Published in final edited form as:

J Stat Plan Inference. 2010 March 1; 140(3): 597–615. doi:10.1016/j.jspi.2009.08.007.

Estimation and Efficiency with Recurrent Event Data under Informative Monitoring

Akim Adekpedjou^{*}, Edsel A. Peña[†], and Jonathan Quiton[‡]

^{*}A. Adekpedjou (akima@umr.edu) is Assistant Professor, Department of Mathematics and Statistics, Missouri University of Science and Technology, Rolla, MO 65409. He acknowledges research support by NSF Grant DMS 0243594 (PI: J. Lynch) and NIH Grant GM 056182 (PI: E. Peña).

[†]E. Peña (pena@stat.sc.edu) is Professor, Department of Statistics, University of South Carolina, Columbia, SC 29208. He acknowledges research support by NIH Grant GM056182, NIH Grant 1 P20 RR17698, and NSF Grant DMS 0805809.

[‡]J. Quiton (Jonathan.Quiton@wku.edu) is Assistant Professor, Department of Mathematics and Computer Science, Western Kentucky University, Bowling Green, KY 42101-1078. He acknowledges research support by NIH Grant GM 056182 (PI: E. Peña).

Abstract

This article deals with studies that monitor occurrences of a recurrent event for n subjects or experimental units. It is assumed that the i^{th} unit is monitored over a random period $[0, \tau_i]$. The successive inter-event times $T_{i1}, T_{i2}, \dots$, are assumed independent of τ_i . The random number of event occurrences over the monitoring period is $K_i = \max\{k \in \{0, 1, 2, \dots\} : T_{i1} + T_{i2} + \dots + T_{ik} \leq \tau_i\}$. The T_{ij} s are assumed to be i.i.d. from an unknown distribution function F which belongs to a parametric family of distributions $\mathcal{C} = \{F(\cdot; \theta) : \theta \in \Theta \subset \mathbb{R}^p\}$. The τ_i s are assumed to be i.i.d. from unknown distribution function G . The problem of estimating θ , and consequently the distribution F , is considered under the assumption that the τ_i s are informative about the inter-event distribution. Specifically, $1 - G = (1 - F)^\beta$ for some unknown $\beta > 0$, a generalized Koziol-Green (cf., Koziol and Green (1976); Chen, Hollander, and Langberg (1982)) model. Asymptotic properties of estimators of θ , β , and F are presented. Efficiencies of estimators of θ and F are ascertained relative to estimators which ignores the informative monitoring aspect. These comparisons reveal the gain in efficiency when the informative structure of the model is exploited. Concrete demonstrations were performed for F exponential and a two-parameter Weibull.

Keywords

Counting processes; efficiency comparisons; exponential inter-event times; generalized Koziol-Green model; martingales; Weibull inter-event times

1 Introduction

The parametric, semiparametric, and nonparametric estimation of the distribution function of an event time has been the subject of intense research in past decades, especially in settings where there is at most one observed event (so-called *single-event* settings) during the monitoring period per experimental unit. Among the seminal papers dealing with this problem are those of Kaplan and Meier (1958), Efron (1967), Cox (1972), Breslow and Crowley (1974), Aalen (1978), and Borgan (1984); see the books of Fleming and Harrington (1991), Andersen, Borgan, Gill, and Keiding (1993), Kalbfleisch and Prentice (2002), and Aalen, Borgan, and Gjessing (2008). The situation where the event is recurrent so there could be more than one event occurrence per unit has also been dealt with, albeit not as thoroughly yet as the single-event case. In the recurrent event setting, the estimation problem has been considered by Gill (1980, 1981), Vardi (1982a,b), Wang and Chang (1999), and Peña, Strawderman, and Hollander (2001). Gill (1981) dealt with the problem of nonparametric inference for renewal processes in a life testing setting. Vardi (1982a) presented an algorithm for obtaining the maximum likelihood (ML) estimator of the survivor function when the underlying interoccurrence times are arithmetic. Sellke (1988), in the case of a single unit, considered the problem of establishing weak convergence of a Nelson-Aalen-type estimator when the length of the monitoring period increases without bound. Peña et al. (2001) proposed Nelson-Aalen and Kaplan-Meier-type estimators and derived their asymptotic properties when the number of units increases but with the monitoring time for each unit being finite with probability one, in contrast to the setting in Sellke (1988) where the monitoring time increases to infinity.

An important issue that arose in the single-event setting is the impact of an informative censoring mechanism. An analytically tractable informative random censorship model was proposed by Koziol and Green (1976) which assumes proportional hazards for the event time and the censoring time. This model was utilized by Chen et al. (1982) to study exact properties of the Kaplan-Meier estimator. Cheng and Lin (1987) also utilized this model to derive an estimator of the survivor function which exploits the informative censoring structure, and showed that their estimator is more efficient than the Kaplan and Meier (1958) estimator, especially under heavy censoring. Hollander and Peña (1989) also used this Koziol-Green model to obtain a more efficient class of confidence bands for the survivor function.

There are many situations however in the engineering, actuarial, biomedical, public health, social and economic sciences, as well as in business, where the event of interest is recurrent. Examples of such events are machine (mechanical or electronic) malfunction, non-life insurance claim, onset of depression, heart attack, economic recession, marital strife, and commission of a criminal act. In this recurrent event setting, the impact of an informative monitoring period has not been examined extensively. This article is for the purpose of studying the impact of an informative monitoring period especially in the context of efficiency gains and losses in the estimation of the inter-event time parameter and distribution. As pointed out in Peña et al. (2001) and Peña and Hollander (2004), recurrent event data have additional features that require attention in performing statistical inference. Two of these important features are: (i) because of the sum-quota data accrual scheme, the number of observed event occurrences is informative about the inter-event distribution even if G is unrelated to F ; and (ii) the variable that right-censors the last inter-event time at the end of the monitoring period is dependent on the previous inter-event times. Thus, there is both informative and dependent censoring in recurrent event data. Because of these additional features for recurrent event data, there is a need to study the additional impact of having a G informative about F in the estimation of F or its parameters, in particular, in the efficiency gain when the informative structure is exploited.

There has been several models that have been proposed to model informative censoring. William (1989) proposed a model where the censoring variable is related to the frailty of the individual. He showed in particular that in the case of exponential frailty the use of the Kaplan-Meier estimator can lead to errors in estimating the survivor probability. Wang, Qin, and Chiang (2001) proposed various models where the occurrence of recurrent events is modeled by a subject specific nonstationary Poisson process via a latent variable. Siannis (2004) considered a parametric model where the parameter represents the level of dependence between the failure and the censoring process. In this article we employ a generalization to the recurrent event setting of the model studied in Koziol and Green (1976), the so-called Koziol-Green (KG) model. This KG model has been most utilized in studying efficiency aspects under informative censoring in single-event settings; see for instance Chen et al. (1982) which obtains exact properties of the Kaplan-Meier estimator under this model, and Cheng and Lin (1987) which derives an estimator of the survivor function utilizing the informative structure. We point out that, just as in the case of the single-event setting, the utility of the proposed generalized KG model is not primarily to provide a practical and realistic model, but rather to provide a medium in which to examine analytically properties of inference procedures with recurrent event data.

The major goal of this article is to obtain estimators of the inter-event time distribution and its parameter for this generalized KG model and to ascertain the loss in efficiency if one ignores the informative structure. An outline of this article is as follows. Section 2 introduces relevant processes, describes the generalized KG model and its properties, and develops the estimators. The framework of stochastic processes is adopted to gain generality. Section 3 deals with asymptotic properties of the estimators under the KG model and those estimators derived by ignoring the KG assumption. Section 4 performs efficiency comparisons of the estimators that exploits the informative structure relative to those which were derived ignoring the structure. In particular, the efficiency of a fully nonparametric estimator of the inter-event distribution is examined. Section 5 presents the results of simulation studies which studies small- to moderate-sample properties of estimators for models in which closed-form analytical expressions are not possible, specifically when the inter-event distribution is a two-parameter Weibull. Finally, Section 6 provides some concluding remarks.

2 Model of Interest and Estimators

2.1 Random Entities

All random entities are defined on a basic probability space $(\Omega, \mathcal{F}, P)$. We suppose that there are n subjects in the study. For the i^{th} subject, $\{S_{ij}, j = 1, 2, \dots\}$ are the successive calendar times of event occurrences, while $\{T_{ij}, j = 1, 2, \dots\}$ are the successive inter-event times. Thus, we have $S_{i0} = 0$, $S_{ij} = \sum_{k=1}^j T_{ik}$ and $T_{ij} = S_{ij} - S_{i,j-1}$. The T_{ij} s are assumed to be i.i.d. nonnegative r.v.s with a common absolutely continuous distribution function F . In this paper we restrict to the i.i.d. inter-event times setting, while the possibly more relevant model for biostatistical applications with correlated inter-event times, specifically with the association induced by frailty components, will be dealt with in a separate paper.

The renewal function associated with F is

$$\rho_F(t) = \sum_{n=1}^{\infty} F^{*(n)}(t) I\{t \geq 0\}, \quad (1)$$

where $F^{*(n)}$ is the n^{th} convolution of F , the distribution of S_{in} . We assume that $F \in \mathcal{C} = \{F(\cdot; \theta) : \theta \in \Theta \subset \mathbb{R}^p\}$. The hazard rate function of $F(\cdot; \theta)$ is

$$\lambda_F(\cdot; \theta) = \frac{f(\cdot; \theta)}{1 - F(\cdot; \theta)} = \frac{f(\cdot; \theta)}{\bar{F}(\cdot; \theta)},$$

where $\bar{F} = 1 - F$ (analogously, $\bar{G} = 1 - G$). The cumulative hazard function is

$$\Lambda_F(\cdot; \theta) = \int_0^\cdot \lambda_F(v; \theta) dv$$

(similarly λ_G and Λ_G). For subject i , $i = 1, 2, \dots, n$, the recurrent event is observed over a random period $[0, \tau_i]$, where the τ_i s represent the end of the monitoring periods and are i.i.d. according to some distribution function G . These τ_i s may also be viewed as the right-censoring variables, though the right-censoring structure is somewhat different from the usual right-censoring structure with a single event per unit as explained in the next paragraph. Furthermore, the τ_i s and T_{ij} s are mutually independent. With $\mathbf{Z}_{+,0} = \{0, 1, 2, \dots\}$, the (random) number of event occurrences observed for the i th unit is $K_i = \max\{k \in \mathbf{Z}_{+,0} : S_{ik} \leq \tau_i\}$. Therefore, the random observables for the i th subject is represented by the vector

$$\mathbf{D}_i = (K_i, \tau_i, T_{i1}, \dots, T_{iK_i}, \tau_i - S_{iK_i}), i = 1, 2, \dots, n. \quad (2)$$

Let us examine in more detail the censoring structure for this model. Note that there will always be one right-censored inter-event time per unit. This inter-event time, denoted by T_{i,K_i+1} , is right-censored by the variable $\tau_i - S_{iK_i}$. Observe that the right-censoring variable depends on the previous inter-event times through S_{iK_i} , and in fact, since both T_{i,K_i+1} and S_{iK_i} depend on the random variable K_i , then they are dependent. More interestingly, observe that T_{i,K_i+1} has a distribution different from T_{i1} owing to the randomness of K_i . This is evident from the resulting sum-quota constraint given by

$$\sum_{j=1}^{K_i} T_{ij} \leq \tau_i < \sum_{j=1}^{K_i} T_{ij} + T_{i,K_i+1},$$

and, in fact, T_{i,K_i+1} is stochastically larger than T_{i1} , a length-biased phenomenon. The intuition is that the inter-event time that covered the monitoring time tends to be longer, similar to the well-known phenomenon that a passenger arriving on a bus stop has a greater chance of riding a bus that waits longer! In this recurrent event model we thus have the situation where the right-censoring structure is both dependent and informative. Observe that the sample space for $\mathbf{D}_i$ is the subset

$$\mathcal{D}_i \subset \cup_{k=0}^{\infty} (\{k\} \times \mathbb{R}_+ \times \mathbb{R}_+^k \times \mathbb{R}_+)$$

such that $\mathbf{d} = (k, \tau, t_1, t_2, \dots, t_k, c) \in \mathcal{D}_i$ satisfies $t_j \leq \tau$, ($j=1, 2, \dots, k$), $c = \tau - \sum_{j=1}^k t_j$.

