Ann Inst Stat Math (2011) 63:791–825
DOI 10.1007/s10463-009-0252-2

Fisher information in window censored renewal process data and its applications

Yanxing Zhao · **H. N. Nagaraja**

Received: 27 November 2007 / Revised: 1 May 2009 / Published online: 23 July 2009

Abstract Suppose we have a renewal process observed over a fixed length of time starting from a random time point and only the times of renewals that occur within the observation window are recorded. Assuming a parametric model for the renewal time distribution with parameter $\boldsymbol{\theta}$, we obtain the likelihood of the observed data and describe the exact and asymptotic behavior of the Fisher information (FI) on $\boldsymbol{\theta}$ contained in this window censored renewal process. We illustrate our results with exponential, gamma, and Weibull models for the renewal distribution. We use the FI matrix to determine optimal window length for designing experiments with recurring events when the total time of observation is fixed. Our results are useful in estimating the standard errors of the maximum likelihood estimators and in determining the sample size and duration of clinical trials that involve recurring events associated with diseases such as lupus.

Keywords Renewal process · Window censoring · Fisher information · Simulation · Exponential distribution · Gamma distribution · Experimental design

This research was supported in part by NIH grant P01 DK55546.

Y. Zhao
Novartis Pharmaceuticals, One Health Plaza, East Hanover, NJ 07936, USA
e-mail: yanxingzhao@hotmail.com

H. N. Nagaraja (✉)
Department of Statistics, The Ohio State University, 404 Cockins Hall,
1958 Neil Ave., Columbus, OH 43210-1247, USA
e-mail: hnn@stat.osu.edu

1 Introduction

We consider a renewal process (RP) represented by a sequence $\{X_n, n \geq 1\}$ of independent and identically distributed (iid) positive continuous random variables. The first unit is placed in operation at time zero; it fails at time X_1 and is immediately replaced by a new unit which then fails at time $X_1 + X_2$, and so on. Let f and F denote the probability density function (pdf) and cumulative distribution function (cdf) of the *renewal variable* X representing the common distribution of the X_i and we assume that $\mu_X = E(X)$ is finite.

Now suppose the RP has evolved over time when we start observing the process for a fixed period of length w. We name the process inside the observation window a window censored renewal process (WCRP). Waiting time for the first renewal within this window is the forward recurrence time (FRT) random variable Y and its long-run pdf is given by Karlin and Taylor (1975)

$$g(y) = \frac{1 - F(y)}{\mu_X}, \quad y \geq 0. \tag{1}$$

Let N denote the number of renewals observed in the window and $\mathcal{D}$ denote the data set obtained from one WCRP. The data set has three different profiles that depend on N. When $N = 0$, no event is observed inside the observation window as shown in Fig. 1a. When $N = 1$, as in Fig. 1b, we observe one Y, one Z and the data set is $\mathcal{D} = (N = 1, Y, Z)$, with Z denoting a right-censored renewal variable representing the time between the last renewal and the end of observation period. When $N \geq 2$ as in Fig. 1c, we observe one Y, one Z, $N - 1$ X's and the data set is $\mathcal{D} = (N = n, Y, \mathbf{X}, Z)$, with $\mathbf{X} = (X_1, \ldots, X_{N-1})$ denoting the renewal variables.

When we observe m multiple WCRPs, we assume that they are independent. If they have a common renewal distribution and window length, we have m iid WCRPs.

Renewal processes can be used to model recurrent events data that arise in a variety of real-life situations. Nelson (2003) gives several examples from product repairs and disease recurrences, and discusses graphical and formal non-parametric approaches for analyzing WCRP data. Vardi (1982) introduced a non-parametric method for estimating F from multiple iid WCRPs and Denby and Vardi (1985) discussed a short-cut method for estimating F based on iid WCRPs. They obtain the Kaplan–Meier

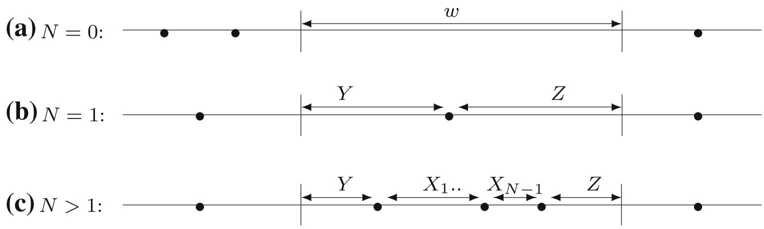

Fig. 1 Data profile of a WCRP; *filled circle* indicates a renewal and *vertical line* mark the start and the end of an observation window of length w

estimator after excluding observation windows without renewals. Soon and Woodroofe (1996) extended Vardi's work to non-arithmetic F and proved the consistency of non-parametric maximum likelihood estimates (MLE). Rigdon and Basu (2000) discuss parametric inference procedures for repairable systems with complete repairs and consider models based on homogeneous Poisson processes. Time truncated case there corresponds to a special WCRP data where the renewal distribution is exponential. Their models apply to data sets where the systems are observed from time zero. Thus left-censoring is not addressed in their models. Alvarez (2006) has recently considered an alternating renewal process (ARP) model in a parametric setting. He has discussed the maximum likelihood estimation of the parameters in a model with two states where the underlying distributions of 'on' and 'off' lifetimes are either geometric or exponential distributions. Alvarez (2005) considers ARP with more than two states and Weibull lifetime distributions. Here we focus on the Fisher information (FI) from WCRPs observed from parametric renewal distributions.

Let $\boldsymbol{\theta}$ denote the parameter vector associated with the renewal cdf F. In Sect. 2, we obtain the likelihood function for the data $\mathcal{D}$ and an expression for the FI about $\boldsymbol{\theta}$ in a WCRP. We describe the asymptotic properties of the FI matrix and a method to approximate its elements. In Sects. 3 and 4, we focus on the exponential and gamma renewal distributions, respectively, and study the properties of FI from these distributions. Section 5 is devoted to a similar study for the Weibull distribution. In Sect. 6, we illustrate the uses of FI in designing longitudinal studies involving RP data. The last section contains some concluding remarks. Proofs are collected in the Appendix.

Let $I(\theta; X)$ be the FI in a random variable X on the parameter θ and $I(\boldsymbol{\theta}; \mathcal{D})$ be the FI matrix on $\boldsymbol{\theta}$ in the data set $\mathcal{D}$. We denote an exponential random variable with mean β by Expo(β), and a gamma random variable with shape parameter α and scale parameter β by Gamma(α, β).

