(t)$ (see Figs 2 (b1)-(b2) and (d1)-(d2)), but when truncation is light (giving large m) it is biased and has considerably larger mean square error than $\hat{\Lambda}_{SM}(t)$ and $\hat{\Lambda}_c(t)$ (see Figs 2 (a1)-(a2) and (c1)-(c2)). This is discussed in Appendix B.

4.2 Effect of imprecise information about $G(.)$ and M

If we know $G(.)$ and M, much more efficient estimation of $\Lambda(t)$ is possible than if we have only zero-truncated data, especially when truncation is heavy. In addition, the estimators given in Section 3.1 and 4.1, which are based on known M and $G(.)$, are robust to departures from a Poisson process of events, although in the case of $\hat{\Lambda}_{TK}(t)$ the robustness fails when truncation is light. We now examine the effect of misspecifying either $G(.)$ or M, since they will often be known only approximately.

First, suppose that upon our knowledge the total number of units is M^*, but the true value is M_0. This does not affect the estimator $\hat{\lambda}_T(t)$, but affects the others. Now,

$$\hat{\lambda}_c(t) = \frac{\frac{n.(t)}{M_0}\frac{M_0}{M^*}}{\frac{1}{M^*}\sum_{i=1}^{M^*}\delta_i(t)}, \quad \hat{\lambda}_{SM}(t) = \frac{\frac{n.(t)}{M_0}}{\frac{M^*}{M_0}\bar{G}(t)}, \quad \text{and}$$

$$\hat{\lambda}_{TK}(t) = \frac{\frac{n.(t)}{M_0}}{\frac{1}{M_0}\sum_{i=1}^{m}\delta_i(t) + (\frac{M^*}{M_0} - \frac{m}{M_0}\frac{\hat{\Lambda}_{TK}(t)}{\hat{\Lambda}_{TK}(1)})},$$

are then not consistent even under the Poisson model. However, if $\frac{M^*}{M_0}$ is close to 1, or, if we think of M^* an approximately unbiased estimate of M_0 with small variance, then $\hat{\Lambda}_c(t)$, $\hat{\Lambda}_{SM}(t)$, and $\hat{\Lambda}_{TK}(t)$ are approximately unbiased and efficient.

Over- or under-estimation of M_0 will result in under- or over-estimation of the $\lambda(t)$'s in any given situation. To assess the average behavior of the estimates, we did the following simulations. $M_{0j}, j = 1, \ldots, n = 100$ were independently generated from $N(150, 5^2)$. For each j, $N_{ji}(t)$'s, $i = 1, \ldots, M_{0j}$, were independent time homogeneous Poisson processes with $\lambda = 0.01, 0.001$, respectively; τ_{ji}'s were from $U(0, 300]$. We took M equal to $M^* = 150$ to obtain the estimates based on each simulated data set ($j = 1, \ldots, 100$). Fig.3 presents the sample means and mean square errors of the estimates for the two cases. These figures show that, in such a situation, all estimates are roughly unbiased, and $\hat{\Lambda}_c(t)$, $\hat{\Lambda}_{SM}(t)$, and $\hat{\Lambda}_{TK}$ still behave similarly and are more efficient than $\hat{\Lambda}_T(t)$ in the case with $\lambda = 0.001$. In the case $\lambda = 0.01$, where $m \doteq 102$, $\hat{\Lambda}_T(t)$ works almost as well as $\hat{\Lambda}_c(t)$ does while the other two are a little less efficient.

Misspecification of $G(.)$ can also be investigated. If we know M, but assume the distribution of observation time is $G^*(.)$ while the true one is $G_0(.)$, $\hat{\lambda}_T(t)$ and $\hat{\lambda}_c(t)$ are unaffected since they do not depend on $G(.)$. However, we note that

$$E\{\frac{\partial l_{TK}}{\partial\lambda(t)}\} = M\Pr_0(N_i(\tau_i) = 0)[\frac{\partial \log \Pr_0(N_i(\tau_i) = 0)}{\partial\lambda(t)} - \frac{\partial \log \Pr^*(N_i(\tau_i) = 0)}{\partial\lambda(t)}]$$

does not in general equal zero if $G^*(.) \neq G_0(.)$, nor does the expectation of (19),

$$E\{\sum_{i=1}^{M} \delta_i(t)n_i(t) - \bar{G}^*(t)\lambda(t)\} = M\lambda(t)[\bar{G}_0(t) - \bar{G}^*(t)].$$

The estimates $\hat{\Lambda}_{TK}(t)$ and $\hat{\Lambda}_{SM}(t)$ would then be inconsistent. We can investigate the extent of the bias in the estimators; this will not be great if $G^*(.)$ is a good approximation to $G_0(.)$.

To examine the effect of misspecification of $G(.)$, we performed the following simulations. We generated τ_i from $N(150, 70^2) = G_0(.)$, and $N_i(t)$ in the same way as the simulations in Section 3, for $i = 1, \ldots, 150$. Estimates of $\Lambda(t)$ were obtained by taking $G(.)$ as $G^*(.) = U(0, 300]$. The simulations were repeated 100 times, respectively. Fig.4 gives the sample means and sample mean square errors of the estimators. We observe that $\hat{\Lambda}_{TK}(t)$ and $\hat{\Lambda}_{SM}(t)$ are not badly biased, and their efficiencies are similar to $\hat{\Lambda}_c(t)$'s, especially when truncation is heavy (see Figs 4 (b1)-(b2)).