2.2 Generalized KG Model for Recurrent Events

The generalized Koziol-Green model for this recurrent event setting postulates that there exists a $\beta > 0$ such that $\bar{G}(t; \theta, \beta) = \bar{F}(t; \theta)^\beta$. This condition is equivalent to $\Lambda_G(\cdot; \theta, \beta) = \beta \Lambda_F(\cdot; \theta)$. In the single-event setting, the parameter β is referred to as the censoring parameter since Pr

$(\tau_i < T_i) = \beta/(1 + \beta)$ and τ_i right-censors T_i . In contrast, for the recurrent event setting, the parameter β determines the length of the monitoring period relative to the inter-event times. For example, when $F(t; \theta) = 1 - \exp(-\theta t)$ for $t \geq 0$, then

$$\frac{E(\tau_i|\theta, \beta)}{E(T_{i1}|\theta)} = \frac{1/(\theta\beta)}{1/\theta} = \frac{1}{\beta}.$$

More precisely, for this exponential model, $1/\beta$ is the expected number of event occurrences during the monitoring period. This is so since, given τ_i , the expected number of event occurrences over $[0, \tau_i]$ is $\theta\tau_i$, and the result follows since $E(\tau_i) = 1/(\theta\beta)$. For this recurrent event setting, the case where $\beta \in (0, 1]$ is of more practical relevance since it leads to more observed recurrences, though $\beta > 1$ is also an admissible value. The property $\Pr(\tau_i < T_{i1}) = \beta/(1 + \beta)$ still holds true for this recurrent event setting, but it seems inappropriate to refer to β in this setting as the ‘censoring parameter’ since as pointed out earlier, the effective right-censoring variable for T_{i,K_i+1} is $\tau_i - S_{iK_i}$, whose distribution not only depends on β but also on F . Thus, it is perhaps more proper to refer to β as the ‘monitoring parameter!’

A crucial independence property that was exploited in Chen et al. (1982) and the other papers dealing with this model in the single-event setting is the fact that $\min(\tau_i, T_i)$ and $I\{T_i \leq \tau_i\}$ are independent if and only if the KG model holds (cf., Allen (1963)). This allowed the derivation of the exact means and variances of the Kaplan-Meier estimator in Chen et al. (1982), and also facilitated the derivation of the semiparametric estimator of the survivor function in Cheng and Lin (1987). For the recurrent event setting in this paper, such an independence property does not come directly into play. For single-event settings, it has been pointed out, e.g., Csörgő and Faraway (1998), that the KG model may not occur much in practice, though its utility is more from the theoretical point of view as it provides a model allowing for the analytical examinations of properties of procedures. Our proposing the generalized KG model for recurrent event data is also not meant primarily as a practical model, but rather it is meant to provide a model specimen for the examination of analytical properties of procedures utilizing recurrent event data.

2.3 Estimators

The first goal of this article is, given the data in (2), to develop estimators of θ and β , and consequently, an estimator of F . For generality we adopt the approach of counting processes and martingales (cf., Fleming and Harrington (1991) and Andersen et al. (1993)).

For $i = 1, 2, \dots, n$ and $s \in \mathbb{R}_+$, let

$$N_i^\dagger(s) = \sum_{j=1}^{\infty} I\{S_{ij} \leq s \wedge \tau_i\}, \quad Y_i^\dagger(s) = I\{\tau_i \geq s\}, \quad \text{and} \quad N_i^r(s) = I\{\tau_i \leq s\}. \quad (3)$$

The processes in (3) counts the number of failures for unit i at time s ; indicates whether unit i is still under observation at time s ; and indicates whether unit i is past its monitoring time at time s , respectively. We augment the probability space by the filtration $\mathbf{F} = \{\mathcal{F}_s; s \geq 0\}$, where $\mathcal{F}_s$ is given by

$$\mathcal{F}_s = \mathcal{F}_0 \vee \left\{ \bigvee_{i=1}^n \sigma\{N_i^\dagger(v), Y_i^\dagger(v), 0 \leq v \leq s\} \right\},$$

with $\mathcal{F}_0$ representing the σ -field containing all the information available at time 0, and $\mathcal{F}_s$ is the σ -field containing all information generated on all subjects up to time s . Also, for each $i = 1, 2, \dots, n$, define the backward recurrence process via

$$R_i(s) = s - S_{iN_i^\dagger(s-)}, \quad (4)$$

which is the time elapsed at time s since the last event occurrence. This is an $\mathbf{F}$ -adapted and a left-continuous process, hence is an $\mathbf{F}$ -predictable process. For $s \in \mathfrak{R}_+$, let

$$A_i^\dagger(s; \theta) = \int_0^s Y_i^\dagger(v) \lambda_F(R_i(v); \theta) dv \quad \text{and} \quad A_i^\tau(s; \theta, \beta) = \int_0^s Y_i^\dagger(v) \beta \lambda_F(v; \theta) dv. \quad (5)$$

From stochastic integration theory, $A_i^\dagger(s; \theta)$ and $A_i^\tau(s; \theta, \beta)$ are, respectively, the compensators of $N_i^\dagger(s)$ and $N_i^\tau(s)$, so

$$M_i^\dagger(s; \theta) = N_i^\dagger(s) - A_i^\dagger(s; \theta) \quad \text{and} \quad M_i^\tau(s; \theta, \beta) = N_i^\tau(s) - A_i^\tau(s; \theta, \beta) \quad (6)$$

are, for each i , square-integrable $\mathbf{F}$ -martingales. From the results of Jacod (1974/75), the full likelihood process is, for $s \in \mathfrak{R}_+$,

$$L(s; \theta, \beta) = \prod_{i=1}^n \prod_{v=0}^s \left\{ Y_i^\dagger(v)^{\Delta N_i^\dagger(v) + \Delta N_i^\tau(v)} \lambda_F(R_i(v); \theta)^{\Delta N_i^\dagger(v)} \{\beta \lambda_F(v; \theta)\}^{\Delta N_i^\tau(v)} \right\} \\ \times \exp \left\{ - \sum_{i=1}^n \int_0^s Y_i^\dagger(v) [\lambda_F(R_i(v); \theta) + \beta \lambda_F(v; \theta)] dv \right\}. \quad (7)$$

Taking logarithm we obtain the log-likelihood process

$$l(s; \theta, \beta) = \sum_{i=1}^n \left\{ \int_0^s \log(Y_i^\dagger(v) \beta \lambda_F(v; \theta)) dN_i^\tau(v) + \int_0^s \log(Y_i^\dagger(v) \lambda_F(R_i(v); \theta)) dN_i^\dagger(v) \right\} \\ - \sum_{i=1}^n \left\{ \int_0^s Y_i^\dagger(v) \beta \lambda_F(v; \theta) dv + \int_0^s Y_i^\dagger(v) \lambda_F(R_i(v); \theta) dv \right\}. \quad (8)$$

For a vector $\mathbf{a}$, let $\mathbf{a}^0 = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^t$, and define the operator $\nabla_\theta = \partial/\partial\theta \equiv (\partial/\partial\theta_j, j = 1, 2, \dots, p)^t$. Let

$$\varphi(v; \theta) = \nabla_\theta \log \lambda_F(v; \theta). \quad (9)$$

Under the regularity condition that the order of differentiation with respect to θ and β and integration with respect to Lebesgue measure can be interchanged, we obtain the score processes for θ and β by taking the derivatives of (8) with respect to θ and β , respectively, to be

$$U_\theta(s; \theta, \beta) = \sum_{i=1}^n \left\{ \int_0^s \varphi(v; \theta) M_i^\tau(dv; \theta, \beta) + \int_0^s \varphi(R_i(v); \theta) M_i^\dagger(dv, \theta) \right\}; \quad (10)$$

$$U_{\beta}(s; \theta, \beta) = \sum_{i=1}^n \int_0^s \frac{1}{\beta} M_i^{\tau}(dv; \theta, \beta). \quad (11)$$

To make notation compact, let

$$\mathbf{H}_i(v; \theta, \beta) = \begin{pmatrix} \varphi(v; \theta) & \varphi(R_i(v); \theta) \\ \frac{1}{\beta} & 0 \end{pmatrix};$$

$$\mathbf{M}_i(v; \theta, \beta) = (M_i^{\tau}(v; \theta, \beta), M_i^{\dagger}(v; \theta))^t \quad \text{and} \quad \mathbf{U}(v; \theta, \beta) = (U_{\theta}^t(v; \theta, \beta), U_{\beta}(v; \theta, \beta))^t. \quad (12)$$

Then, the vector of score processes becomes

$$\mathbf{U}(s; \theta, \beta) = \sum_{i=1}^n \int_0^s \mathbf{H}_i(v; \theta, \beta) d\mathbf{M}_i(v; \theta, \beta), \quad s \in \mathbb{R}_+. \quad (13)$$

Let $s^* \in \mathbb{R}_+$. Equating the score vector in (13), evaluated at $s = s^*$, to $\mathbf{0}$ and solving for θ and β , we obtain the maximum likelihood (ML) estimators $\widehat{\theta} = \widehat{\theta}(s^*)$ and $\widehat{\beta} = \widehat{\beta}(s^*)$ of θ and β , respectively. In general, it is not possible to obtain closed-form expressions for these estimators, so in practice numerical methods such as the Newton-Raphson (NR) algorithm or the Nelder-Mead simplex algorithm could be used to obtain the ML estimates.