2 Likelihood function and Fisher information

2.1 Likelihood function

The likelihood function of a data set $\mathcal{D}$ from a WCRP depends on N. Let δ_n represent the indicator of n renewals inside the observation window, i.e.,

$$\delta_n = \begin{cases} 1 & \text{if } N = n \\ 0 & \text{otherwise,} \end{cases}$$

Theorem 1 *For the data set $\mathcal{D}$ generated from a WCRP with window length w, the likelihood function $L(\mathcal{D}; \boldsymbol{\theta})$ is given by*

$$L(\mathcal{D}; \boldsymbol{\theta}) = \sum_{n=0}^{\infty} \delta_n L_n(\mathcal{D}; \boldsymbol{\theta}), \tag{2}$$

where

$$
\begin{aligned}
L_0(\mathcal{D};\boldsymbol{\theta}) &= 1 - G(w),\\
L_1(\mathcal{D};\boldsymbol{\theta}) &= g(y)[1 - F(z)], \text{ where } z = w - y, \text{ and}\\
L_n(\mathcal{D};\boldsymbol{\theta}) &= g(y)f(x_1)\cdots f(x_{n-1})[1 - F(z)],\\
&\qquad \text{with } z = w - y \text{ if } n = 1 \text{ and } z = w - y - \sum_{i=1}^{n-1} x_i \text{ if } n \geq 2,\\
&\qquad 0 < y, z, x_i < w, i = 1, \ldots, n-1, \text{ for } n \geq 2.
\end{aligned}
\tag{3}
$$

The proof of Theorem 1 is given in Appendix A.1. Note that for any $\mathcal{D}$ only one term in the series in (2) is non-zero.

The log-likelihood corresponding to (2) is

$$
\ell(D;\boldsymbol{\theta}) = \delta_0 \log L_0(D;\boldsymbol{\theta}) + \delta_1 \log L_1(D;\boldsymbol{\theta}) + \sum_{n=2}^{\infty} \delta_n \log L_n(D;\boldsymbol{\theta}), \tag{4}
$$

and will be used in calculating the FI in $\mathcal{D}$.

The likelihood function of the data set $\mathbf{D}$ obtained from m iid WCRPs of length w is given by

$$
\begin{aligned}
L(\mathbf{D};\boldsymbol{\theta}) &= \prod_{i=1}^{m} L(\mathcal{D}_i;\boldsymbol{\theta})\\
&= \prod_{i=1}^{m} \Big[\delta_0[1 - G(w)] + \delta_1 g(y_i)[1 - F(z_i)]\\
&\quad + \sum_{n=2}^{\infty} \delta_n g(y_i) f(x_{i1}) \cdots f(x_{i(n-1)})[1 - F(z_i)]\Big].
\end{aligned}
$$

Since in each of the infinite series, only one term is non-zero, the joint likelihood is a product of finitely many terms. Now suppose n_w is the number of windows that do not have any renewals. From the remaining $(m - n_w)$ windows we collect all the X variables, Y variables, and Z variables. Let the number of X variables be denoted by n_x. The number of Y or Z variables is $(m - n_w)$. We relabel the variables as X_k, $k = 1, \ldots, n_x$, $Y_k, k = 1, \ldots, m - n_w$, and $Z_k, k = 1, \ldots, m - n_w$. After rearranging the factors, $L(\mathbf{D};\boldsymbol{\theta})$ can be written as

$$
L(\mathbf{D};\boldsymbol{\theta}) = [1 - G(w)]^{n_w} \times \prod_{k=1}^{m-n_w} g(y_k) \times \prod_{k=1}^{n_x} f(x_k) \times \prod_{k=1}^{m-n_w} [1 - F(z_k)]. \tag{5}
$$

Thus the log-likelihood function of a data set from m iid WCRPs can be expressed as

$$\begin{aligned}\ell(\mathbf{D};\boldsymbol{\theta}) &= \log L(\mathbf{D};\boldsymbol{\theta})\\ &= n_w \log[1-G(w)] + \sum_{k=1}^{m-n_w}\log g(y_k) + \sum_{k=1}^{n_x}\log f(x_k)\\ &\quad + \sum_{k=1}^{m-n_w}\log[1-F(z_k)]. \end{aligned} \tag{6}$$

Similar expressions for $L(\mathbf{D};\boldsymbol{\theta})$ and $\ell(\mathbf{D};\boldsymbol{\theta})$ can be written when the window lengths vary.

2.2 Fisher information in a WCRP

Theorem 2 *Consider a WCRP data $\mathcal{D}$ with renewal distribution parameter $\boldsymbol{\theta}$ and assume that all the needed regularity conditions are satisfied. Then the (k,l)th element of the FI matrix $\mathbf{I}(\boldsymbol{\theta};\mathcal{D}) = [I(\theta_k,\theta_l;\mathcal{D})]$ is given by*

$$\begin{aligned}I(\theta_k,\theta_l;\mathcal{D}) &= E\left[\left(\frac{\partial}{\partial\theta_k}\log L(\mathcal{D};\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_l}\log L(\mathcal{D};\boldsymbol{\theta})\right)\right]\\ &= A_0(\theta_k,\theta_l;\mathcal{D}) + \int_0^w A_1(\theta_k,\theta_l;\mathcal{D})g(y)[1-F(z)]\mathrm{d}y\\ &\quad + \sum_{n=2}^{\infty}\int_{y=0}^{w}\int_{x_1=0}^{w-y}\cdots\int_{x_{n-1}=0}^{w-y-\sum_{i=1}^{n-2}x_i}\\ &\quad \times A_n(\theta_k,\theta_l;\mathcal{D})L_n(\mathcal{D};\boldsymbol{\theta})\mathrm{d}x_{n-1}\cdots\mathrm{d}x_1\mathrm{d}y \end{aligned} \tag{7}$$