The methods in this paper assume that $G(.)$ and M are known exactly, but $G(.)$, and sometimes M, is usually estimated from supplementary data, as in the example in Section 6. If the estimates are fairly precise there is little harm in using the variance estimates for $\hat{\Lambda}_{TK}(t)$ and $\hat{\Lambda}_{SM}(t)$ presented in Section 3 and 4. More generally, we may add a term to the variance estimates of the form $\hat{\mathrm{Var}}_{\hat{G}}\{\hat{\Lambda}(t)|\hat{G}\}$ to account for the imprecision of $\hat{G}(.)$. An alternative procedure that is adequate for most practical purposes is to check on the sensitivity of estimates $\hat{\Lambda}(t)$ and variance estimates $\hat{\mathrm{V}}(t)$ in Sections 3 and 4 to variations in $G(.)$ or M, and to use this to modify confidence limits for $\Lambda(t)$ in an informal way.

5 Data Aggregated Across Units

Sometimes there is no linkage of events for individual units, in which case all we observe is the pair (t, τ) for each observed event. For example, warranty claims may be recorded

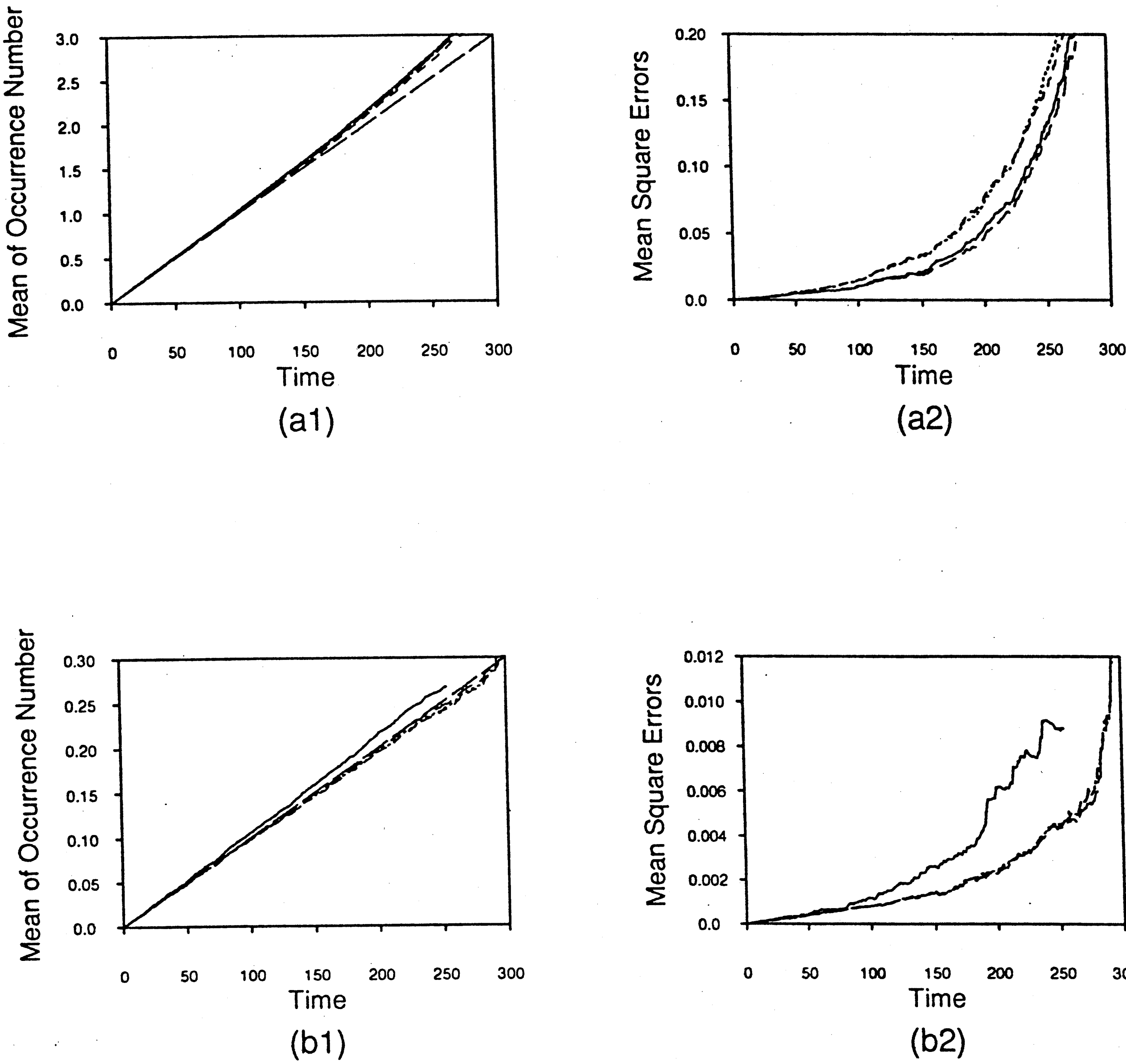

Fig.3. Simulation with $M \sim N(150, 5^2)$, $G(.) = U(0, 300)$, and $(N(t), t > 0)$ being a time homogeneous Poisson process, where $E\{N(t)\} = \lambda t$: (a1) plot of sample means of the estimates for the simulation with $\lambda = 0.01$ ($E\{m\} \doteq 102$) $n = 100$ times repeated; (a2) plot of the sample mean square errors of the estimates corresponding to (a1); (b1) and (b2) plot of the results given by the simulation corresponding to (a1) with $\lambda = 0.001$ instead ($E\{m\} \doteq 21$). (______ ,$\hat{\Lambda}_T(t)$; .... ,$\hat{\Lambda}_{TK}(t)$; – – – – , $\hat{\Lambda}_{SM}(t)$; ———,$\hat{\Lambda}_c(t)$; ____ ____, $\Lambda(t)$)