The observed Fisher information process, which is the negative of the second partial derivatives with respect to the parameters is

$$\mathbf{I}(s; \theta, \beta) = \begin{bmatrix} \mathbf{I}_{11}(s; \theta, \beta) & \mathbf{I}_{12}(s; \theta, \beta) \\ \mathbf{I}_{21}(s; \theta, \beta) & \mathbf{I}_{22}(s; \theta, \beta) \end{bmatrix} = - \begin{bmatrix} \frac{\partial}{\partial \theta \partial \theta^t} & \frac{\partial}{\partial \theta \partial \beta} \\ \frac{\partial}{\partial \beta \partial \theta^t} & \frac{\partial}{\partial \beta^2} \end{bmatrix} l(s; \theta, \beta), \quad (14)$$

where, with

$$\begin{aligned} \dot{\lambda}_F(\cdot; \theta) &= \frac{\partial}{\partial \theta} \lambda_F(\cdot; \theta) \quad \text{and} \quad \ddot{\lambda}_F(\cdot; \theta) = \frac{\partial^2}{\partial \theta \partial \theta^t} \lambda_F(\cdot; \theta); \\ \dot{\varphi}_F(\cdot; \theta) &= \frac{\partial}{\partial \theta} \varphi_F(\cdot; \theta) = \left(\ddot{\lambda}_F(\cdot; \theta) / \lambda_F(\cdot; \theta, \beta) \right) - \varphi_F(\cdot; \theta)^{\otimes 2}, \end{aligned}$$

we have

$$\begin{aligned} \mathbf{I}_{11}(s; \theta, \beta) &= \sum_{i=1}^n \int_0^s Y_i^{\dagger}(v) \left(\beta \ddot{\lambda}_F(v; \theta) + \ddot{\lambda}_F(R_i(v); \theta) \right) dv - \\ &\quad \sum_{i=1}^n \int_0^s \left(\dot{\varphi}_F(v; \theta) dN_i^{\tau}(v) + \dot{\varphi}_F(R_i(v); \theta) dN_i^{\dagger}(v) \right); \\ \mathbf{I}_{12}(s; \theta, \beta) &= \mathbf{I}_{21}(s; \theta, \beta)^t = \sum_{i=1}^n \int_0^s Y_i^{\dagger}(v) \dot{\lambda}_F(v; \theta) dv; \\ \mathbf{I}_{22}(s; \theta, \beta) &= \frac{1}{\beta^2} \sum_{i=1}^n N_i^{\tau}(s). \end{aligned}$$

Provided that the matrix inverse of the observed Fisher information matrix exists, the Newton-Raphson iteration below for numerically computing the parameter estimates may be implemented.

$$(\widehat{\theta}^{new}, \widehat{\beta}^{new})^t \leftarrow (\widehat{\theta}^{old}, \widehat{\beta}^{old})^t + \mathbf{I}(s^*; \widehat{\theta}^{old}, \widehat{\beta}^{old})^{-1} \mathbf{U}(s^*; \widehat{\theta}^{old}, \widehat{\beta}^{old}).$$

A more convenient computational implementation is achieved when we let $s^* \rightarrow \infty$ since the score equation for the ML estimator of β satisfies

$$\widehat{\beta}(\infty) = \widehat{\beta} = \frac{n}{\sum_{i=1}^n \Lambda_F(\tau_i; \widehat{\theta})}. \quad (15)$$

As a consequence, the estimator of θ , denoted by $\widehat{\theta}(\infty) = \widehat{\theta}$, satisfies

$$\sum_{i=1}^n \left\{ \varphi_F(\tau_i; \widehat{\theta}) + \sum_{j=1}^{K_i} \varphi_F(T_{ij}; \widehat{\theta}) \right\} = n \frac{\sum_{i=1}^n \dot{\Lambda}_F(\tau_i; \widehat{\theta})}{\sum_{i=1}^n \Lambda_F(\tau_i; \widehat{\theta})} + \sum_{i=1}^n \left\{ \sum_{j=1}^{K_i} \dot{\Lambda}_F(T_{ij}; \widehat{\theta}) + \dot{\Lambda}_F(\tau_i - S_{iK_i}; \widehat{\theta}) \right\}, \quad (16)$$

since $\int_0^t \varphi_F(w; \theta) \lambda_F(w; \theta) dw = \dot{\Lambda}_F(t; \theta)$. Equation (16) may be solved through numerical methods, e.g., Newton-Raphson iteration.

Having obtained the estimate $\widehat{\theta}(s^*)$ of θ , then the parametric estimator of F is provided by

$$\widehat{F}(\cdot) = \bar{F}(\cdot; \widehat{\theta}(s^*)). \quad (17)$$

We will obtain the asymptotic distribution of this estimator in Section 3 and compare this with a fully nonparametric estimator in Section 4.

2.4 Computational Forms

To facilitate the numerical implementation of the procedures to obtain estimates, we observe the forms below for the integrals with respect to the martingale processes. Let $\psi: \mathfrak{R}_+ \rightarrow \mathfrak{R}$ which could possibly depend also on (θ, β) , and

$$\Psi(s; \theta, \beta) = \int_0^s \psi(v; \theta, \beta) \lambda_F(v; \theta) dv.$$

Then, we have the following identities.

$$\begin{aligned} \sum_{i=1}^n \int_0^\infty \psi(v; \theta, \beta) M_i^\tau(dv; \theta, \beta) &= \sum_{i=1}^n [\psi(\tau_i; \theta, \beta) - \beta \Psi(\tau_i; \theta, \beta)]; \\ \sum_{i=1}^n \int_0^\infty \psi(R_i(v); \theta, \beta) M_i^\dagger(dv; \theta, \beta) &= \sum_{i=1}^n \left[\sum_{j=1}^{K_i} \psi(T_{ij}; \theta, \beta) - \left\{ \sum_{j=1}^{K_i} \Psi(T_{ij}; \theta, \beta) + \Psi(\tau_i - S_{iK_i}; \theta, \beta) \right\} \right]. \end{aligned}$$

These expressions are used in re-expressing the score functions as $s \rightarrow \infty$. With regards to the terms in the observed Fisher information, we also observe the following identities:

$$\dot{\lambda}_F = \varphi_F \lambda_F \quad \text{and} \quad \ddot{\lambda}_F = \left(\dot{\varphi}_F + \varphi_F^{\otimes 2} \right) \lambda_F$$

with $\dot{\varphi}_F = \left(\partial / \partial \theta' \right) \varphi_F = \left(\partial^2 / \partial \theta \partial \theta' \right) \log \lambda_F$.

2.4.1 Exponential Inter-Event Times—Implementing to the case of exponentially-distributed inter-event times with $\lambda_F(t; \theta) = \theta$ so $\phi_F(t; \theta) = 1/\theta$, we find that

$$U_\theta(\infty; \theta, \beta) = \frac{n + \sum_{i=1}^n K_i}{\theta} - (1 + \beta) \sum_{i=1}^n \tau_i \quad \text{and} \quad U_\beta(\infty; \theta, \beta) = \frac{n}{\beta} - \theta \sum_{i=1}^n \tau_i. \quad (18)$$

Equating both to zeros and solving for θ and β yield the estimators

$$\hat{\theta} = \frac{\sum_{i=1}^n K_i}{\sum_{i=1}^n \tau_i} \quad \text{and} \quad \hat{\beta} = \frac{n}{\sum_{i=1}^n K_i}, \quad (19)$$

so the ML estimator of θ is just the occurrence-exposure rate.

For this exponential case, with $\bar{K} = \sum_{i=1}^n K_i / n$ and $\bar{\tau} = \sum_{i=1}^n \tau_i / n$, we also obtain the observed Fisher information matrix to be

$$\mathbf{I}(\infty; \theta, \beta) = n \begin{bmatrix} \frac{1 + \bar{K}}{\theta^2} & \frac{\bar{\tau}}{\theta} \\ \frac{\bar{\tau}}{\theta} & \frac{1}{\beta^2} \end{bmatrix}. \quad (20)$$

Dividing by n and taking the limit as $n \rightarrow \infty$, noting that $\bar{K} \xrightarrow{p} E(N_1^+(\tau_1)) = 1/\beta$ and $\bar{\tau} \xrightarrow{p} E(\tau_1) = 1/(\theta\beta)$, then the in-probability limiting matrix is

$$\Sigma(\infty; \theta, \beta) = \begin{bmatrix} \frac{1 + \beta}{\theta^2 \beta} & \frac{1}{\theta \beta} \\ \frac{1}{\theta \beta} & \frac{1}{\beta^2} \end{bmatrix}.$$

The inverse of this matrix is

$$\Sigma(\infty; \theta, \beta)^{-1} = \begin{bmatrix} \theta^2 \beta & -\theta \beta^2 \\ -\theta \beta^2 & \beta^2 (1 + \beta) \end{bmatrix}. \quad (21)$$

2.4.2 Weibull Inter-Event Times—In this subsubsection we present the estimation procedure when F is Weibull with shape parameter θ_1 and scale parameter θ_2 , so that $\theta = (\theta_1, \theta_2)' \in \mathfrak{R}_+^2$. For this case we have

$$\Lambda_F(t; \theta) = (\theta_2 t)^{\theta_1} \quad \text{and} \quad \lambda_F(t; \theta) = (\theta_1 \theta_2) (\theta_2 t)^{\theta_1 - 1}. \quad (22)$$

We observe that given $\theta = (\theta_1, \theta_2)$ and β in this Weibull situation, the ratio of $E(\tau)$ and $E(T_1)$ equals $1/(\beta^{1/\theta_1})$. Thus, if one desires approximately k_0 observed recurrences per unit, then the monitoring parameter could be chosen approximately equal to $1/(k_0^{\theta_1})$.

From (22), it follows that

$$\varphi_F(t; \theta) = \left[\frac{\frac{1}{\theta_1} [1 + \log \Lambda_F(t; \theta)]}{\frac{\theta_1}{\theta_2}} \right] \quad \text{and} \quad \dot{\Lambda}_F(t; \theta) = \left[\frac{\frac{1}{\theta_1} \log \Lambda_F(t; \theta)}{\frac{\theta_1}{\theta_2}} \right] \Lambda_F(t; \theta). \quad (23)$$

The second equation arising in (16) simplifies to

$$\sum_{i=1}^n K_i = \sum_{i=1}^n \sum_{j=1}^{K_i+1} \Lambda_F(T_{ij}; \hat{\theta})$$

where, for brevity of notation but with a slight conflict with our earlier usage of T_{i, K_i+1} , from hereon we shall let

$$T_{iK_i+1} \equiv \tau_i - S_{iK_i}.$$

Consequently, $\hat{\theta}_2$ in terms of $\hat{\theta}_1$ is given by

$$\hat{\theta}_2(\hat{\theta}_1) = \left\{ \frac{\sum_{i=1}^n K_i}{\sum_{i=1}^n \sum_{j=1}^{K_i+1} T_{ij}^{\hat{\theta}_1}} \right\}^{1/\hat{\theta}_1} = \left\{ \frac{K_{\bullet}}{W(\hat{\theta}_1)} \right\}^{1/\hat{\theta}_1}, \quad (24)$$

where

$$K_{\bullet} = \sum_{i=1}^n K_i \quad \text{and} \quad W(\theta_1) = \sum_{i=1}^n \sum_{j=1}^{K_i+1} T_{ij}^{\theta_1}.$$

Let us then define, for $i = 1, 2, \dots, n$,

$$\tau_i^*(\theta_1) = K_{\bullet} \frac{\tau_i^{\theta_1}}{W(\theta_1)} \quad \text{and} \quad T_{ij}^*(\theta_1) = K_{\bullet} \frac{T_{ij}^{\theta_1}}{W(\theta_1)}, \quad j=1, 2, \dots, K_i+1.$$

Substituting the expression in (24) in the first equation arising in (16), we obtain that $\hat{\theta}_1$ is the solution in θ_1 of the equation

$$(n+K_{\bullet}) + \sum_{i=1}^n \left\{ \log \tau_i^*(\theta_1) + \sum_{j=1}^{K_i} \log T_{ij}^*(\theta_1) \right\} =$$

$$n \frac{\sum_{i=1}^n \tau_i^*(\theta_1) \log \tau_i^*(\theta_1)}{\sum_{i=1}^n \tau_i^*(\theta_1)} + \sum_{i=1}^n \sum_{j=1}^{K_i+1} T_{ij}^*(\theta_1) \log T_{ij}^*(\theta_1). \quad (25)$$

The solution to this uni-dimensional equation maybe obtained numerically by plotting, direct search, or a Newton-Raphson iteration. The last approach will be implemented in performing simulations associated with this Weibull inter-event times. For the purpose of implementing this procedure, say in the R package, define the function $w: \mathfrak{R}_+^2 \rightarrow \mathfrak{R}$ via

$$w(t, \theta) = t^\theta \quad \text{so that} \quad \dot{w}(t, \theta) \equiv \frac{\partial}{\partial \theta} w(t, \theta) = w(t, \theta) \log t.$$

In terms of the w -function, we also have

$$W(\theta) = \sum_{i=1}^n \sum_{j=1}^{K_i+1} w(T_{ij}, \theta), \quad \tau_i^*(\theta) = K_{\bullet} \frac{w(\tau_i, \theta)}{W(\theta)} \quad \text{and} \quad T_{ij}^*(\theta) = K_{\bullet} \frac{w(T_{ij}, \theta)}{W(\theta)}.$$