where

$$\begin{aligned}A_0(\theta_k,\theta_l;\mathcal{D}) &= \frac{\left(\frac{\partial}{\partial\theta_k}G(w)\right)\left(\frac{\partial}{\partial\theta_l}G(w)\right)}{1-G(w)},\\ A_1(\theta_k,\theta_l;\mathcal{D}) &= \left(\frac{\frac{\partial}{\partial\theta_k}g(y)}{g(y)} - \frac{\frac{\partial}{\partial\theta_k}F(z)}{1-F(z)}\right)\left(\frac{\frac{\partial}{\partial\theta_l}g(y)}{g(y)} - \frac{\frac{\partial}{\partial\theta_l}F(z)}{1-F(z)}\right),\\ A_n(\theta_k,\theta_l;\mathcal{D}) &= \left(\frac{\frac{\partial}{\partial\theta_k}g(y)}{g(y)} + \sum_{i=1}^{n-1}\frac{\frac{\partial}{\partial\theta_k}f(x_i)}{f(x_i)} - \frac{\frac{\partial}{\partial\theta_k}F(z)}{1-F(z)}\right)\\ &\quad \times\left(\frac{\frac{\partial}{\partial\theta_l}g(y)}{g(y)} + \sum_{i=1}^{n-1}\frac{\frac{\partial}{\partial\theta_l}f(x_i)}{f(x_i)} - \frac{\frac{\partial}{\partial\theta_l}F(z)}{1-F(z)}\right),\end{aligned}$$

and $L_n(\mathcal{D};\boldsymbol{\theta})$ and z are given by (3).

The proof of this theorem is given in Appendix A.2.

2.3 Limiting properties of Fisher information

Theorem 3 *Let the renewal random variable X have finite mean μ_X and $I(\theta; \mathcal{D})$ be the FI on θ from one WCRP with window length w. Under certain regularity conditions*

$$\lim_{w \longrightarrow \infty} \frac{I(\theta; \mathcal{D})}{w} = \frac{I(\theta; X)}{\mu_X}.$$

The regularity conditions are:

(RC1) $\frac{\partial}{\partial\theta} \log g(y; \theta)$ and $\frac{\partial^2}{\partial\theta^2} g(y; \theta)$ are finite for all $y > 0$ where g is given by (1).
(RC2) The FI on θ in the FRT variable Y from the RP is finite and the associated regularity conditions hold.

The proof is in Appendix A.3. If there are multiple parameters in the renewal distribution, the off-diagonal terms of the FI matrix have a similar property. If $I(\theta_k, \theta_l; \mathcal{D})$ is the (k, l)th entry of the FI matrix for $\boldsymbol{\theta}$ from a WCRP with window length w,

$$\lim_{w \longrightarrow \infty} \frac{I(\theta_k, \theta_l; \mathcal{D})}{w} = \frac{I(\theta_k, \theta_l; X)}{\mu_X}.$$

This, combined with Theorem 3 leads to the following result.

Theorem 4 *Let $\mathbf{I}(\boldsymbol{\theta}; \mathcal{D})$ be the FI matrix for $\boldsymbol{\theta}$ from a WCRP data with window length w. Then*

$$\lim_{w \longrightarrow \infty} \frac{1}{w} \mathbf{I}(\boldsymbol{\theta}; \mathcal{D}) = \frac{1}{\mu_X} \mathbf{I}(\boldsymbol{\theta}; X).$$

Any continuous function of the elements of the FI matrix possesses a similar limiting property. It holds for the trace, and in the case of the determinant,

$$\lim_{w \longrightarrow \infty} \frac{\det(\mathbf{I}(\boldsymbol{\theta}; \mathcal{D}))}{w^p} = \frac{\det(\mathbf{I}(\boldsymbol{\theta}; X))}{\mu_X^p},$$

where p is the dimension of $\boldsymbol{\theta}$.

Remarks Let $\hat{\boldsymbol{\theta}}$ be the MLE of the p-dimensional parameter $\boldsymbol{\theta}$ based on the data set from m iid WCRPs. If w is held fixed and $m \to \infty$, then $\sqrt{m}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \stackrel{d}{\to} N_p(\mathbf{0}, [\mathbf{I}(\boldsymbol{\theta}; \mathcal{D})]^{-1})$. On the other hand, if m remains finite and $w \to \infty$, or if both approach infinity, $\sqrt{mw}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \stackrel{d}{\to} N_p(\mathbf{0}, [\mathbf{I}(\boldsymbol{\theta}; X)]^{-1})$. Here N_p denotes a p-variate normal distribution. Hence the FI matrix or an approximation to it can be used to construct confidence intervals and tests of hypotheses based on the large-sample properties of the MLEs. For example, when the renewal distribution Expo(β), the FI from these m WCRPs is $\frac{mw}{\beta^3}$ (see Lemma 5 to follow). When mw is large, the MLE $\hat{\beta}$ from such a data set is approximately $N_1(\beta, \beta^3/mw)$.

When the renewal distribution is either gamma, Weibull, or lognormal with two parameters, the FI matrix is approximated by simulation with large m and thus the asymptotic distribution of the MLE is approximated by a bivariate normal distribution as follows:

$$\hat{\boldsymbol{\theta}} \dot{\sim} N_2(\boldsymbol{\theta}, \hat{\Sigma}), \quad \hat{\Sigma} = \frac{1}{m}[\hat{\mathbf{I}}(\boldsymbol{\theta}; \mathcal{D})]^{-1}.$$

2.4 Approximating the Fisher information

When the FI matrix $\mathbf{I}(\boldsymbol{\theta}; \mathcal{D})$ does not have a closed-form expression, assuming $\boldsymbol{\theta}$ is known, we can simulate m iid WCRPs and compute the average of the observed FI matrices, defined as the product of the first derivatives of the log-likelihood function. This can serve as an estimate of the FI matrix. We now describe the simulation process.

2.4.1 Simulation of a renewal process

To simulate a WCRP, we need to first simulate a Y with pdf in (1). This pdf can be viewed as (Cox 1962)

$$g(y) = \frac{1 - F(y)}{\mu_X} = \int_y^\infty \frac{1}{v} \frac{vf(v)}{\mu_X} \mathrm{d}v, \tag{8}$$

and hence Y can be simulated using the inherent length-biased sampling process represented by the pdf $vf(v)/\mu_X$ and a random contraction created by a Uniform(0, 1) random variable. Thus, the FRT random variable Y can be generated using the following steps:

1. Generate a random variable V with pdf $h(v) = vf(v)/\mu_X$.
2. Generate independently a U from the Uniform(0, 1) distribution.
3. Set $Y = UV$.

Sometimes V and X belong to the same family and thus the simulation processes will be similar. For example, when X is Gamma(α, β), V is a Gamma($\alpha + 1, \beta$) random variable.