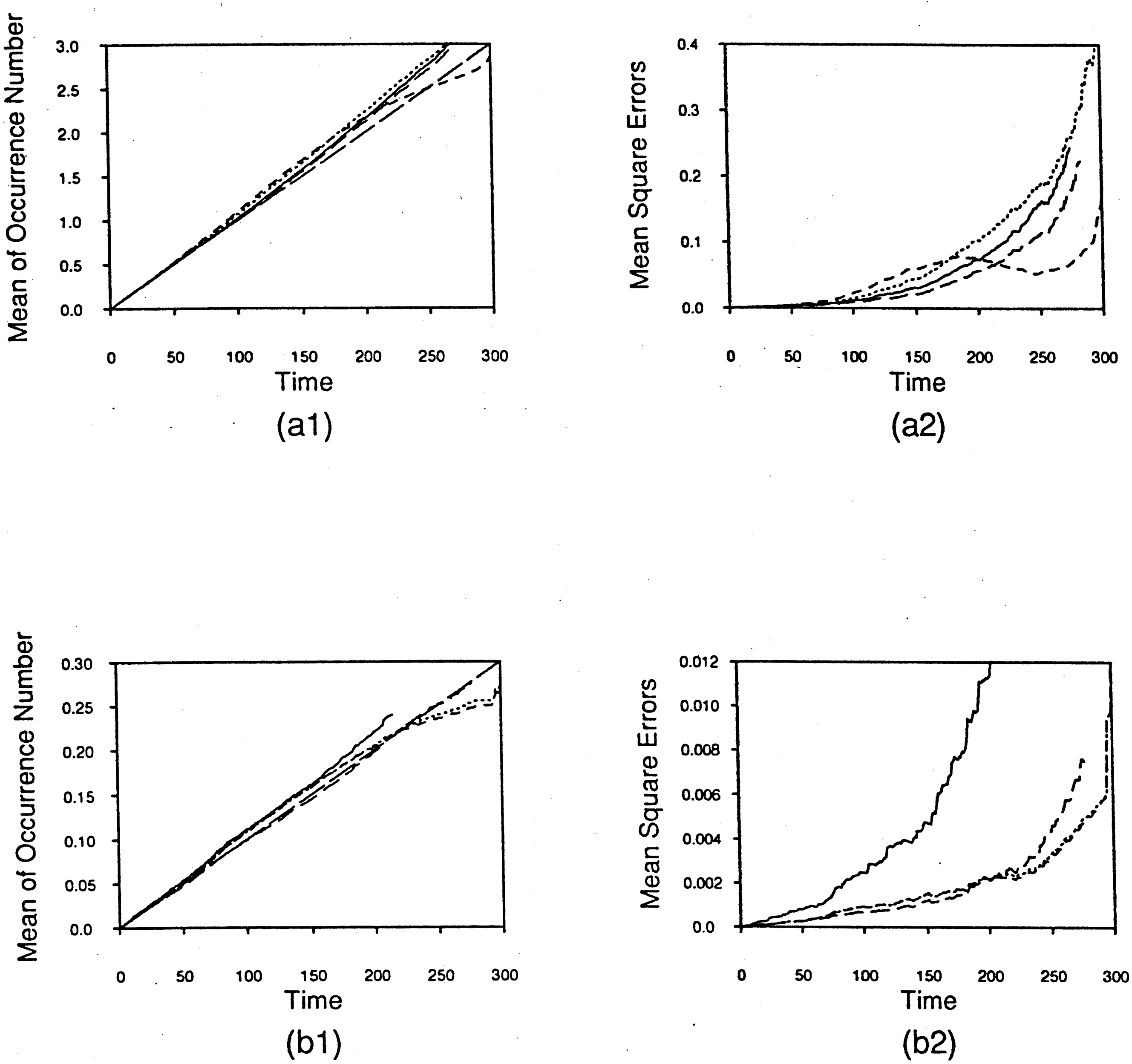

Fig.4. Simulation with $M = 150$, $G(.) = N(150, 70^2)$, and $(N(t), t > 0)$ being a time homogeneous Poisson process, where $E\{N(t)\} = \lambda t$: (a1) plot of sample means of the estimates for the simulation with $\lambda = 0.01$ ($E\{m\} \doteq 107$) $n = 100$ times repeated; (a2) plot of the sample mean square errors of the estimates corresponding to (a1); (b1) and (b2) plot of the results given by the simulation corresponding to (a1) with $\lambda = 0.001$ instead ($E\{m\} \doteq 22$). (———— ,$\hat{\Lambda}_T(t)$; .... ,$\hat{\Lambda}_{TK}(t)$; - - - - , $\hat{\Lambda}_{SM}(t)$; ————,$\hat{\Lambda}_c(t)$; —— —— , $\Lambda(t)$)

according to the time of the claim and the time at which the product unit was sold, with no historical record kept for each unit. In this case, if M and $G(.)$ are unknown, we are able to obtain only the likelihood function based on $\Pr(t_l^*|t_l^* \leq \tau_l^*, \tau_l^*)$ which, assuming events occur according to a Poisson process, is

$$L_{T_2} = \prod_{l=1}^{n} \frac{\lambda(t_l^*)}{\Lambda(\tau_l^*)} = \prod_{i=1}^{m} \prod_{j=1}^{n_i} \frac{\lambda(t_{ij})}{\Lambda(\tau_i)}, \tag{23}$$

where $n = \sum_{i=1}^{m} n_i$ is the total number of events across all units, and (t_l^*, τ_l^*)'s are the observed (t, τ) values.