Furthermore, define $r: \mathfrak{R}_+^2 \rightarrow \mathfrak{R}$ via

$$r(t, \theta) = \frac{w(t, \theta)}{W(\theta)} \left[\log t - \frac{\dot{W}(\theta)}{W(\theta)} \right],$$

where

$$\dot{W}(\theta) \equiv \frac{dW(\theta)}{d\theta} = \sum_{i=1}^n \sum_{j=1}^{K_i+1} \dot{w}(T_{ij}, \theta).$$

Then, in terms of this r -function,

$$\dot{\tau}_i^*(\theta) \equiv \frac{\partial}{\partial \theta} \tau_i^*(\theta) = K_{\bullet} r(\tau_i, \theta) \quad \text{and} \quad \dot{T}_{ij}^*(\theta) \equiv \frac{\partial}{\partial \theta} T_{ij}^*(\theta) = K_{\bullet} r(T_{ij}, \theta).$$

Define the function $q: \mathfrak{R}_+ \rightarrow \mathfrak{R}$ according to

$$q(\theta) = n + K_{\bullet} + \sum_{i=1}^n \log \tau_i^*(\theta) + \sum_{i=1}^n \sum_{j=1}^{K_i} \log T_{ij}^*(\theta) -$$

$$n \frac{\sum_{i=1}^n \tau_i^*(\theta) \log \tau_i^*(\theta)}{\sum_{i=1}^n \tau_i^*(\theta)} - \sum_{i=1}^n \sum_{j=1}^{K_i+1} T_{ij}^*(\theta) \log T_{ij}^*(\theta), \quad (26)$$

so that $\widehat{\theta}_1$ solves $q(\widehat{\theta}_1) = 0$ from (25). The derivative of this q -function is

$$\begin{aligned} \dot{q}(\theta) \equiv \frac{dq(\theta)}{d\theta} = & \sum_{i=1}^n \frac{\dot{\tau}_i^*(\theta)}{\tau_i^*(\theta)} + \sum_{i=1}^n \sum_{j=1}^{K_i} \frac{\dot{T}_{ij}^*(\theta)}{T_{ij}^*(\theta)} - n \frac{\sum_{i=1}^n \dot{\tau}_i^*(\theta) [1 + \log \tau_i^*(\theta)]}{\sum_{i=1}^n \tau_i^*(\theta)} + \\ & n \left[\frac{\sum_{i=1}^n \tau_i^*(\theta) \log \tau_i^*(\theta)}{\sum_{i=1}^n \tau_i^*(\theta)} \right] \left[\frac{\sum_{i=1}^n \dot{\tau}_i^*(\theta)}{\sum_{i=1}^n \tau_i^*(\theta)} \right] - \sum_{i=1}^n \sum_{j=1}^{K_i+1} \dot{T}_{ij}^*(\theta) [1 + \log T_{ij}^*(\theta)]. \end{aligned} \quad (27)$$

Using these function definitions, the Newton-Raphson iteration for obtaining the estimate of the Weibull shape parameter θ_1 is given by

$$\widehat{\theta}_1^{new} \leftarrow \widehat{\theta}_1^{old} - \frac{q(\widehat{\theta}_1^{old})}{\dot{q}(\widehat{\theta}_1^{old})}. \quad (28)$$

Upon obtaining the estimate of θ_1 , the estimates of the scale parameter θ_2 and the monitoring parameter β could then be obtained using (24) for $\widehat{\theta}_2$ and, from (15),

$$\widehat{\beta} = \frac{n}{\sum_{i=1}^n (\widehat{\theta}_2 \tau_i)^{\widehat{\theta}_1}}. \quad (29)$$

3 Asymptotic Properties

In this section we study the asymptotic properties of the ML estimators. We will make use of the results in Borgan (1984), which deals with the consistency and asymptotic normality of ML estimators in parametric counting process models. We will consider the case where the number of subjects is increasing to infinity ($n \rightarrow \infty$) in contrast to the situation where only one subject is considered and the time of monitoring increases to infinity ($\tau \rightarrow \infty$) as in Sellke (1988). Some of the regularity conditions in Borgan (1984) will be reformulated in terms of gap-times, which will enable obtaining more useful analytical conditions.

3.1 Reformulated Processes

Following an idea exploited in Sellke (1988) and also in Peña et al. (2001), we define stochastic processes $Z_i: \mathcal{R}_+^2 \rightarrow \{0, 1\}$ via

$$Z_i(s, t) = I\{R_i(s) \leq t\} \quad \text{for } i=1, 2, \dots, n. \quad (30)$$

The first time parameter corresponds to calendar time, while the second time parameter represents gap or inter-event time. Note that $Z_i(s, t)$ indicates whether at calendar time s at most t units of time have elapsed since the last event occurrence. Because it is $\mathbf{F}$ -adapted and has left-continuous paths, then $Z_i(\cdot, t)$ is an $\mathbf{F}$ -predictable process, aside from bounded.

To facilitate our asymptotic analysis we introduce a generalized likelihood process involving two time indices $L_G: \mathcal{R}_+^2 \times \Theta \times \mathcal{R}_+ \rightarrow \mathcal{R}_+$ defined via

$$L_G(s, t; \theta, \beta) = \prod_{i=1}^n \left\{ \prod_{v=0}^s \lambda_G(v; \theta, \beta)^{\Delta N_i^+(v)} \right\} \exp \left\{ - \int_0^s Y_i^+(v) \lambda_G(v; \theta, \beta) dv \right\} \times \left\{ \prod_{v=0}^s \lambda_F(R_i(v); \theta)^{Z_i(v, t) \Delta N_i^+(v)} \right\} \exp \left\{ - \int_0^s Y_i^+(v) Z_i(v, t) \lambda_F(R_i(v); \theta) dv \right\}. \quad (31)$$

Notice here that the likelihood in Sellke (1988) could not be used directly for asymptotic properties purposes because it does not contain an informative censoring part since the distribution of the censoring time does not depend on that of the inter-event time. So the most important difference between the two likelihood is the contribution of the censored observations. Observe that $L(s; \theta, \beta) = \lim_{t \rightarrow \infty} L_G(s, t; \theta, \beta)$ since $\lim_{t \rightarrow \infty} Z_i(s, t) = 1$. Therefore, functions and estimators derived from $L(s; \theta, \beta)$ are limits of analogous functions and estimators obtained from $L_G(s, t; \theta, \beta)$ as $t \rightarrow \infty$. As we will see, however, dealing with L_G is more convenient analytically.

Taking the logarithm of L_G to obtain the generalized log-likelihood $l_G = \log L_G$, and then the partial derivatives of l_G with respect to θ and β , we obtain the generalized score process

$$\mathbf{U}_G(s, t; \theta, \beta) = \left[\begin{array}{c} \partial / \partial \theta \\ \partial / \partial \beta \end{array} \right] l_G(s, t; \theta, \beta) = \sum_{i=1}^n \int_0^s \mathbf{H}_i(v; \theta, \beta) \mathbf{M}_i(dv, t; \theta, \beta), \quad (32)$$

where $\mathbf{H}_i$ is as defined in (12), and $\mathbf{M}_i(s, t; \theta, \beta) = (M_i^+(s; \theta, \beta), M_i(s, t; \theta))^t$ with

$$M_i(s, t; \theta) = \int_0^s Z_i(v, t) M_i^+(dv; \theta).$$

If (θ, β) are the true parameter values, then for fixed $t \in \mathfrak{R}_+$, $\mathbf{M}_i(\cdot, t; \theta, \beta)$ is a vector of square-integrable martingales with $\mathbf{F}$ -predictable quadratic variation (PQV) process given by

$$\langle \mathbf{M}_i(\cdot, t; \theta, \beta) \rangle(s) = \text{Dg} \left(\int_0^s Y_i^+(v) \beta \lambda_F(v; \theta) dv, \int_0^s Y_i^+(v) Z_i(v, t) \lambda_F(R_i(v); \theta) dv \right), \quad (33)$$

where $\text{Dg}(\mathbf{a})$ is the diagonal matrix with diagonal elements being the elements of the vector $\mathbf{a}$. Re-scaling by $1/n$ and simplifying, we have

$$\Sigma_n(s, t; \theta, \beta) \equiv \frac{1}{n} \langle \mathbf{U}_G(\cdot, t; \theta, \beta) \rangle(s) = \left[\begin{array}{cc} \Sigma_{11n}(s, t; \theta, \beta) & \Sigma_{12n}(s; \theta, \beta) \\ \Sigma_{21n}(s; \theta, \beta) & \Sigma_{22n}(s; \theta, \beta) \end{array} \right]$$

where

$$\begin{aligned} \Sigma_{11n}(s, t; \theta, \beta) &= \Sigma_{11An}(s; \theta, \beta) + \Sigma_{11Bn}(s, t; \theta, \beta); \\ \Sigma_{11An}(s; \theta, \beta) &= \frac{1}{n} \sum_{i=1}^n \int_0^s \varphi_F(v; \theta)^{\otimes 2} Y_i^+(v) \beta \lambda_F(v; \theta) dv; \\ \Sigma_{11Bn}(s, t; \theta, \beta) &= \frac{1}{n} \sum_{i=1}^n \int_0^s \varphi_F(R_i(v); \theta)^{\otimes 2} Y_i^+(v) Z_i(v, t) \lambda_F(R_i(v); \theta) dv; \\ \Sigma_{12n}(s; \theta, \beta) &= \Sigma_{21n}(s; \theta, \beta)^t = \frac{1}{n} \sum_{i=1}^n \int_0^s \frac{1}{\beta} \varphi_F(v; \theta) Y_i^+(v) \beta \lambda_F(v; \theta) dv; \\ \Sigma_{22n}(s; \theta, \beta) &= \frac{1}{n} \sum_{i=1}^n \int_0^s \frac{1}{\beta^2} Y_i^+(v) \beta \lambda_F(v; \theta) dv. \end{aligned}$$

The matrix $\Sigma_{11Bn}(s, t; \theta, \beta)$ has the alternative representation given by

$$\Sigma_{11Bn}(s, t; \theta, \beta) = \frac{1}{n} \sum_{i=1}^n \int_0^t \varphi_F(w; \theta)^{\otimes 2} Y_i(s, w) \lambda_F(w; \theta) dw,$$

where

$$Y_i(s, t) = \sum_{j=1}^{N_i^\dagger(s-)} I\{T_{ij} \geq t\} + I\left\{(s \wedge \tau_i) - S_{iN_i^\dagger(s-)} \geq t\right\} \quad (34)$$

is the generalized at-risk process (cf., Peña et al. (2001)). It is of interest to obtain the limit of $\Sigma_n(s, t; \theta, \beta)$ as $n \rightarrow \infty$ to be able to use Borgan's results. For this purpose, we need the following lemma.

Lemma 1 For $(s^*, t^*) \in (0, \infty)^2$, then, as $n \rightarrow \infty$,

- i. $\sup_{v \in [0, s^*]} \left| \frac{1}{n} \sum_{i=1}^n Y_i^\dagger(v) - \bar{G}(v) \right| \xrightarrow{p} 0;$
- ii. $\sup_{(v, w) \in [0, s^*] \times [0, t^*]} \left| \frac{1}{n} \sum_{i=1}^n Y_i(v, w) - y(v, w; \theta, \beta) \right| \xrightarrow{p} 0,$

where the function $y(s, t; \theta, \beta)$ is given by (cf., Peña et al. (2001))

$$y(s, t; \theta, \beta) = \bar{F}(t; \theta) \bar{G}(t; \theta, \beta) \left\{ 1 + \frac{1}{\bar{G}(t; \theta, \beta)} \int_t^s \rho_F(w - t; \theta) G(dw; \theta, \beta) \right\} I\{t \leq s\}, \quad (35)$$

where $\rho_F(\cdot; \theta)$, the renewal function, is given in (1).