The steady-state WCRP with window length w can now be simulated using the intrinsic properties of the RP as described in the following algorithm:

1. Simulate Y using the process described above. If $Y \geq w$, we stop and set $\mathcal{D} = (N = 0, w)$. If $Y < w$, we save Y and go to step 2.
2. Generate X_1 from the renewal pdf $f(x)$. If $Y + X_1 \geq w$, we set $\mathcal{D} = (Y, Z = w - X_1)$. If $Y + X_1 < w$, save Y and X_1, and go to step 3.
3. Continue independently generating X_j from $f(x)$ up to X_n such that $Y + \sum_{j=1}^{n-1} X_j \leq w$ and $Y + \sum_{j=1}^{n} X_j > w$. Set $\mathcal{D} = (N = n, Y, \mathbf{X} = (X_1, \ldots, X_{n-1}), Z = w - (Y + \sum_{j=1}^{n-1} X_j))$.

Repeat this cycle for, say, m times.

The (k, l)th entry of the observed FI matrix from the ith WCRP in the simulated data set is

$$\begin{aligned}
&\hat{I}(\theta_k, \theta_l; \mathcal{D}_i) \\
&= \left[\frac{\partial}{\partial\theta_k}\ell(\mathcal{D}_i; \boldsymbol{\theta})\right] \times \left[\frac{\partial}{\partial\theta_l}\ell(\mathcal{D}_i; \boldsymbol{\theta})\right] \\
&= \delta_0 \frac{\frac{\partial}{\partial\theta_k}G(w)\frac{\partial}{\partial\theta_l}G(w)}{(1-G(w))^2} + \delta_1\left(\frac{\frac{\partial}{\partial\theta_k}g(y_i)}{g(y_i)} - \frac{\frac{\partial}{\partial\theta_k}F(z_i)}{1-F(z_i)}\right)\left(\frac{\frac{\partial}{\partial\theta_l}g(y_i)}{g(y_i)} - \frac{\frac{\partial}{\partial\theta_l}F(z_i)}{1-F(z_i)}\right) \\
&\quad + \sum_{n=2}^{\infty}\delta_n\left[\left(\frac{\frac{\partial}{\partial\theta_k}g(y_i)}{g(y_i)} + \left(\sum_{j=1}^{n-1}\frac{\frac{\partial}{\partial\theta_k}f(x_{ij})}{f(x_{ij})}\right) - \frac{\frac{\partial}{\partial\theta_k}F(z_i)}{1-F(z_i)}\right)\right. \\
&\quad \left.\times\left(\frac{\frac{\partial}{\partial\theta_l}g(y_i)}{g(y_i)} + \left(\sum_{j=1}^{n-1}\frac{\frac{\partial}{\partial\theta_l}f(x_{ij})}{f(x_{ij})}\right) - \frac{\frac{\partial}{\partial\theta_l}F(z_i)}{1-F(z_i)}\right)\right]. \qquad (9)
\end{aligned}$$

The approximated FI matrix from one WCRP, denoted by $\hat{\mathbf{I}}(\boldsymbol{\theta}; \mathcal{D}) = [\hat{I}(\theta_k, \theta_l; \mathcal{D})]$ is

$$\hat{\mathbf{I}}(\boldsymbol{\theta}; \mathcal{D}) = \frac{1}{m}\sum_{i=1}^{m}\hat{\mathbf{I}}(\boldsymbol{\theta}; \mathcal{D}_i), \qquad (10)$$

where $\hat{\mathbf{I}}(\boldsymbol{\theta}; \mathcal{D}_i)$ is the observed FI matrix for the ith window. Clearly, this estimate of the FI matrix is unbiased and is consistent as $m \to \infty$.

If the true parameter values are unknown, we replace $\boldsymbol{\theta}$ with the MLE $\hat{\boldsymbol{\theta}}$ in (9) and (10). The new approximation to the FI matrix is denoted by $\hat{\mathbf{I}}(\hat{\boldsymbol{\theta}}; \mathcal{D}) = [\hat{I}(\hat{\theta}_k, \hat{\theta}_l; \mathcal{D})]$. It is a consistent estimator whenever $I(\theta_k, \theta_l; \mathcal{D})$ is continuous in (θ_k, θ_l), since $\hat{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}$.

3 WCRP with exponential renewal distribution

3.1 Likelihood function

If X is distributed as Expo(β), its pdf is given by

$$f(x; \beta) = \frac{1}{\beta}\mathrm{e}^{-x/\beta}, \quad x > 0, \qquad (11)$$

and, in view of (2), the corresponding FRT variable Y is also Expo(β). Plugging these functions into (3) in Theorem 1 we obtain

$$L_0(\mathcal{D}; \beta) = \mathrm{e}^{-w/\beta}, \quad L_n(\mathcal{D}; \beta) = \left(\frac{1}{\beta}\right)^n \mathrm{e}^{-w/\beta} \quad \text{for } n \geq 1.$$

Thus the likelihood function of a data set from one WCRP is

$$L_n = \beta^{-n} \mathrm{e}^{-w/\beta}, \quad n \geq 0, \tag{12}$$

an expression that depends only on n, the number of renewals in the window. This is not surprising given that the exponential distribution has the memoryless property. The marginal distribution of the number of renewals N in a window of length w is obtained by integrating out the Y and the X's from the likelihood function over their supports. Thus,

$$\begin{aligned}
P(N=0) &= 1 - G(w) = \mathrm{e}^{-w/\beta}, \\
P(N=1) &= \int_0^w \frac{1}{\beta}\mathrm{e}^{-w/\beta}\mathrm{d}y = \frac{w}{\beta}\mathrm{e}^{-w/\beta}, \\
P(N=n) &= \int_{y=0}^{w}\int_{x_1=0}^{w-y}\cdots\int_{x_{n-1}=0}^{w-y-\sum_{j=1}^{n-2}x_j}\left(\frac{1}{\beta}\right)^n \mathrm{e}^{-w/\beta}\mathrm{d}x_{n-1}\cdots\mathrm{d}x_1\mathrm{d}y \\
&= \frac{(\frac{w}{\beta})^n \mathrm{e}^{-w/\beta}}{n!}, \quad \text{for } n = 2, 3, \ldots,
\end{aligned}$$

or $N \sim \text{Poisson}(\frac{w}{\beta})$ with expected value $E(N) = \frac{w}{\beta}$. This is well-known; using different arguments, Cox (1962) observed that the number of renewals is Poisson in a WCRP with an exponential renewal distribution.