From (23) it is clear that we can estimate only $\frac{\lambda(t)}{\Lambda(\tau)}$, where $\tau = \max(\tau_l^*)$, and not the absolute values of the $\lambda(t)$'s or $\Lambda(t)$'s. In fact, the estimation problem is equivalent to one for truncated failure time data (*e.g.*, see Kalbfleisch and Lawless 1992), and the nonparametric estimate from (23) can be given in closed form as follows.

Denote $\frac{\lambda(t)}{\Lambda(t)}$ by $\lambda^*(t)$. Then, for $t \leq \tau$,

$$\frac{\lambda(t)}{\Lambda(\tau)} = \lambda^*(t) \prod_{x=t+1}^{\tau} [1 - \lambda^*(x)].$$

We can rewrite (23) as

$$L_{T_2} = \prod_{l=1}^{n} \frac{\lambda^*(t_l^*)}{\Lambda^*(\tau_l^*)} = \prod_{t=1}^{\tau} \lambda^*(t)^{n(t)}[1 - \lambda^*(t)]^{n^*(t)}, \tag{24}$$

where $n(t) = \#\{l : t_l^* = t\} = n_.(t)$, and $n^*(t) = \#\{l : t_l < t \leq \tau_l\}$. The *m.l.e.*of $\lambda^*(t)$ based on (24) is

$$\hat{\lambda}^*(t) = \frac{n_.(t)}{n_.(t) + n^*(t)}.$$

If M and $G(.)$ are known, then, under the discrete Poisson process assumption, the data consisting of the pairs (t_l^*, τ_l^*) give the likelihood function

$$L_{T_{k2}} = \{\prod_{l=1}^{n} \lambda(t_l^*) g(\tau_l^*)\} \exp\{-M \sum_{s=1}^{\tau} \sum_{t=1}^{s} \lambda(t) g(s)\},$$

which is proportional to

$$\prod_{t=1}^{\tau} \lambda(t)^{n_.(t)} \exp\{-M \sum_{t=1}^{\tau} \lambda(t) \bar{G}(t)\}. \tag{25}$$

Maximization of (25) gives the estimates

$$\hat{\lambda}(t) = \frac{n_.(t)}{M \bar{G}(t)},$$

the same as $\hat{\lambda}_{SM}(t)$ in Section 4.

6 An Example

We will illustrate the methods introduced above by estimating the repair occurrence rate for a system on a particular car model, based on real warranty data. We consider a group of 8,394 cars manufactured over a two month period; this is a subset of a larger group discussed by Kalbfleisch, Lawless and Robinson (1991).

As of the final data base update all $M = 8,394$ cars had been sold. They generated a total of 1134 claims from 831 different cars. (For the discussion here, claims before the cars were sold are ignored.) Let δ_i equal 1 if car i had a warranty claim and 0 otherwise, and $N_i(t)$ indicate the total number of claims from car i up to time t, where t can be either age (*i.e.*, days since the car was sold) or mileage of the car. Denote (a_{ij}, m_{ij}) as the age and mileage of car i at the j^{th} claim. If the observation time period for car i is denoted by $(0, \tau_i]$, we have $\{\tau_i, (a_{ij}, m_{ij} : j = 1, \ldots, N_i(\tau_i))\}$ as the data for car i, provided $\delta_i = 1$, *i.e.*, there is at least one claim. Otherwise all we know is the date of sale of the car and that there was no claim from it. The age and mileage limits of the warranty plan are one year and 12,000 miles, *i.e.*, $A_0 = 365$ and $W_0 = 12,000$. These and the rates at which cars accumulate miles determine the τ_i's. In the following, we will obtain estimates of $\Lambda(t) = E\{N_i(t)\}$, the expected cumulative number of repairs per car as a function of age. Rates and means as a function of mileage could also be given. Information about the accumulation of mileage in the population of cars is available from customer surveys; this allows us to estimate $G(.)$.

We assume that the mileage accumulation rate for car i is u_i, so that $m_i(a) = u_i a$ is the mileage at age a. Although this ignores fluctuations in mileage accumulation it is a reasonably practical assumption for many cars in their first two or three years. With a one year/12,000 miles warranty the end of observation time for car i is $\tau_i = \min(X - x_i, 365, \frac{12000}{u_i})$, where age is measured in days, x_i is the day the car was sold, and X is the day of the final data update. For these cars $X - x_i$ exceeded 365 days for 7356 out of 8394 cars.

Customer surveys had been carried out for cars of the same type as in the warranty data base, in which the mileage at one year was obtained for each car sampled. We use this supplementary information to estimate $\bar{G}_2(t) = \Pr\{\frac{12000}{u_i} \geq t\}$ and then estimate $\bar{G}(t) = \Pr\{\tau_i \geq t\}$ as $\hat{\bar{G}}_2(t)$ times $\hat{\bar{G}}_1(t) = \frac{\sum_{i=1}^{M} I[365 \bigwedge (X - x_i) \geq t]}{M}$, an estimate of $\bar{G}_1(t) = \Pr\{365 \bigwedge (X - x_i) \geq t\}$, where $a \bigwedge b$ is used to denote $\min(a, b)$. (We could stratify cars according to their dates of sale for the sake of precision, but it turns out that this gives almost exactly the same estimates as below; hence we present the slightly simpler unstratified analysis.)