Proof: Result (i) is just the Glivenko-Cantelli theorem, whereas result (ii) was established in Peña et al. (2001). $\square$

Using this lemma, it is immediate that as $n \rightarrow \infty$,

$$\sup_{(v, w) \in [0, s^*] \times [0, t^*]} \|\Sigma_n(v, w; \theta, \beta) - \Sigma(v, w; \theta, \beta)\| \xrightarrow{p} 0$$

where the limiting matrix

$$\Sigma(s, t; \theta, \beta) = \begin{bmatrix} \Sigma_{11}(s, t; \theta, \beta) & \Sigma_{12}(s; \theta, \beta) \\ \Sigma_{21}(s; \theta, \beta) & \sigma_{22}(s; \theta, \beta) \end{bmatrix} \quad (36)$$

has components given by

$$\begin{aligned}
\Sigma_{11}(s, t; \theta, \beta) &= \Sigma_{11A}(s; \theta, \beta) + \Sigma_{11B}(s, t; \theta, \beta); \\
\Sigma_{11A}(s; \theta, \beta) &= \int_0^s \varphi_F(v; \theta)^{\otimes 2} \beta \lambda_F(v; \theta) \exp\{-\beta \Lambda_F(v; \theta)\} dv; \\
\Sigma_{11B}(s, t; \theta, \beta) &= \int_0^t \varphi_F(w; \theta)^{\otimes 2} y(s, w; \theta, \beta) \lambda_F(w; \theta) dw; \\
\Sigma_{12}(s; \theta, \beta) &= \Sigma_{21}(s; \theta, \beta)^t = \int_0^s \frac{1}{\beta} \varphi_F(v; \theta) \beta \lambda_F(v; \theta) \exp\{-\beta \Lambda_F(v; \theta)\} dv; \\
\sigma_{22}(s; \theta, \beta) &= \int_0^s \frac{1}{\beta^2} \beta \lambda_F(v; \theta) \exp\{-\beta \Lambda_F(v; \theta)\} dv.
\end{aligned}$$

Also, observe that the limiting matrix in (36) can be decomposed into

$$\Sigma(s, t; \theta, \beta) = \int_0^s \begin{bmatrix} \varphi(v; \theta) \\ 1/\beta \end{bmatrix}^{\otimes 2} \beta \lambda_F(v; \theta) \exp\{-\beta \Lambda_F(v; \theta)\} dv + \int_0^t \begin{bmatrix} \varphi(w; \theta) \\ 0 \end{bmatrix}^{\otimes 2} y(s, w; \theta, \beta) \lambda_F(w; \theta) dw.$$

Furthermore, with

$$\mathbf{I}_G(s, t; \theta, \beta) = - \begin{bmatrix} \frac{\partial^2}{\partial \theta \partial \theta^t} & \frac{\partial^2}{\partial \theta \partial \beta} \\ \frac{\partial^2}{\partial \beta \partial \theta^t} & \frac{\partial^2}{\partial \beta^2} \end{bmatrix} \mathbf{I}_G(s, t; \theta, \beta) \quad (37)$$

the generalized observed Fisher information process, it is straightforward to show that as $n \rightarrow \infty$,

$$\sup_{(v, w) \in [0, s^*] \times [0, t^*]} \left\| \frac{1}{n} \mathbf{I}_G(v, w; \theta, \beta) - \Sigma(v, w; \theta, \beta) \right\| \xrightarrow{p} 0.$$

The limiting results pertaining to Σ_n and $\mathbf{I}_G$ are conditions in Borgan (1984)'s set of regularity conditions, and from his theorem the following results for the recurrent event setting follow.

Theorem 1 Let $(s^*, t^*) \in \mathcal{R}_+^2$ such that $\Sigma(s^*, t^*; \theta, \beta)$ is positive definite. Under the 'usual regularity conditions' in Borgan (1984),

- i. (Consistency) There exists a sequence of solutions $(\widehat{\theta}_n(s^*, t^*), \widehat{\beta}_n(s^*, t^*))^t$ to the sequence of equations $\mathbf{U}_G(s^*, t^*; \theta, \beta) = \mathbf{0}$ satisfying, as $n \rightarrow \infty$,

$$\left\| \begin{pmatrix} \widehat{\theta}_n(s^*, t^*) \\ \widehat{\beta}_n(s^*, t^*) \end{pmatrix} - \begin{pmatrix} \theta \\ \beta \end{pmatrix} \right\| \xrightarrow{p} 0;$$

- ii. (Asymptotic Normality) As $n \rightarrow \infty$,

$$\sqrt{n} \left[\begin{pmatrix} \widehat{\theta}_n(s^*, t^*) \\ \widehat{\beta}_n(s^*, t^*) \end{pmatrix} - \begin{pmatrix} \theta \\ \beta \end{pmatrix} \right] \xrightarrow{d} MVN\left(0, \Sigma(s^*, t^*; \theta, \beta)^{-1}\right),$$

where MVN means multivariate normal. Furthermore, $\frac{1}{n} \mathbf{I}_G(s^*, t^*; \widehat{\theta}_n(s^*, t^*), \widehat{\beta}_n(s^*, t^*))$ is a consistent estimator of $\Sigma(s^*, t^*; \theta, \beta)$.

The following corollary is immediate from Theorem 1.

Corollary 1 Under the conditions of Theorem 1, as $n \rightarrow \infty$,

$$\sqrt{n} \left(\widehat{\theta}_n(s^*, t^*) - \theta \right) \xrightarrow{d} MVN(0, \Sigma^{11}(s^*, t^*; \theta, \beta)),$$

where $\Sigma^{11}(s^*, t^*; \theta, \beta) = (\Gamma(s^*, t^*; \theta, \beta) + \Xi(s^*, \theta, \beta))^{-1}$ with

$$\Gamma(s, t; \theta, \beta) = \int_0^t \varphi_F(w; \theta)^{\otimes 2} y(s, w; \theta, \beta) \lambda_F(w; \theta) dw; \quad (38)$$

$$\Xi(s; \theta, \beta) = \frac{\int_0^s \varphi_F(v; \theta)^{\otimes 2} \beta \lambda_F(v; \theta) \exp\{-\beta \Lambda_F(v; \theta)\} dv - \left(\int_0^s \varphi_F(v; \theta) \beta \lambda_F(v; \theta) \exp\{-\beta \Lambda_F(v; \theta)\} dv \right)^{\otimes 2}}{\left(\int_0^s \beta \lambda_F(v; \theta) \exp\{-\beta \Lambda_F(v; \theta)\} dv \right)}. \quad (39)$$

Since of main interest to us is the case where $t \rightarrow \infty$, we observe that because $y(s, t; \theta, \beta) = 0$ whenever $s < t$, then the limiting function arising in (38) as $t \rightarrow \infty$ is

$$\Gamma(s; \theta, \beta) \equiv \lim_{t \rightarrow \infty} \Gamma(s, t; \theta, \beta) = \int_0^s \varphi_F(w; \theta)^{\otimes 2} y(s, w; \theta, \beta) \lambda_F(w; \theta) dw. \quad (40)$$

With a view towards investigating the loss in efficiency when one uses a fully nonpara-metric estimator of the inter-event survivor function $\bar{F}$, from Corollary 1 and the δ -method, we are able to get the asymptotic distribution of the parametric estimator $\widehat{\bar{F}}(t)$ of $\bar{F}(t)$ in (17).

Corollary 2 Under the conditions of Corollary 1, as $n \rightarrow \infty$,

$$\sqrt{n} \left(\widehat{\bar{F}}_n(t) - \bar{F}(t; \theta) \right) \xrightarrow{d} N(0, \eta^2(t; \theta, \beta))$$

where the variance function is

$$\eta^2(t; \theta, \beta) = \exp\{-2\Lambda_F(t; \theta)\} (\nabla_{\theta} \Lambda_F(t; \theta))^t [\Gamma(\theta, \beta) + \Xi(\theta, \beta)]^{-1} (\nabla_{\theta} \Lambda_F(t; \theta)), \quad (41)$$

where $\Gamma(\theta, \beta)$ and $\Xi(\theta, \beta)$ are obtained from (38) and (39), respectively, by letting s and t both tend to ∞ .

Note that under regularity conditions,

$$\nabla_{\theta} \Lambda_F(t; \theta) = \int_0^t \varphi_F(v; \theta) \lambda_F(v; \theta) dv.$$

We remark here that the result in Corollary 2 could be strengthened to the weak convergence of the process $\sqrt{n} \left(\widehat{\bar{F}}_n(\cdot) - \bar{F}(\cdot; \theta) \right)$ to a Gaussian process, see Adekpédjou (2007)'s dissertation for this extended result; however, for the specific purpose of performing efficiency comparisons, the weaker version in Corollary 2 suffices.

4 Efficiency Comparisons

The second goal of this article is to examine efficiency issues in the presence of an informative monitoring period. The major reason for introducing the generalized KG model for recurrent events is in order to have an analytically tractable model which facilitates the investigation of some efficiency questions with recurrent event data. Through the informative monitoring model, we seek to provide concrete answers to the following questions. (i) If one employs an estimator of θ that was derived ignoring the informativeness of G on F , how much efficiency loss is incurred? (ii) If one uses an estimator of θ which was derived ignoring the additional event recurrences after the first event, but which takes into account the informativeness of G on F , what is the cost in terms of efficiency? This further leads to the question of how much information is contributed by the event occurrences past the first occurrence. (iii) If one adopts the fully nonparametric estimator of $F(t)$ studied in Peña et al. (2001), but the true underlying model is the generalized KG model, how much efficiency is sacrificed?

4.1 Ignoring Informative Monitoring

We now address the first question. We suppose that $\tilde{\theta}_n(s^*)$ is the estimator of θ that was derived ignoring the informativeness of G on θ . Then, it is not difficult to see that the limiting variance of $\sqrt{n}\tilde{\theta}_n(s^*)$ equals $\Gamma(s^*; \theta, \beta)^{-1}$, where the matrix Γ is defined in (38). With $\det \mathbf{Q}$ denoting the determinant of a matrix $\mathbf{Q}$, a measure of the change in asymptotic relative efficiency of the sequence $\{\tilde{\theta}_n(s^*)\}$ relative to the sequence $\{\hat{\theta}_n(s^*)\}$ is

$$\begin{aligned} \Delta \text{ARE} \left(\left\{ \hat{\theta}_n(s^*) \right\} : \left\{ \tilde{\theta}_n(s^*) \right\} \right) &= \det \left[\Gamma(s^*; \theta, \beta)^{-1} (\Gamma(s^*; \theta, \beta) + \Xi(s^*; \theta, \beta)) - \mathbf{I} \right] \\ &= \det \left[\Gamma(s^*; \theta, \beta)^{-1} \Xi(s^*; \theta, \beta) \right] = \frac{\det \Xi(s^*; \theta, \beta)}{\det \Gamma(s^*; \theta, \beta)}. \end{aligned}$$