3.2 Fisher information

Lemma 5 *Consider a RP with renewal distribution Expo(β). The FI on the parameter β in the data set from a WCRP with window length w is*

$$I(\beta; \mathcal{D}) = \frac{w}{\beta^3}. \tag{13}$$

One can use Theorem 2, but a direct proof is easier in view of the likelihood given in (12). Note that

$$\frac{\partial \log L_n}{\partial \beta} = -\frac{1}{\beta}\left(n - \frac{w}{\beta}\right),$$

and since $N \sim \text{Poisson}(\frac{w}{\beta})$,

$$I(\beta; \mathcal{D}) = \frac{1}{\beta^2}\text{Var}(N) = \frac{w}{\beta^3}.$$

Thus the FI on β is a multiple of the window length w. Since the FI on β from $X \sim \text{Expo}(\beta)$ is $I(\beta; X) = \frac{1}{\beta^2}$, it follows that

$$I(\beta; \mathcal{D}) = \frac{w}{\beta^3} = w\frac{I(\beta; X)}{\mu_X} = E(N)I(\beta; X). \tag{14}$$

These relationships are due to the special properties of the exponential distribution.

4 Gamma renewal distribution

4.1 Likelihood function

If X is distributed as Gamma(α, β), it has pdf

$$f(x; \alpha, \beta) = \frac{\mathrm{e}^{-x/\beta}x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}, \quad x \geq 0, \quad \alpha, \beta > 0. \tag{15}$$

The cdf of X does not have a closed form in general, and is given by

$$F(x; \alpha, \beta) = \int_0^x \frac{\mathrm{e}^{-t/\beta}t^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}\mathrm{d}t. \tag{16}$$

The pdf and cdf of the associated FRT random variable Y, respectively, are

$$g(y; \alpha, \beta) = \frac{1 - F(y; \alpha, \beta)}{\alpha\beta} = \frac{1 - \int_0^y \frac{\mathrm{e}^{-x/\beta}x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}\mathrm{d}x}{\alpha\beta}, \text{ and} \tag{17}$$

$$\begin{aligned} G(y; \alpha, \beta) &= \int_0^y g(t; \alpha, \beta)\mathrm{d}t \\ &= \int_0^y \frac{1 - \int_0^t \frac{\mathrm{e}^{-x/\beta}x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}\mathrm{d}x}{\alpha\beta}\mathrm{d}t. \end{aligned} \tag{18}$$

The likelihood function for the data set from a WCRP with this renewal distribution is obtained by plugging (15)–(18) into (3) in Theorem 1. With $\boldsymbol{\theta} = (\alpha, \beta)$,

$$\begin{aligned} L(D; \boldsymbol{\theta}) &= \delta_0 L_0(D; \boldsymbol{\theta}) + \delta_1 L_1(D; \boldsymbol{\theta}) + \sum_{n=2}^{\infty} \delta_n L_n(D; \boldsymbol{\theta}) \\ &= \delta_0 \left[1 - \int_0^w \frac{1 - \int_0^t f(x)\mathrm{d}x}{\alpha\beta}\mathrm{d}t\right] + \delta_1 \left[\frac{1 - \int_0^y f(x)\mathrm{d}x}{\alpha\beta}\left(1 - \int_0^z f(t)\mathrm{d}t\right)\right] \\ &\quad + \sum_{n=2}^{\infty} \delta_n \left[\frac{1 - \int_0^y f(x)dx}{\alpha\beta} f(x_1)\cdots f(x_{n-1})\left(1 - \int_0^z f(t)\mathrm{d}t\right)\right] \end{aligned}$$

where f is given by (15).

4.2 Fisher information

Clearly, the FI matrix from a WCRP with a Gamma(α, β) renewal distribution does not have a closed-form expression. As described in Sect. 2.4, the FI matrix can be approximated using simulations and (9) and (10). We need the following expressions to estimate the FI on α from a WCRP using $\hat{I}(\alpha, \alpha; \mathcal{D}_i)$, given in (9).

$$\begin{aligned}
\frac{\partial}{\partial\alpha}\log(1-G(w)) &= \frac{\int_w^\infty \int_0^y f(x)\left[\frac{\frac{\partial}{\partial\alpha}\Gamma(\alpha)}{\Gamma(\alpha)} + \log\beta - \log x\right]\mathrm{d}x\,\mathrm{d}y}{\int_w^\infty [1-F(y)]\,\mathrm{d}y} - \frac{1}{\alpha},\\
\frac{\partial}{\partial\alpha}\log f(x_{ij}) &= -\frac{\frac{\partial}{\partial\alpha}\Gamma(\alpha)}{\Gamma(\alpha)} - \log\beta + \log x_{ij},\\
\frac{\partial}{\partial\alpha}\log g(y_i) &= \frac{-\int_0^{y_i} f(x)\log x\,\mathrm{d}x + \left[\frac{\frac{\partial}{\partial\alpha}\Gamma(\alpha)}{\Gamma(\alpha)} + \log\beta\right]F(y_i)}{1-F(y_i)} - \frac{1}{\alpha},\\
\frac{\partial}{\partial\alpha}\log(1-F(z_i)) &= \frac{-\int_0^{z_i} f(x)\log x\,\mathrm{d}x + \left[\frac{\frac{\partial}{\partial\alpha}\Gamma(\alpha)}{\Gamma(\alpha)} + \log\beta\right]F(z_i)}{1-F(z_i)}.
\end{aligned} \tag{19}$$

Similarly, the following expressions are needed for estimating the FI in β from the ith WCRP by $\hat{I}(\beta, \beta; D_i)$.

$$\begin{aligned}
\frac{\partial}{\partial\beta}\log(1-G(w)) &= \frac{\frac{\alpha}{\beta}\int_w^\infty [F(y) - F(y;\alpha+1,\beta)]\mathrm{d}y}{\int_w^\infty [1-F(y)]\mathrm{d}y} - \frac{1}{\beta},\\
\frac{\partial}{\partial\beta}\log f(x_{ij}) &= -\frac{\alpha}{\beta} + \frac{x_{ij}}{\beta^2},\\
\frac{\partial}{\partial\beta}\log g(y_i) &= \frac{\frac{\alpha}{\beta}[F(y_i) - F(y_i;\alpha+1,\beta)]}{1-F(y_i)} - \frac{1}{\beta},\\
\frac{\partial}{\partial\beta}\log(1-F(z_i)) &= \frac{\frac{\alpha}{\beta}[F(z_i) - F(z_i;\alpha+1,\beta)]}{1-F(z_i)}.
\end{aligned}$$

In the above expressions $F(x; \alpha+1, \beta)$ denotes the cdf of a Gamma($\alpha+1, \beta$) variable.