The estimated $\bar{G}(t)$ is shown in Fig.5. Note that $\hat{\bar{G}}(365) = 0.53$, but that $\hat{\bar{G}}(t) = 0$ for

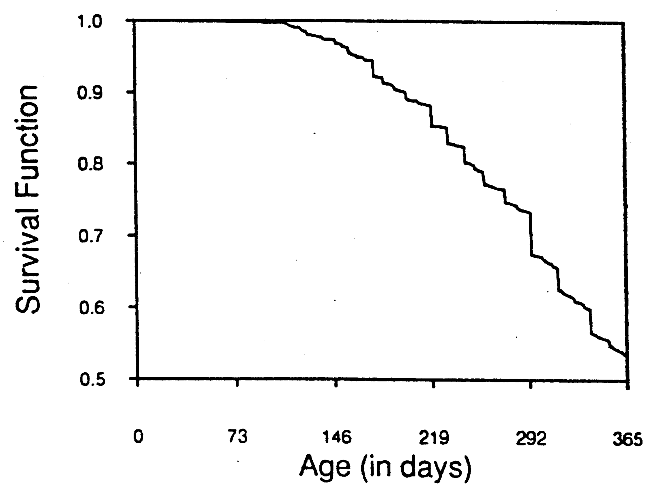

Fig.5. Plot of the estimated $\bar{G}(t)$, $t \in [0, 365]$.

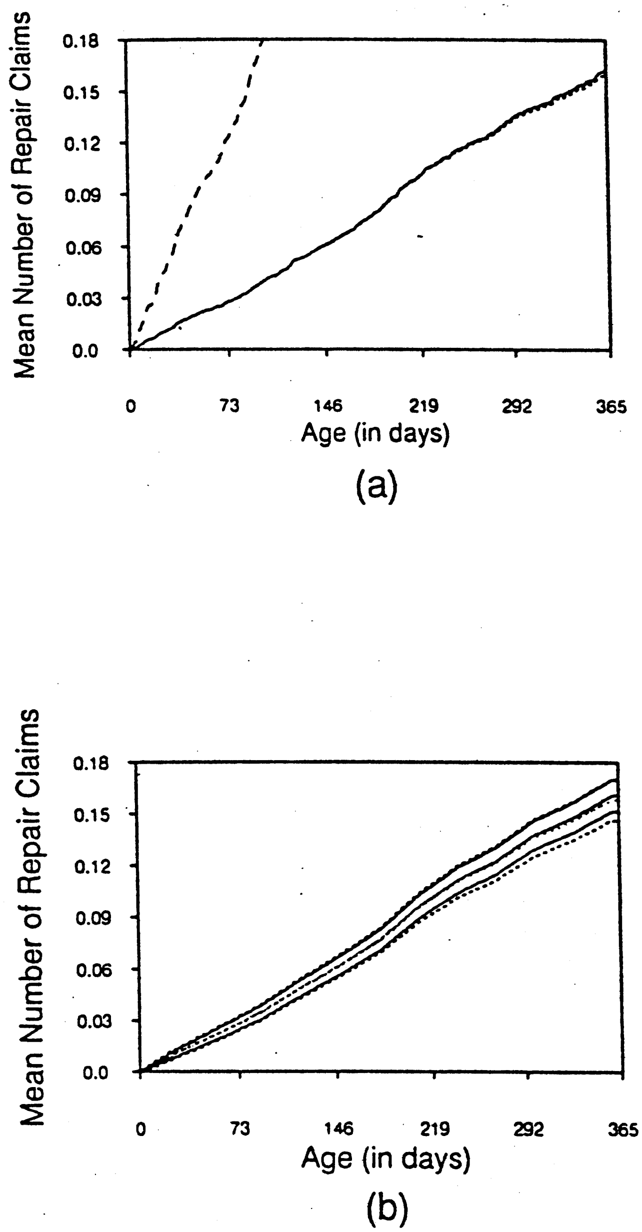

Fig.6. (a) Plot of the estimates for $\Lambda(t)$, the mean of car repair claims, based on the data in the example (———— ,$\hat{\Lambda}_{TK}(t)$; .... ,$\hat{\Lambda}_{SM}(t)$; – – – – , $\hat{\Lambda}_{T}(t)$); (b) plot of the pointwise approximate 95% confidence intervals for the mean of car repair claims based on the data in the example. (———— , $\hat{\Lambda}_{TK}(t) \pm 1.96\hat{V}_{TK}(t)^{\frac{1}{2}}$; .... , $\hat{\Lambda}_{SM}(t) \pm 1.96\hat{V}_{SM}(t)^{\frac{1}{2}}$)

$t > 365$; we show $\hat{\hat{G}}(t)$ only for $t \leq 365$.