Let us examine further the situation when $s^* \rightarrow \infty$. Denote by

$$\bar{\varphi}_F(\cdot; \theta) = \varphi_F(\Lambda_F^{-1}(\cdot; \theta); \theta) \quad (42)$$

with $\Lambda_F^{-1}(\cdot; \theta)$ being the inverse of $\Lambda_F(\cdot; \theta)$. Then, making the variable substitution $z = \beta \Lambda_F(v; \theta)$ in the expression for $\Xi(\theta, \beta)$, we obtain the compact expression

$$\begin{aligned} \Xi(\theta, \beta) &= \int_0^\infty \bar{\varphi}_F(z/\beta; \theta)^{\otimes 2} \exp(-z) dz - \left(\int_0^\infty \bar{\varphi}_F(z/\beta; \theta) \exp(-z) dz \right)^{\otimes 2} \\ &= \text{Cov} \left[\bar{\varphi}_F \left(\frac{Z}{\beta}; \theta \right) \right], \end{aligned} \quad (43)$$

where Z is a unit exponential random variable. Note in particular that $\Xi(\theta, \beta)$, being a covariance matrix, is nonnegative definite, so $\det \Xi(\theta, \beta) \geq 0$. Since $\Gamma(\theta, \beta)$ is also a limiting covariance matrix, hence is nonnegative definite, then this establishes that

$$\Delta \text{ARE} \left(\left\{ \hat{\theta}_n(\infty) \right\} : \left\{ \tilde{\theta}_n(\infty) \right\} \right) \geq 0,$$

as is to be expected. Analogously, by straightforward manipulations and with Z still a unit exponential variable, we are able to find the more compact expression for $\Gamma(\theta, \beta)$ given by

$$\Gamma(\theta, \beta) = \left(\frac{1}{1+\beta} \right) \text{Cov} \left[\bar{\varphi}_F \left(\frac{Z}{1+\beta}; \theta \right) \right] + \int_0^\infty \bar{\varphi}_F(v; \theta)^{\otimes 2} \exp(-v) \left[\int_v^\infty \rho_F \left(\Lambda_F^{-1}(z; \theta) - \Lambda_F^{-1}(v; \theta); \theta \right) \beta \exp(-\beta z) dz \right] dv. \quad (44)$$

In the special case where the recurrent event accrual follows a homogeneous Poisson process so that $\Lambda_F(t; \theta) = \theta t$, we are able to obtain closed form expressions for the above quantities. In this situation, $\phi_F(t; \theta) = 1/\theta$ and the renewal function is $\rho_F(t; \theta) = \theta t I(t \geq 0)$. Since $\phi_F(\cdot; \theta)$ is constant in t , this immediately shows that $\Xi(\theta, \beta) = 0$. Straightforward calculations also show that in this exponential case,

$$\Gamma(\theta, \beta) = \frac{1}{\theta^2 \beta}. \quad (45)$$

Observe that the inverse (reciprocal) of this quantity equals the (1, 1)th element of $\Sigma(\infty; \theta, \beta)$ given in (21) which was obtained via a limiting operation on the observed Fisher information.

Therefore, when F is the exponential distribution, $\Delta \text{ARE}(\hat{\theta}; \tilde{\theta}) = 0$, that is, there is no loss of efficiency by ignoring the fact that $\bar{G}(t; \theta, \beta) = \exp(-\beta \theta t)$ is informative about $\bar{F}(t; \theta) = \exp(-\theta t)$. Surprising as it may look at first, this actually is a logical result for in this case, from (19) and by an easy calculation,

$$\hat{\theta}_n = \tilde{\theta}_n = \frac{\sum_{i=1}^n K_i}{\sum_{i=1}^n \tau_i}, \quad (46)$$

that is, both estimators of θ are just the occurrence-exposure rate. However, for other non-exponential F where $\phi_F(\cdot; \theta)$ is not constant, then $\det \Xi(\theta, \beta)$ need not be zero, hence there could be loss in efficiency by ignoring the informativeness of G for F , as in the situation of Weibull inter-event times dealt with via simulations in Section 5.

4.2 Single-Event versus Recurrent Event Methods

To address the second efficiency question, denote by $\check{\theta}_n$ the estimator of θ based only on the possibly right-censored first event times, that is, the single-event estimator, but taking into account the informativeness of G for F . Then, the only technical change is that the appropriate $y(\infty, t; \theta, \beta)$ function to utilize in the $\Gamma(\theta, \beta)$ expression is

$$y_1(\infty, t; \theta, \beta) = \bar{F}(t; \theta) \bar{G}(t; \theta, \beta) = \exp \{ -(1+\beta) \Lambda_F(t; \theta) \}. \quad (47)$$

As a consequence, the limiting variance of $\sqrt{n} \check{\theta}_n$ is $[\Gamma_1(\theta, \beta) + \Xi(\theta, \beta)]^{-1}$, where

$$\begin{aligned} \Gamma_1(\theta, \beta) &= \int_0^\infty \varphi_F(v; \theta)^{\otimes 2} \exp \{ -(1+\beta) \Lambda_F(v; \theta) \} \lambda_F(v; \theta) dv \\ &= \left(\frac{1}{1+\beta} \right) \text{Cov} \left[\bar{\varphi}_F \left(\frac{Z}{1+\beta}; \theta \right) \right], \end{aligned} \quad (48)$$

with Z a unit exponential random variable. With $\Gamma_2(\theta, \beta)$ denoting the second term in the expression for $\Gamma(\theta, \beta)$ in (44), that is,

$$\Gamma_2(\theta, \beta) = \int_0^\infty \varphi_F(v; \theta)^{\otimes 2} \exp(-v) \times \left[\int_v^\infty \rho_F(\Lambda_F^{-1}(z; \theta) - \Lambda_F^{-1}(v; \theta); \theta) \beta \exp(-\beta z) dz \right] dv, \quad (49)$$

then the change in asymptotic relative efficiency of the sequence $\{\widehat{\theta}_n\}$ relative to $\{\check{\theta}_n\}$ is

$$\begin{aligned} \Delta \text{ARE}(\{\widehat{\theta}_n\} : \{\check{\theta}_n\}) &= \det \left[(\Gamma_1(\theta, \beta) + \Xi(\theta, \beta))^{-1} (\Gamma_1(\theta, \beta) + \Gamma_2(\theta, \beta) + \Xi(\theta, \beta)) - \mathbf{I} \right] \\ &= \frac{\det \Gamma_2(\theta, \beta)}{\det (\Gamma_1(\theta, \beta) + \Xi(\theta, \beta))}. \end{aligned} \quad (50)$$

Again, this will always be positive, indicating that there will always be gain in efficiency by utilizing the additional event occurrences.

In the special case where F is the exponential distribution, simple calculations reveal that

$$\Gamma_1(\theta, \beta) = \frac{1}{\theta^2(1+\beta)} \quad \text{and} \quad \Gamma_2(\theta, \beta) = \frac{1}{\theta^2\beta(1+\beta)}.$$

As a consequence, when the inter-event times are exponentially distributed,

$$\Delta \text{ARE}(\{\widehat{\theta}_n\} : \{\check{\theta}_n\}) = \frac{1}{\beta}, \quad (51)$$

which as noted in Section 2 is the expected number of event occurrences in each unit's monitoring period. This result could further be interpreted as follows. When $\beta \rightarrow 0$, then the monitoring period lengthens, consequently more recurrences will be observed (per unit), which will provide more information, thereby making $\{\widehat{\theta}_n\}$ more efficient relative to $\{\check{\theta}_n\}$ at the incremental order of $1/\beta$. Whereas, when $\beta \rightarrow \infty$, then the monitoring period shortens so there will either be no event observed or more likely just a single event observed (per unit), so in this situation, the two estimators becomes very close, hence the gain in efficiency goes down to zero.

4.3 Optimal Design Choice

Viewed in a different light, this result for the exponentially-distributed inter-event times implies that if experimenter #1 has a sample of size n_1 and uses the recurrences for estimating θ , then experimenter #2 will need a sample of size $n_2 = n_1(1 + 1/\beta)$ to gain the same (asymptotic) precision if he/she ignores the recurrences and use only the possibly right-censored times to first event occurrence, that is, single-event methods. The flip-side to these two approaches in terms of cost is that experimenter #1 will take more time to perform the study compared to experimenter #2, assuming that all units are entered into the study at the same time. A choice of which design to utilize entails taking into consideration costs associated with the experimental units and the duration of performing the study.

To amplify on the choice of study design, let us assume that C_1 is the cost per experimental unit, while C_2 is the cost incurred per unit of time while the study is ongoing. Let there be n_1 units in study design #1 which monitors all event occurrences over the units monitoring periods $[0, \tau_i]$ s. For this study, the overall study duration will be $\max_{i \leq n_1} \tau_i$. On the other hand, let n_2 be the number of units for study design #2 which monitors only the occurrence of the first event time or when it gets right-censored by τ_i . The overall study duration in this case will be

$\max_{i \leq n_2} (\tau_i \wedge T_{i1})$. Therefore, the expected total costs for each of these study designs will be, respectively,

$$\begin{aligned} TC_1(n_1; \theta, \beta) &= C_1 n_1 + C_2 E \left\{ \max_{i \leq n_1} \tau_i \right\}; \\ TC_2(n_2; \theta, \beta) &= C_1 n_1 + C_2 E \left\{ \max_{i \leq n_2} (\tau_i \wedge T_{i1}) \right\}. \end{aligned}$$

In the case of an exponential F , note that $\tau_i \stackrel{d}{=} V_i / (\beta\theta)$ and $\tau_i \wedge T_{i1} \stackrel{d}{=} V_i / ((\beta+1)\theta)$ where the V_i s are i.i.d. unit exponential variables. By recalling that

$$E\{V_{(n)}\} = \sum_{j=1}^n \frac{1}{n-j+1} \approx \log(n)$$

where $V_{(n)}$ is the largest order statistic among $V_1, V_2, \dots, V_n$ i.i.d. $\text{EXP}(1)$, then

$$\begin{aligned} TC_1(n_1; \theta, \beta) &\approx C_1 n_1 + C_2 \log(n_1) / (\beta\theta); \\ TC_2(n_2; \theta, \beta) &\approx C_1 n_2 + C_2 \log(n_2) / ((\beta+1)\theta). \end{aligned}$$

Since the two study designs will lead to the same asymptotic precision when $n_2 = n_1(1 + 1/\beta)$, then study design #1 would be cost-preferable to study design #2 if $TC_1(n_1; \theta, \beta) < TC_2(n_1(1 + 1/\beta); \theta, \beta)$. Using the above approximations, this inequality will occur if

$$\frac{C_1}{C_2} > \frac{1}{n_1 \theta (\beta+1)} \log \left[n_1 \left(\frac{\beta}{\beta+1} \right)^{\frac{\beta}{\beta+1}} \right]. \quad (52)$$

Of course, for this decision criterion to be usable, some prior or pilot estimates of θ and β will be needed.

4.4 Efficiency of Generalized PLE

We now address the third question posed earlier which pertains to the efficiency of the fully nonparametric estimator of the inter-event distribution F relative to an estimator derived using information about the structure of F and G . The KG model in the single-event settings was used for these efficiency studies, so the extended KG model for our recurrent event setting is a justifiably reasonable model to perform analogous efficiency studies. Peña et al. (2001) obtained the generalized product-limit estimator (GPLe) in this recurrent event setting. This estimator of the inter-event time survivor function $\bar{F}(t)$ is given by

$$\hat{\bar{F}}_n(s^*, t) = \prod_{w \leq t} \left[1 - \frac{\sum_{i=1}^n N_i(s^*, dw)}{\sum_{i=1}^n Y_i(s^*, w)} \right] \quad (53)$$

where $N_i(s, t) = \int_0^s Z_i(v, t) N_i^\dagger(dv)$ and $Y_i(\cdot, \cdot)$ is the generalized at-risk process defined in (34).

It was established in their paper that $\hat{\bar{F}}_n(s, t)$ is asymptotically normal with mean $\bar{F}(t)$ and asymptotic variance $\sigma^2(s, t; \theta, \beta)/n$, where

$$\sigma^2(s, t; \theta, \beta) = \bar{F}(t; \theta)^2 \int_0^t \frac{\Lambda_F(dw; \theta)}{y(s, w; \theta, \beta)} \quad (54)$$

with $y(s, t; \theta, \beta)$ is the function given in (35). We compare this limiting variance with the limiting variance of the parametric estimator of $\bar{F}(t; \theta)$ which is $\eta^2(t; \theta, \beta)$ in (41). The comparison is when $s^* \rightarrow \infty$.