In order to demonstrate the estimation of FI, we simulate $\hat{\mathbf{I}}(\boldsymbol{\theta}; D)$ with $\boldsymbol{\theta} = (\alpha, \beta)$ using the statistical software R. The integrations in the expressions above are calculated using the 'integrate' function in R. Due to the instability of $\log x.f(x)$, R misrepresents the inner integral of the double integral on the right side of (19) as divergent for some parameter values. The following transformation is made to evaluate the inner integral defined on a bounded region $(a, b) = (0, y)$ to make the double integrals converge. We take $r = 0.9$ and $x = t^{\frac{1}{1-r}} + a$ and note that for any function $B(\cdot)$,

$$\int_a^b B(x)\mathrm{d}x = \frac{1}{1-r}\int_0^{(b-a)^{1-r}} t^{\frac{r}{1-r}} B\left(t^{\frac{1}{1-r}} + a\right)\mathrm{d}t.$$

In our simulation, we chose $\alpha = 0.5, 0.8, 1, 2, 5$, and $\beta = 0.5, 1, 5$. The renewal distribution has mean $\mu_X = \alpha\beta$ and standard deviation $\sigma_X = \sqrt{\alpha\beta^2}$. For each (α, β) value, five window lengths $0.5\mu_X$, μ_X, $\mu_X + \sigma_X$, $\mu_X + 2\sigma_X$ and $\mu_X + 3\sigma_X$ are chosen. Note that a change in the value of α or β affects the mean and standard deviation of the renewal distribution and hence changes the window length w. For each value of (α, β) and each window length picked, m iid WCRPs each with a renewal distribution Gamma(α, β) are simulated and the approximated FI is calculated using (10).

We plot the curves of the FI in α, β and the determinant of the FI matrix in Figs. 2, 3 and 4. Each figure has a fixed β and the curves are drawn by connecting the points formed by pairs of window length and FI for each of the five choices of α. In each figure, plot (a) presents $\hat{I}(\alpha, \alpha; \mathcal{D})$ versus w, (b) presents $\hat{I}(\beta, \beta; \mathcal{D})$ versus w, and (c) presents the determinant of our approximated FI matrix, $\widehat{\det}[\mathbf{I}(\alpha, \beta; \mathcal{D})]$, versus w. The values of α and associated symbols (given in parentheses) in each subplot are $0.5(\circ)$, $0.8(\triangle)$, $1.0(\times)$, $2.0(\triangledown)$, and $5.0(*)$. For instance, the curve with circles $(\circ)$ on it in plot (a) of Fig. 2 shows the FI in α for a WCRP with renewal distribution Gamma$(\alpha = 0.5, \beta = 0.5)$. The renewal distribution has mean $\mu_X = 0.5 \times 0.5 = 0.25$ and standard deviation $\sigma_X = \sqrt{0.5 \times 0.5^2} = 0.354$. The window lengths at which simulations are run are $0.5 \times 0.25 = 0.625$, 0.25, $0.25 + 0.354 = 0.6036$, $0.25 + 2 \times 0.354 = 0.957$, and $0.25 + 3 \times 0.354 = 1.310$. The other four curves in the same plot are obtained similarly with α equalling 0.8, 1, 2, and 5, respectively. Plots (b) and (c) of Fig. 2 are drawn similarly; only the ordinates are different.

Across different values of β the shapes of the FI curves are nearly the same. The FI curves in plots (a) appear to be linear; in plots (b) they appear to show mildly nonlinear trends within the range of our simulation and the determinant curves in plot (c) show quadratic trends. As expected, both FI and the determinant of the FI matrix increase monotonically as the window length increases. We will refine the examination of these plots in Sect.6.

5 Weibull renewal distribution

Suppose $X \sim$ Weibull(r, β) with pdf

$$f(x; r, \beta) = \frac{r}{\beta} x^{r-1} \mathrm{e}^{-x^r/\beta}, \quad x \geq 0, \quad \beta, r > 0,$$

where r is the shape parameter and β is the scale parameter. The associated FRT variable Y has the pdf

$$g(y; r, \beta) = \frac{\mathrm{e}^{-y^r/\beta}}{\beta^{1/r}\Gamma(1 + \frac{1}{r})},$$

and this random variable can be simulated as the length-biased version of the random variable V distributed as a $\{\text{Gamma}(1 + (1/r), \beta)\}^{1/r}$ random variable (Zhao 2006, p. 15). Now one can follow the algorithm presented in Sect. 2.4.1 to simulate the needed WCRP data.