We obtained the estimate $\hat{\Lambda}_T(t)$ as described in Section 2, and, using the estimate $\hat{\hat{G}}(t)$, the estimates $\hat{\Lambda}_{TK}(t)$ and $\hat{\Lambda}_{SM}(t)$ described in Sections 3 and 4, respectively. The estimates are shown in Fig.6 (a), where it is seen that $\hat{\Lambda}_{TK}(t)$ and $\hat{\Lambda}_{SM}(t)$ are almost identical but that $\hat{\Lambda}_T(t)$ is very different. We know from the simulations of Section 4 that when the event processes are non-Poisson (in particular, when overdispersion is present), the estimator $\hat{\Lambda}_T(t)$ tends to significantly overestimate $\Lambda(t)$. Diagnostics based on fitting a negative binomial model to the $(N_i(t), 0 < t \leq \tau_i)$'s indicate that there is substantial overdispersion present and consequently, we discard $\hat{\Lambda}_T(t)$ from further consideration. Fig.6 (b) shows pointwise approximate 95% confidance intervals for $\Lambda(t)$ given by $\hat{\Lambda}_{SM}(t) \pm 1.96\hat{V}_{SM}(t)^{\frac{1}{2}}$ and $\hat{\Lambda}_{TK}(t) \pm 1.96\hat{V}_{TK}(t)^{\frac{1}{2}}$, respectively. The variance estimate $\hat{V}_{SM}(t)$ is the robust estimate (22). For $\hat{\Lambda}_{TK}(t)$ we calculated the variance estimate $\hat{V}_{TK}(t)$ based on a piecewise constant intensity function, $\lambda(t) = \lambda_j$ for $t \in (a_{j-1}, a_j]$ where $a_0 = 0, a_1 = 30, \ldots, a_{K-1} = 330, a_K = 365$, as suggested at the end of Section 2. Appendix C outlines the calculations. The robust limits from $\hat{\Lambda}_{SM}(t)$ are slightly wider than the Poisson limits from $\hat{\Lambda}_{TK}(t)$.

Finally, we remark that the estimates of $\Lambda(t)$ given here depend on the assumption that the τ_i's are independent of the warranty claim processes. For the vehicle system under consideration, it is possible that claim rates $\lambda(t)$ are higher for the cars that accumulate mileage more rapidly. If this is so, then the τ_i's will not be strictly independent of the claim processes, because cars with high u_i's (mileage accumulation rates) give smaller τ_i's. Diagnostic checks, to be presented elsewhere, suggest a slight dependence of the claim rates on the u_i's. In this case it may be seen that $\hat{\Lambda}_{SM}(t)$ estimates the expected number of warranty claims observed, defined by $\sum_{s=1}^{t} E\{n_i(s)|\tau_i \geq s\}$ rather that $\sum_{s=1}^{t} E\{n_i(s)\}$. For this example it appears that the two functions do not differ greatly, so there is little harm in treating $\hat{\Lambda}_{SM}(t)$ and $\hat{\Lambda}_{TK}(t)$ as estimating $\Lambda(t)$.

7 Conclusion

Our results indicate that when recurrent event data are zero-truncated, care is needed. If the events follow a Poisson process the truncated data likelihood L_T of Section 2 is available, but simulations and large sample calculations show that the *m.l.e.* $\hat{\Lambda}_T(t)$ can be badly biased when events do not follow a Poisson process. In many applications there is information about the distribution of observation times (τ_i's) as well as the total number of the units, however, and in this case the paper shows (i) that the *m.l.e.* based on the Poisson likelihood L_{TK}

(9) is reasonably robust under heavy truncation, and (ii) that the simple moment estimator $\hat{\Lambda}_{SM}(t)$ in Section 4.1 is both robust against Poisson departures generally and quite efficient under the Poisson model.

Several further investigations would be worthwhile. One that is of theoretical and practical interest is the determination of asymptotic variances and variance estimates for $\hat{\Lambda}_T(t)$ and $\hat{\Lambda}_{TK}(t)$ under a continuous-time Poisson model. Methods of estimating $\Lambda(t)$ under non-Poisson models when there is no supplementary information about $G(\tau)$ would also be useful.

The methods in this paper assume that the observation times are determined independently of the event processes. This will sometimes be violated, for example if individuals that experience many events are "withdrawn" in some way; an example is in the case of equipment which fails so often that it is taken out of service. In the case of automobile warranties, as discussed in Section 6, the τ_i's might be related to claims if the rate of claims depends both on age and the usage rate (mileage accumulation) for the car. In such situations it is necessary to model the relationship between observation times and the event processes, if we are to estimate $\Lambda(t)$ or $\lambda(t)$ consistently. Such models, and diagnostic methods for assuming whether observation times and event processes are independent, will be presented elsewhere.

Finally, covariates may be introduced into the models used here. This would be worthy of study, because individual-level covariates are frequently of interest. We note that one way to assess the independence of observation times is to introduce a covariate based on them.

Appendix A: Asymptotics for $\hat{\Lambda}_{SM}(t)$

Theorem 1 *Suppose $N_i(t), i = 1, \ldots, M$ are M independent counting processes with $E\{N_i(t)\} = \Lambda(t)$. Observation times, τ_i's, are i.i.d.random variables with distribution $G(.)$, and τ_i is independent of $N_i(t)$. If $Cov(dN_i(t), dN_i(s)) = c(t,s)dtds$ exists when $s, t \in (0, \tau^*]$, $\tau^* = \sup\{s : \bar{G}(s) > 0\}$, then*

(A1) $\lim_{M\to\infty} \hat{\Lambda}_{SM,M}(t) = \Lambda(t)$, a.s. $\forall t \in (0, \tau^*]$;

(A2) $\sqrt{M}(\hat{\Lambda}_{SM,M}(t) - \Lambda(t)) \xrightarrow{d} \mathcal{G}_{SM}(t)$, as $M \to \infty$, where $(\mathcal{G}_{SM}(t), t \in (0, \tau^*])$ a Gaussian process with mean 0 and covariance process $Cov(\mathcal{G}_{SM}(t), \mathcal{G}_{SM}(s))$ which will be given below.