Theorem 2 For the generalized KG model in this recurrent event setting, the asymptotic relative efficiency of the GPLE $\tilde{\bar{F}}$ relative to the parametric estimator $\tilde{\bar{F}}$ at time point $\Lambda_F^{-1}(t)$ is

$$ARE\left(\tilde{\bar{F}} : \widehat{\bar{F}}\right) = \frac{\left(\int_0^t \bar{\varphi}_F(w; \theta) dw\right)^t [\Gamma(\theta, \beta) + \Xi(\theta, \beta)]^{-1} \left(\int_0^t \bar{\varphi}_F(w; \theta) dw\right)}{\int_0^t \{[1 + Q(w; \theta, \beta)] \exp(-(1+\beta)w)\}^{-1} dw},$$

where

$$Q(w; \theta, \beta) = \int_0^\infty \rho_F \left[\Lambda_F^{-1}(v+w; \theta) - \Lambda_F^{-1}(w; \theta) \right] \beta \exp(-\beta v) dv. \quad (55)$$

Proof: The desired asymptotic relative efficiency is the ratio between $\eta^2(\Lambda_F^{-1}(t; \theta); \theta, \beta)$ and $\sigma^2(\Lambda_F^{-1}(t; \theta); \theta, \beta)$. The expressions in the theorem follow via straightforward manipulations of the expressions for $\eta^2(t; \theta, \beta)$ and $\sigma^2(t; \theta, \beta)$, by using the variable transformation $v = \Lambda_F(t)$ in the integrals, and noting that $\nabla_\theta \Lambda_F(t; \theta) = \int_0^t \psi(w; \theta) \lambda_F(w; \theta) dw$. $\parallel$

Corollary 3 For $\bar{F}(t; \theta) = \exp(-\theta t)$, the exponential case, the ARE of the GPLE relative to the parametric estimator at time point $\Lambda_F^{-1}(t)$ is

$$ARE\left(\tilde{\bar{F}} : \widehat{\bar{F}}\right) = \frac{[(1+\beta)t]^2}{\exp\{(1+\beta)t\} - 1} = \frac{p(t; \beta)}{1 - p(t; \beta)} (-\log p(t; \beta))^2$$

with $p(t; \beta) = \Pr\{(\tau_i \wedge T_{i1}) > \Lambda_F^{-1}(t; \theta)\} = \bar{F}(\Lambda_F^{-1}(t)) \bar{G}(\Lambda_F^{-1}(t))$.

Proof: In this exponential setting, we have already seen from earlier calculations that $\Xi(\theta, \beta) = 0$, while $\Gamma(\theta, \beta) = 1/(\theta^2 \beta)$. Furthermore, $\Lambda_F^{-1}(t; \theta) = t/\theta$, $\rho_F(t; \theta) = \theta t$, $\phi_F(t; \theta) = 1/\theta$, so that $\bar{\varphi}_F(t; \theta) = 1/\theta$. The expressions follow upon simplifying the ARE and noting the above facts. $\parallel$

From the ARE expression in Corollary 3, the following properties are easily established.

Corollary 4 The $ARE(\tilde{\bar{F}} : \widehat{\bar{F}})$ in Corollary 3 satisfies (i) $\lim_{\beta \rightarrow \infty} ARE(\tilde{\bar{F}} : \widehat{\bar{F}}) = 0$; (ii) $\lim_{\beta \rightarrow 0} ARE(\tilde{\bar{F}} : \widehat{\bar{F}}) = t/[\exp(t) - 1]$; and (iii) if p_0 is the solution of $\exp(p - 1) - \sqrt{p} = 0$, then

$$\sup_{p(t;\beta) \in [0,1]} ARE(\tilde{F} : \widehat{F}) = \frac{p_0(-\log p_0)^2}{1 - p_0} \approx 0.65.$$

The ARE expression in Corollary 3 for the exponential inter-event time distribution as a function of $p(t; \beta)$ is plotted in Figure 1. In Corollary 4, the result in (i) indicates that the shorter the monitoring periods become, which happens when $\beta \rightarrow \infty$, then the more inefficient the GPLE becomes. This may seem surprising at first sight since one may think that there will be very few observed recurrences; however, because of the informativeness of G for F , then the right-censored observations will contain information about F , and this is being exploited by the parametric estimator. On the other hand, when $\beta \rightarrow 0$, then the monitoring periods become longer, and in this case the impact of the right-censored observations will tend to be negligible because more and more complete observations will be observed. Indeed, this is manifested by observing that the limiting ARE expression in (ii) of Corollary 4 is the ARE of the parametric estimator $\widehat{F}(t) = \exp(-t/\bar{T})$, where $\bar{T}$ is the sample mean, relative to the nonparametric empirical survivor function estimator $\tilde{F}(t) = \sum_{i=1}^n I\{T_i > t\}/n$ based on an i.i.d. sample $T_1, T_2, \dots, T_n$ from a unit exponential distribution. It is quite interesting to observe the similarity with the efficiency results in single-event settings. The upper bound for the ARE in (iii) is similar to that obtained by Cheng and Lin (1987) concerning the Kaplan-Meier estimator when compared to the estimator that exploits the informative structure of the KG model in the single-event setting.

More generally, it is of interest to know if the ARE expression in Theorem 2 is always bounded above by unity. A partial answer is provided by the following theorem. Below, for a matrix $\mathbf{A}$, $\text{tr}(\mathbf{A})$ is its trace.

Theorem 3 The $ARE(\tilde{F} : \widehat{F})$ expression in Theorem 2 is bounded above by

$$\text{tr} \left\{ \left(\mathbf{I} + \Gamma(\theta, \beta)^{-1} \Xi(\theta, \beta) \right)^{-1} \right\} = \sum_{j=1}^p \frac{1}{1 + e_j(\theta, \beta)},$$

where $e_j(\theta, \beta)$, $j = 1, 2, \dots, p$, are the eigenvalues of $\Gamma(\theta, \beta)^{-1} \Xi(\theta, \beta)$. In particular, if the parameter θ is one-dimensional, then $ARE(\tilde{F} : \widehat{F}) \leq 1$.

Proof: To establish the result, we first observe that if $\mathbf{f} : \mathcal{R} \rightarrow \mathcal{R}^p$ and $g : \mathcal{R} \rightarrow \mathcal{R}$ with square-integrable components with respect to a measure μ , and if $\mathbf{S}$ is a positive definite $p \times p$ symmetric matrix, then

$$\left(\int \mathbf{f}^t g d\mu \right) \mathbf{S}^{-1} \left(\int \mathbf{f} g d\mu \right) \leq \left(\int g^2 d\mu \right) \text{tr} \left(\mathbf{S}^{-1} \int \mathbf{f} \mathbf{f}^t d\mu \right). \quad (56)$$

This inequality follows by first letting $\tilde{\mathbf{f}} = \mathbf{S}^{-1/2} \mathbf{f}$, and then noting that the left-hand side of (56) equals

$$\left(\int \mathbf{f}^t g d\mu \right) \mathbf{S}^{-1} \left(\int \mathbf{f} g d\mu \right) = \left(\int \tilde{\mathbf{f}}^t g d\mu \right)^t \left(\int \tilde{\mathbf{f}} g d\mu \right) = \sum_{j=1}^p \left(\int \tilde{f}_j g d\mu \right)^2.$$

Applying the Cauchy-Schwartz Inequality to each of the p terms, we obtain

$$\sum_{j=1}^p \left(\int \bar{f}_j g d\mu \right)^2 \leq \sum_{j=1}^p \left(\int \bar{f}_j^2 d\mu \right) \left(\int g^2 d\mu \right).$$

The inequality in (56) follows by then noting that

$$\sum_{j=1}^p \left(\int \bar{f}_j^2 d\mu \right) = \int \mathbf{f}^T \mathbf{f} d\mu = \int \mathbf{f}^T \mathbf{S}^{-1} \mathbf{f} d\mu = \int \text{tr}(\mathbf{S}^{-1} \mathbf{f} \mathbf{f}^T) d\mu = \text{tr}(\mathbf{S}^{-1} \int \mathbf{f} \mathbf{f}^T d\mu).$$

To utilize this result in the context of the ARE expression, define the following.

$$\begin{aligned} \mathbf{f}(w; \theta) &= \varphi_F(w; \theta) \sqrt{y(\infty, w; \theta, \beta)}; \\ g(w; \theta) &= I\{w \leq t\} / \sqrt{y(\infty, w; \theta, \beta)}; \\ \mathbf{S}(\theta, \beta) &= \Gamma(\theta, \beta) + \Xi(\theta, \beta); \\ d\mu(w; \theta) &= \lambda_F(w; \theta) dw. \end{aligned}$$

Then, a direct application of (56) to the ARE expression in Theorem 3 leads to the upper bound in the statement of the theorem. The case of $p = 1$ is an immediate consequence. $\parallel$

Observe therefore that if the parameter of the inter-event distribution is one-dimensional, then the parametric estimator of the inter-event survivor function will never be less efficient, asymptotically, than the generalized product-limit estimator under this generalized KG model. However, if the parameter vector θ is more than one-dimensional, then it is possible that the ARE will not have an upper bound of unity.

5 Simulation Studies

Obtaining exact analytical efficiency expressions under inter-event distributions other than the exponential distribution is difficult since closed form expressions for the renewal function, which appears in the variance expression, is not generally available. To examine non-exponential inter-event distributions, we resorted to computer simulation studies. Specifically, we considered in the simulation a Weibull inter-event distribution with shape parameter θ_1 and scale parameter θ_2 . The purpose of the simulation study was to compare the efficiency of the estimators $\hat{\theta}$, $\tilde{\theta}$, and $\check{\theta}$ as well as to ascertain the efficiency of the GPLE $\tilde{F}(t)$ relative to the parametric estimator $\hat{F}(t)$.

The simulation code was in the R language (Ihaka and Gentleman (1996)), and a Newton-Raphson procedure as described in the subsection dealing with Weibull inter-event distribution was implemented. A total of $M = 5000$ simulation replications were performed, where for each replication, a recurrent event data following the generalized KG model was generated for combinations of values of n , θ_1 , θ_2 , and β . For each of the resulting recurrent event data, the estimates of $\hat{\theta}$, $\tilde{\theta}$, and $\check{\theta}$, as well as $\hat{\beta}$ and $\check{\beta}$ were obtained. The estimates $\hat{F}(t)$ and $\tilde{F}(t)$ of $F(t)$ were also obtained for t -values coinciding with the percentiles of the true Weibull distribution.

As a measure of the efficiency of $\hat{\theta}$ over $\tilde{\theta}$, we computed the estimate of its (generalized) mean-squared error (MSE) given by

$$\widehat{MSE}(\widehat{\theta}, \theta_0) = \frac{1}{M} \sum_{m=1}^M \left[(\widehat{\theta}_m - \theta_0)(\widehat{\theta}_m - \theta_0)^t \right] \quad (57)$$

where $\widehat{\theta}_m = (\widehat{\theta}_{1m}, \widehat{\theta}_{2m})^t$ is the estimate obtained from the m th replication. The true θ value is denoted by $\theta_0 = (\theta_{10}, \theta_{20})^t$. The estimate of the MSE for $\tilde{\theta}$ and $\check{\theta}$ are similarly defined. The estimate of the efficiency of $\widehat{\theta}$ over $\tilde{\theta}$ is then defined via

$$\text{Eff}(\widehat{\theta} : \tilde{\theta}) = \frac{\det[\widehat{MSE}(\tilde{\theta}, \theta_0)]}{\det[\widehat{MSE}(\widehat{\theta}, \theta_0)]}. \quad (58)$$

$\text{Eff}(\widehat{\theta} : \check{\theta})$ is analogously defined. In an analogous manner, the measure of efficiency of $\tilde{F}$ over $\hat{F}$ at time point t is the ratio of the MSEs of $F(t)$ and $\tilde{F}(t)$ computed over the M replications. The simulation was performed for combinations of

$$n \in \{20, 50, 100\}, \quad \theta_1 \in \{0.9, 1.0, 1.2, 1.5, 2.0\},$$

As is to be expected, the results were somewhat invariant with respect to a change in the scale parameter θ_2 , so we only report here those associated with $\theta_1 = 1.0$. To conserve space, we also only report the cases $n \in \{20, 50\}$ since the conclusions are unchanged.