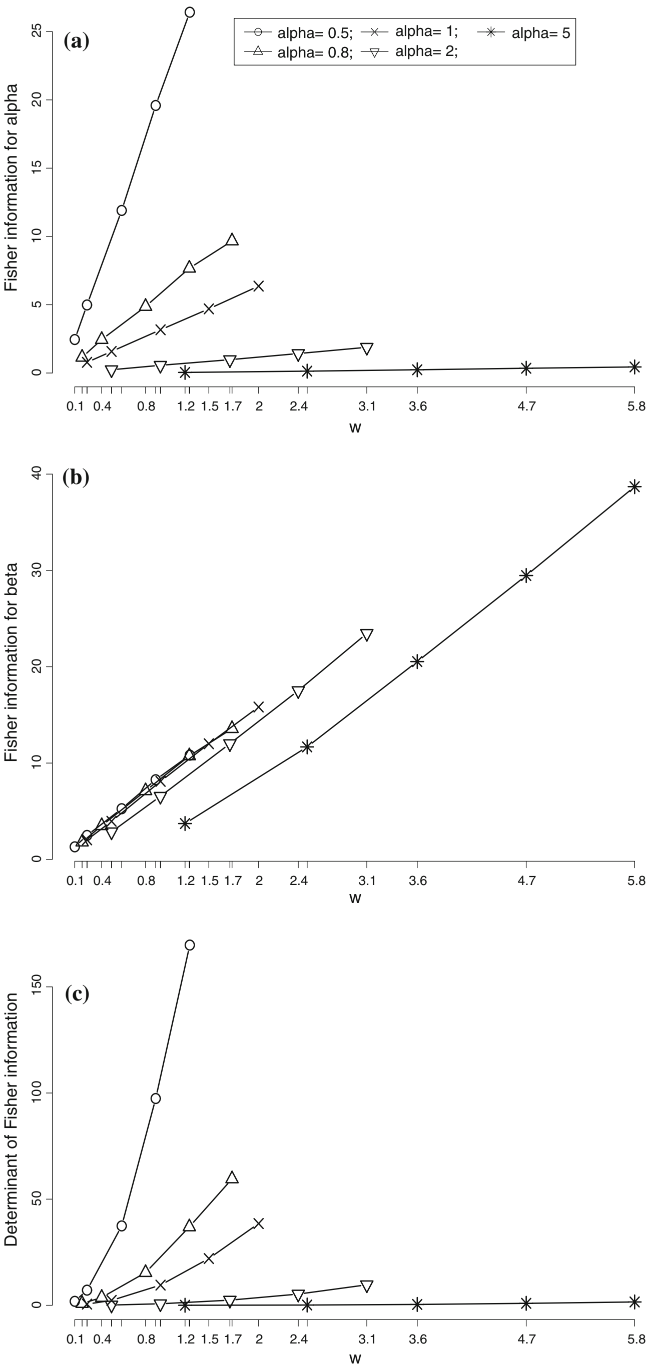

Fig. 2 Estimates of the FI in a WCRP with renewal distribution Gamma(α, β) for $\beta = 0.5$ and selected α and window-length. Figure legend is given in **a**

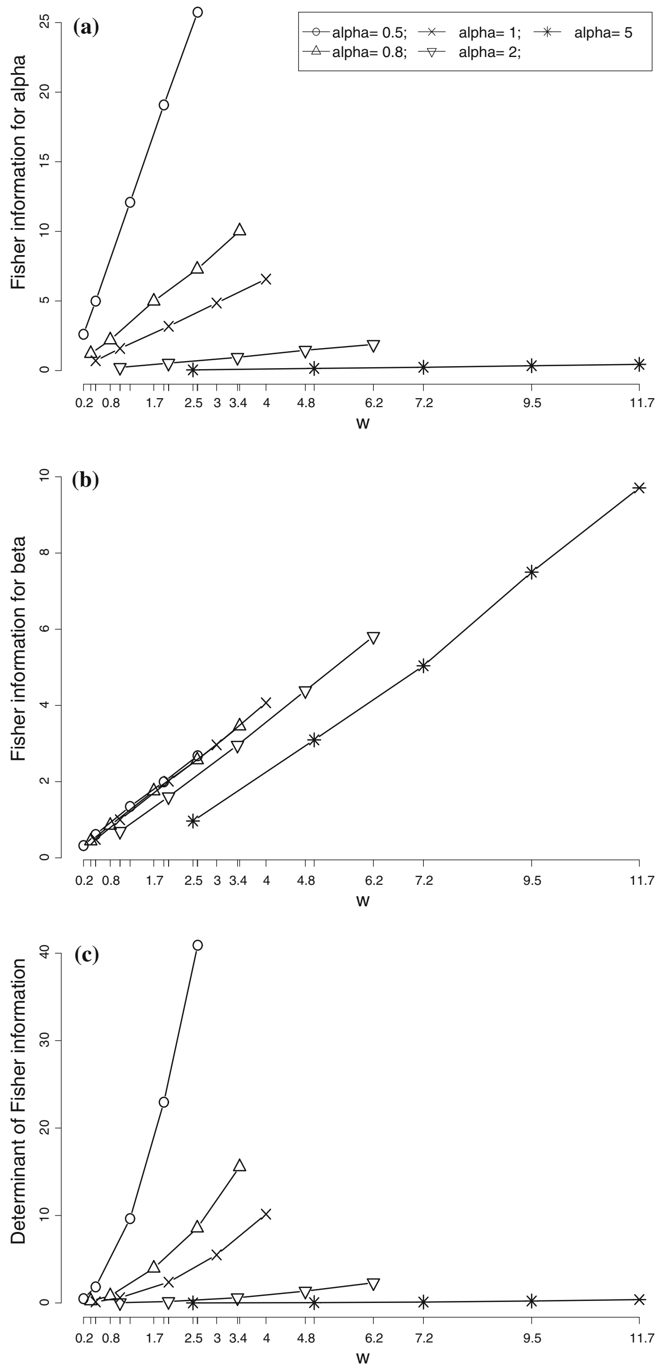

Fig. 3 Estimates of the FI in a WCRP with renewal distribution Gamma(α, β) for $\beta = 1$ and selected α and window-length. Figure legend is given in **a**

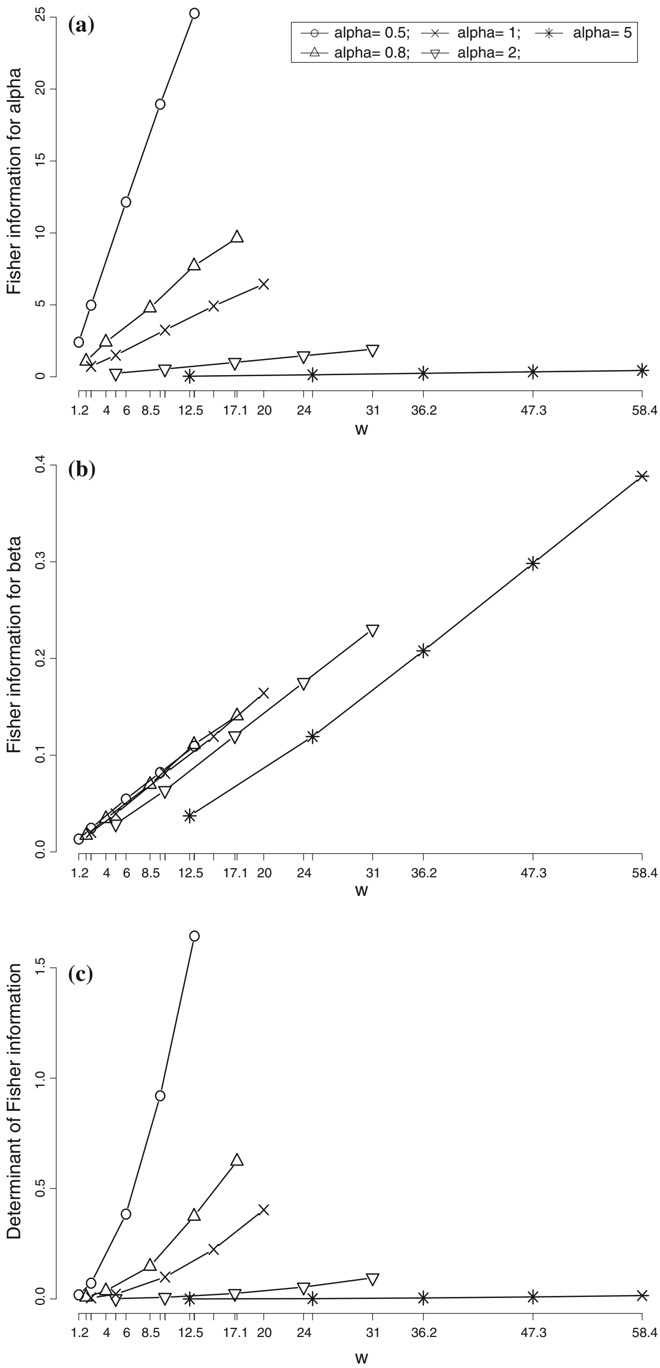

Fig. 4 Estimates of the FI in a WCRP with renewal distribution Gamma(α, β) for $\beta = 5$ and selected α and window-length. Figure legend is given in **a**