PROOF: Note that, $\forall t \in (0, \tau^*]$,

$$\hat{\Lambda}_{SM,M}(t) = \frac{1}{M}\sum_{i=1}^{M} X_i^{(SM)}(t)$$

with $X_i^{(SM)}(t) = \int_0^t \frac{\delta_i(s)}{\bar{G}(s)} dN_i(s)$, $i = 1, \ldots, M$, are *i.i.d.* random variables, where $\bar{G}(s) = \Pr\{\tau \geq s\}$. Moreover,

$$E\{X_i^{(SM)}(t)\} = \Lambda(t),$$

and

$$\begin{aligned} &\mathrm{Cov}(X_i^{(SM)}(t), X_i^{(SM)}(s)) \\ = &\int_0^t \int_0^s \frac{\bar{G}(u \vee v-)}{\bar{G}(u-)\bar{G}(v-)}[c(u,v) + \lambda(u)\lambda(v)]dudv - \int_0^t \int_0^s \lambda(u)\lambda(v)dudv, \end{aligned}$$

denoted by $\int_0^t \int_0^s c_{SM}(u,v)dudv$. According to the SLLN and CLT, we know

$$\hat{\Lambda}_{SM,M}(t) \stackrel{a.s.}{\to} \Lambda(t),$$

$\forall t \in (0, \tau^*]$. Also $\forall \alpha, \beta > 0$,

$$\begin{aligned} &\alpha\sqrt{M}(\hat{\Lambda}_{SM,M}(t) - \Lambda(t)) + \beta\sqrt{M}(\hat{\Lambda}_{SM,M}(s) - \Lambda(s)) \\ = &\frac{1}{\sqrt{M}}\sum_{i=1}^{M}[\alpha(X_i^{(SM)}(t) - \Lambda(t)) + \beta(X_i^{(SM)}(t) - \Lambda(t))] \stackrel{d}{\to} N(0, \sigma_{SM}^{\alpha,\beta}(t,s)), \end{aligned}$$

where $\sigma_{SM}^{\alpha,\beta}(t,s) = \alpha^2 \int_0^t \int_0^t c_{SM}(u,v)dudv + 2\alpha\beta \int_0^t \int_0^s c_{SM}(u,v)dudv + \beta^2 \int_0^s \int_0^s c_{SM}(u,v)dudv$.

Therefore, (A2) holds. □

Appendix B: Bias of $\hat{\Lambda}_T(t)$ and $\hat{\Lambda}_{TK}(t)$ under departures from the Poisson process

From (5) we see that $\hat{\lambda}_T(t)$ satisfies

$$\hat{\lambda}_T(t) = \frac{\frac{n.(t)}{M}}{\frac{1}{M}\sum_{i=1}^{m} \frac{\delta_i(t)}{1-\exp[-\hat{\Lambda}_T(\tau_i)]}}, \qquad \text{(B1)}$$

and from (12) that $\hat{\lambda}_{TK}(t)$ satisfies

$$\hat{\lambda}_{TK}(t) = \frac{\frac{n.(t)}{M}}{\frac{1}{M}\{\sum_{i=1}^{m} \delta_i(t) + (M-m)\frac{\hat{\Lambda}_{TK}(t)}{\hat{\Lambda}_{TK}(0)}\}}, \qquad \text{(B2)}$$

where $\hat{A}_{TK}(t) = \int_t^\infty \exp[-\Lambda_{TK}(s)]dG(s)$. If we define

$$\bar{\bar{G}}_T(t;\lambda(.)) = \frac{1}{M}\sum_{i=1}^{m}\frac{\delta_i(t)}{1-\exp[-\Lambda(\tau_i)]}, \text{ and}$$

$$\bar{\bar{G}}_{TK}(t;\lambda(.)) = \frac{1}{M}\{\sum_{i=1}^{m}\delta_i(t) + (M-m)\frac{A(t)}{A(0)}\},$$

then some insight into the behavior of $\hat{\lambda}_T(t)$ and $\hat{\lambda}_{TK}(t)$ can be gained by examining how well $\bar{\bar{G}}_T(t;\lambda(.))$ and $\bar{\bar{G}}_{TK}(t;\lambda(.))$ estimate $\bar{G}(t)$. Calculations in Hu (1994) show that under the mixed Poisson models we considered in Section 4.1 $\bar{\bar{G}}_T(t;\lambda(.))$ underestimates $\bar{G}(t)$ substantially, thus suggesting that (B1) will overestimate $\lambda(t)$; this is what was observed in Fig.3, portraying the simulation results of Section 4.1. Calculations also show that $\bar{\bar{G}}_{TK}(t;\lambda(.))$ estimates $\bar{G}(t)$ well under the mixed Poisson models when truncation is heavy ($\frac{m}{M}$ is small) and that $\hat{\lambda}_{TK}(t)$ estimates $\lambda(t)$ well. With substantial extra-Poisson variation and $\frac{m}{M}$ over 0.20 or so, however, $\hat{\lambda}_{TK}(t)$ tends to overestimate $\lambda(t)$, as is also indicated by the simulation results.