Table 1 presents the efficiencies of the θ -estimators for the different cases. We have also provided the column *MeanEvs* which represents the mean number of events observed for each of the subjects. Note that when β is decreased, then there are more events observed. It is clear from the results of this simulation that the $\widehat{\theta}$ estimator is always more efficient than the other two estimators, as is to be expected. It is interesting to observe that as β increases, the efficiency gain of $\widehat{\theta}$ over $\tilde{\theta}$ increases, whereas it is the opposite direction for the comparison with the $\check{\theta}$ estimator. This could intuitively be explained by the fact that when there are fewer event occurrences, then the information coming from the τ_i s, which are being used in the $\widehat{\theta}$ estimator but which are ignored by the $\tilde{\theta}$ estimator, becomes more important. On the other hand, for the $\check{\theta}$ estimator, a decrease in β leads to more event occurrences but which are ignored by this estimator, hence the increase in efficiency of the $\widehat{\theta}$ estimator over the $\check{\theta}$ estimator. Interestingly, the efficiency behavior are very similar over changes in n for the $\widehat{\theta}$ and $\tilde{\theta}$ comparison. One may wonder why the $\widehat{\theta}$ estimator is still more efficient than the $\check{\theta}$ estimator even when the average number of events per subject is less than unity. An explanation for this is that the $\widehat{\theta}$ estimator still utilizes the values of τ , which provide additional information, even if there was only one event observed on a subject. The estimator $\check{\theta}$, on the other hand, only utilizes the minimum between the τ_i -value and the time-to-first event occurrence T_{i1} , together with the indicator of which was smaller. We point out that the exact closed-form expressions of the asymptotic covariance matrix for the estimator $\check{\theta}$ is obtainable for this Weibull case by exploiting the independence between the censoring indicator and the $\min\{T_{i1}, \tau_i\}$. We compared these asymptotic variances with the simulated variances and there is indeed good agreement when the sample size is large ($n \in \{50, 100\}$).

Figure 2 provides the plots of the efficiency of $\tilde{F}(t)$ over $\hat{F}(t)$ for different θ_1 and n values. For each of the plot frames, we superimposed the plots associated with the different values of β so

as to see the effect of changing β -value. We have plotted these graphs with the true value of $F(t)$ in the abscissa as a way to standardize the graphs. From these plots, it is evident that the efficiency behavior are basically very similar for the different combinations of θ_1 and n , as well as for varying values of θ_2 , which are not shown here. It is also clear that the relative efficiency of the GPLE over the parametric estimator never exceeds 0.70. It would be a mathematical challenge to know the exact upper bound of this relative efficiency in analogy to the approximate 0.65 that was found for the exponential inter-event times, but this appears to be a difficult problem owing to a non-closed form for the renewal function of a Weibull distribution. With regards to the impact of the monitoring parameter β , when this parameter is increased, then the relative efficiency of the GPLE decreases for larger values of t . This could be attributed to the fact that there will be fewer observations whose inter-event times are in this region, hence the GPLE suffers, aside from the fact that it does not utilize information coming from the τ_i s, whereas the parametric estimator is able to use information from all the observations, including the τ_i s, for estimating $F(t)$ for larger t s.

6 Concluding Remarks

In this paper we have examined efficiency aspects of estimators of the inter-event distribution in a recurrent event setting. A generalized Koziol-Green model was introduced to provide an analytically tractable model of an informative monitoring period. This enables the analytical assessment of efficiencies of estimators. Of particular interest was to study the loss in efficiency if the informative monitoring structure is ignored in the estimation procedure, and to see the gain in efficiency when one utilizes the event recurrences instead of just simply using the time-to-first, possibly right-censored, event occurrence. The generalized product-limit estimator of the inter-event distribution, which does not utilize the informative monitoring aspect, was also examined in terms of its loss in efficiency relative to the parametric estimator, the latter exploiting the informative structure. In a sequel article, a semiparametric estimator of the inter-event distribution based on recurrent event data, and which exploits the generalized KG structure, will be compared to the generalized product-limit estimator.

Acknowledgments

The authors wish to thank Dr. Alex McLain and Dr. Laura Taylor for helpful discussions, and also wish to thank a Referee and the Editor for their comments.

References

- Aalen O. Nonparametric inference for a family of counting processes. *Ann. Statist* 1978;6:701–726.
- Aalen, OO.; Borgan, Ø.; Gjessing, HK. *Survival and event history analysis: A process point of view.* Springer; New York: 2008. *Statistics for Biology and Health.*
- Adekpedjou, A. Ph.D. thesis. University of South Carolina; Columbia, South Carolina: 2007. *Estimation with Recurrent Event Data Under an Informative Monitoring Period.*
- Allen W. A note on conditional probability of failure when hazards are proportional. *Operations Research* 1963;11:658–659.
- Andersen, P.; Borgan, O.; Gill, R.; Keiding, N. *Statistical Models Based on Counting Processes.* Springer-Verlag; New York: 1993.
- Borgan Ø. Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scand. J. Statist* 1984;11:1–16.
- Breslow N, Crowley J. A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics* 1974;2:437–453.
- Chen Y, Hollander M, Langberg N. Small-sample results for the Kaplan-Meier estimator. *Journal of the American Statistical Association* 1982;77:141–144.
- Cheng PE, Lin GD. Maximum likelihood estimation of a survival function under the Koziol-Green proportional hazards model. *Statist. Probab. Lett* 1987;5:75–80.

- Cox D. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society* 1972;34:187–220.
- Csörgő S, Faraway JJ. The paradoxical nature of the proportional hazards model of random censorship. *Statistics* 1998;31:67–78.
- Efron, B. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Prentice-Hall; New York: 1967. The two-sample problem with censored data; p. 831–853.
- Fleming, T.; Harrington, D. *Counting Processes and Survival Analysis*. Wiley; New York: 1991.
- Gill RD. Nonparametric estimation based on censored observations of a Markov renewal process. *Z. Wahrsch. Verw. Gebiete* 1980;53:97–116.
- Gill RD. Testing with replacement and the product limit estimator. *Ann. Statist* 1981;9:853–860.
- Hollander M, Peña E. Families of confidence bands for the survival function under the general random censorship model and the Koziol-Green model. *Canadian Journal of Statistics* 1989;17:59–74.
- Ihaka R, Gentleman R. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 1996;5:299–314.
- Jacod J. Multivariate point processes: predictable projection, Radon-Nikodým derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 1974;31:235–253.
- Kalbfleisch, JD.; Prentice, RL. *The statistical analysis of failure time data*. 2nd ed. Wiley-Interscience [John Wiley & Sons]; Hoboken, NJ: 2002. Wiley Series in Probability and Statistics.
- Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc* 1958;53:457–481.
- Koziol J, Green S. A Cramer-von Mises statistic for randomly censored data. *Biometrika* 1976;63:139–156.
- Peña, E.; Hollander, M. Models for Recurrent Events in Reliability and Survival Analysis. In: Soyer, R.; Mazzuchi, T.; Singpurwalla, N., editors. *Mathematical Reliability: An Expository Perspective*. Kluwer Academic Publishers; 2004. p. 105–123.chap. 6
- Peña E, Strawderman R, Hollander M. Nonparametric estimation with recurrent event data. *J. Amer. Statist. Assoc* 2001;96:1299–1315.
- Sellke, T. *Statistical decision theory and related topics, IV, Vol. 2* (West Lafayette, Ind., 1986). Springer; New York: 1988. Weak convergence of the Aalen estimator for a censored renewal process; p. 183–194.
- Siannis F. Applications of a parametric model for informative censoring. *Biometrics* 2004;60:704–714. [PubMed: 15339293]
- Vardi Y. Nonparametric estimation in renewal processes. *Ann. Statist* 1982a;10:772–785.
- Vardi Y. Nonparametric estimation in the presence of length bias. *Ann. Statist* 1982b;10:616–620.
- Wang M-C, Chang S-H. Nonparametric estimation of a recurrent survival function. *J. Amer. Statist. Assoc* 1999;94:146–153.
- Wang M-C, Qin J, Chiang C-T. Analyzing recurrent event data with informative censoring. *J. Amer. Statist. Assoc* 2001;96:1057–1065.
- William AL. A Model for Informative Censoring. *J. Amer. Statist. Assoc* 1989;84:749–752.

Figure 1.

Asymptotic relative efficiency of the nonparametric estimator $\hat{\bar{F}}$ of the inter-event survivor function relative to the parametric estimator $\hat{\bar{F}}$ as a function of $p \equiv p(t; \beta) = F(\Lambda_F^{-1}(t; \theta); \theta)^{\beta+1}$ in the HPP case.

Figure 2.

Plots of efficiency of the GPLE $\tilde{F}$ relative to the parametric estimator $\hat{F}$ under a Weibull(θ_1 , θ_2) inter-event distribution for varying values of n and β . The curves are black and solid for $\beta = .3$, red and dashed for $\beta = .5$, and blue and dotted for $\beta = .7$. The value of θ_2 is 1.0. These curves are based on 5000 simulation replications.

Table 1

Relative efficiency (in decimal format) of $\hat{\theta}$ estimator relative to the $\tilde{\theta}$ and $\check{\theta}$ estimators under the Weibull (θ_1, θ_2) inter-event distribution for different β -values, for $n \in \{20, 50\}$, and for $\theta_2 = 1$. $MeanEvs$ represents the mean number of events observed for each of the subjects. Results are based on 5000 simulation replications.

n	θ_1	θ_2	β	MeanEvs	$Eff(\hat{\theta} : \tilde{\theta})$	$Eff(\hat{\theta} : \check{\theta})$
20	0.9	1	0.3	3.89	1.26	37.62
20	0.9	1	0.5	2.25	1.47	17.12
20	0.9	1	0.7	1.57	1.72	9.88
20	1.0	1	0.3	3.34	1.25	30.24
20	1.0	1	0.5	2.01	1.51	12.62
20	1.0	1	0.7	1.43	1.82	8.69
20	1.5	1	0.3	1.98	1.54	11.11
20	1.5	1	0.5	1.34	1.90	6.83
20	1.5	1	0.7	1.02	2.12	4.38
20	2.0	1	0.3	1.47	1.65	6.42
20	2.0	1	0.5	1.06	2.02	4.25
20	2.0	1	0.7	0.84	2.41	3.31
50	0.9	1	0.3	3.90	1.26	30.27
50	0.9	1	0.5	2.25	1.41	13.48
50	0.9	1	0.7	1.57	1.66	7.96
50	1.0	1	0.3	3.34	1.30	23.29
50	1.0	1	0.5	2.00	1.53	9.91
50	1.0	1	0.7	1.42	1.71	6.69
50	1.5	1	0.3	1.97	1.49	9.53
50	1.5	1	0.5	1.34	1.75	5.55
50	1.5	1	0.7	1.02	2.11	3.82
50	2.0	1	0.3	1.47	1.66	5.58
50	2.0	1	0.5	1.06	2.01	3.62
50	2.0	1	0.7	0.84	2.30	3.04