Appendix C: Poisson processes with piecewise constant intensities

Consider the piecewise constant intensity function

$$\lambda(t) = \lambda_j \quad \text{for } a_{j-1} < t \le a_j$$

where $a_0 = 0 < a_1 < \ldots < a_K \le \infty$ are specified. By considering such models with K fairly small (say in the 3-10 range) provides enough flexibility to model most practical situations, and allows the easy calculation of estimates and standard errors, based on either of the likelihood function L_T or L_{TK}. Using L_T given by (1), we obtain the maximum likelihood equations

$$\frac{\partial l_T}{\partial \lambda_j} = \frac{n_{.j}}{\lambda_j} - \sum_{i=1}^{m}\frac{w_{ij}}{1-\exp[-\Lambda(\tau_i)]}, \quad j = 1,\ldots,K$$

where $w_{ij} = \frac{\partial\Lambda(\tau_i)}{\partial\lambda_j} = I(\tau_i > a_{j-1})[(\tau_i \wedge a_j) - a_{j-1}]$. The observed information matrix has entries

$$-\frac{\partial^2 l_T}{\partial\lambda_j\partial\lambda_l} = I(j=l)\frac{n_{.j}}{\lambda_j^2} - \sum_{i=1}^{m}\frac{w_{ij}w_{il}e^{-\Lambda(\tau_i)}}{\{1-\exp[-\Lambda(\tau_i)]\}^2}.$$

The equations $\frac{\partial l_T}{\partial \lambda_j} = 0 (j = 1, \ldots, K)$ are readily solved to yield the estimates $\hat{\lambda}_j$'s, and the inverse of the observed information matrix evaluated at $\hat{\lambda}_1, \ldots, \hat{\lambda}_K$ provides asymptotic variance and covariance estimates for the $\hat{\lambda}_j$'s.

Using L_{TK} given by (9), we get

$$\frac{\partial l_{TK}}{\partial \lambda_j} = \frac{n_{.j}}{\lambda_j} - \sum_{i=1}^{m} w_{ij} - (M - m)\frac{A_j}{A_0}$$

where $A_0 = \int_0^\tau e^{-\Lambda(u)} dG(u)$ and $A_j = -\frac{\partial A_0}{\partial \lambda_j}, j = 1, \ldots, K$. The observed information matrix has entries

$$-\frac{\partial^2 l_{TK}}{\partial \lambda_j \partial \lambda_l} = I(j = l)\frac{n_{.j}}{\lambda_j^2} - (M - m)\frac{A_{jl}A_0 - A_j A_l}{A_0^2},$$

where $A_{jl} = -\frac{\partial A_j}{\partial \lambda_l} = \int_0^\tau I(u > a_{j-1})[(u \wedge a_j) - a_{j-1}]I(u > a_{l-1})[(u \wedge a_l) - a_{l-1}]e^{-\Lambda(u)} dG(u)$. As with L_T, estimates $\hat{\lambda}_1, \ldots, \hat{\lambda}_K$ and variance estimates are readily obtained.

Acknowledgements

This research was supported in part by grants to the second author from the Natural Sciences and Engineering Research Council of Canada and the Manufacturing Research Corporation of Ontario.

References

Andersen, P. and Borgan, O. (1985). Counting Process Models for Life History Data: A Review (with discussion). *Scand J. Statist* **12**, 97-158.

Hu, X. J. (1994) Analysis of Field Reliability Data (Ph.D. thesis, University of Waterloo).

Kalbfleisch, J. D. and Lawless, J. F. (1988). Estimation of Reliability from Field Performance Studies (with discussion). *Technometrics* **30**, 365-388.

______(1993). Statistical Analysis of Warranty Claims Data *IIQP Research Report, Univ. of Waterloo* RR-93-03.

Kalbfleisch, J. D., Lawless, J. F. and Robinson, J. A. (1991). Methods for the Analysis and Prediction of Warranty Claims. *Technometrics* **33**, 273-285.

Lawless, J. F. and Nadeau, J. C. (1993). Some Simple Robust Methods for the Analysis of Recurrent Events *IIQP Research Report, Univ. of Waterloo* RR-93-02.

Lawless, J. F. and Kalbfleisch, J. D. (1992). Some Issues in the Collection and Analysis of

Field Reliability Data. In *Survival Analysis: State of the Art* eds. J. P. Klein and P. K. Goel, 141-152. Dordrecht: Kluwer

Suzuki, K. (1985a). Estimation Method of Lifetime Based on the Record of Failures During the Warranty Period. *J. Amer. Statist Assoc.* **80**, 68-72.

______ (1985b). Nonparametric Estimation of Lifetime Distribution from Incomplete Field Data. *Technometrics* **27**, 263-271.

______ (1987). Analysis of Field Failure Data from a Nonhomogeneous Poisson Process. *Rep. Stat. Appl. Res.* **34**, 6-15.

Suzuki, K. and Epstein, B. (1992). Estimation of Lifetime Distribution Using the Relationship of Calendar Time and Usage Time. *Technical Report, University of Electro-Communications, Department of Communications and Systems Engineering , Japan* UEC-CAS-92-08.

Suzuki, K. and Kasashima, T. (1993). Estimation of Lifetime Distribution from Incomplete Field Data with Different Observational Periods. *Technical Report, University of Electro-Communications, Department of Communications and Systems Engineering , Japan* UEC-CAS-93-02.