While closed form expression is not available for the FI matrix, the partial derivatives of f, F, g, and G with respect to r and β that are needed in (9) are much simpler than for the Gamma parent, and involve only single integrals. In particular, we have

$$-\frac{\frac{\partial}{\partial r}G(w)}{1-G(w)} = \frac{\partial}{\partial r}\log(1-G(w)) = \frac{\int_0^w g(y)\left[\frac{y^r}{\beta}\log y - \frac{\log\beta}{r^2} - \frac{\Gamma'(1+\frac{1}{r})}{r^2\Gamma(1+\frac{1}{r})}\right]\mathrm{d}y}{1-G(w)},$$
$$\frac{\frac{\partial}{\partial r}f(x_{ij})}{f(x_{ij})} = \frac{\partial}{\partial r}\log f(x_{ij}) = \frac{1}{r} + \log x_{ij} - \frac{x_{ij}^r}{\beta}\log x_{ij},$$
$$\frac{\frac{\partial}{\partial r}g(y_i)}{g(y_i)} = \frac{\partial}{\partial r}\log g(y_i) = -\frac{y_i^r}{\beta}\log y_i + \frac{\log\beta}{r^2} + \frac{\frac{\partial}{\partial r}\Gamma(1+\frac{1}{r})}{r^2\Gamma(1+\frac{1}{r})},$$
$$-\frac{\frac{\partial}{\partial r}F(z_i)}{1-F(z_i)} = \frac{\partial}{\partial r}\log(1-F(z_i)) = -\frac{z_i^r}{\beta}\log z_i,$$

and

$$-\frac{\frac{\partial}{\partial\beta}G(w)}{1-G(w)} = \frac{\partial}{\partial\beta}\log(1-G(w)) = \frac{\int_0^w g(y)\left(\frac{1}{\beta r} - \frac{y^r}{\beta^2}\right)\mathrm{d}y}{1-G(w)},$$
$$\frac{\frac{\partial}{\partial\beta}f(x_{ij})}{f(x_{ij})} = \frac{\partial}{\partial\beta}\log f(x_{ij}) = -\frac{1}{\beta} + \frac{x_{ij}^r}{\beta^2},$$
$$\frac{\frac{\partial}{\partial\beta}g(y_i)}{g(y_i)} = \frac{\partial}{\partial\beta}\log g(y_i) = -\frac{1}{r\beta} + \frac{y_i^r}{\beta^2},$$
$$-\frac{\frac{\partial}{\partial\beta}F(z_i)}{1-F(z_i)} = \frac{\partial}{\partial\beta}\log(1-F(z_i)) = \frac{z_i^r}{\beta^2}.$$

We simulated $\hat{\mathbf{I}}(\boldsymbol{\theta}; D)$ with $\boldsymbol{\theta} = (r, \beta)$. The simulations used $r = 0.5, 0.8, 1.0, 1.2, 5$ and $\beta = 0.5, 1, 5$ and w values represented by $0.5\mu_X$, μ_X, $\mu_X + \sigma_X$, $\mu_X + 2\sigma_X$ and $\mu_X + 3\sigma_X$. For each set of (r, β, w) value, approximated FI is calculated using (9), (10) and m iid simulated WCRPs. Using the conventions employed for the gamma RP, we plotted three subplots (with the parameter r taking the role of the α of the gamma distribution): (a) presenting $\hat{I}(r, r; D)$ versus w, (b) showing $\hat{I}(\beta, \beta; D)$ versus w, and (c) presenting the determinant of our approximation to the FI matrix, $\widehat{\det}[\mathbf{I}(r, \beta; D)]$, versus w. The plots are not shown here for economy (see Zhao 2006, Figures 3.4–3.6). The FI curves increase monotonically with linear or quadratic trends and the determinant of FI matrix increases quadratically or exponentially. We will show one standardized FI plot later in Sect. 6.3.

6 Designing experiments with WCRP data

Let us consider a study that plans to observe m iid RPs using windows of length w. The total time spent observing the RPs is the total window lengths, $m \times w$. Let us fix the total time at T based on, say, cost considerations. We want to choose the values of m and w such that $m \times w = T$ under certain optimality condition based on the FI contained in the experiment.

When θ is a scalar parameter, we search for the m and w combination that maximizes the FI. Since the variance of the MLE is approximated by the reciprocal of the FI, this criterion can be viewed as attaining the best precision for the MLE given the total duration of the experiment. For a vector parameter $\boldsymbol{\theta}$ we could consider one of the following:

1. Maximize $\frac{\mathbf{I}(\theta_k,\theta_k;\mathcal{D})}{w}$ for each θ_k.
2. Maximize $\frac{\det[\mathbf{I}(\boldsymbol{\theta};\mathcal{D})]}{w^p}$, where $\boldsymbol{\theta}$ is p-dimensional.

Under criterion 1, different combinations of w and m may turn out to be optimal for different k. Now suppose the renewal distribution has two parameters (as in the gamma case). The determinant of the FI matrix from m WCRPs is

$$m^2 \times \det[\mathbf{I}(\boldsymbol{\theta};\mathcal{D})] = \left(\frac{T}{w}\right)^2 \times \det[\mathbf{I}(\boldsymbol{\theta};\mathcal{D})].$$

As the determinant of the FI matrix increases, the determinant of the covariance matrix of the MLEs, also known as the generalized variance (Anderson 1984, p. 259), decreases. Thus, we want to maximize $(\frac{T}{w})^2 \times \det[\mathbf{I}(\boldsymbol{\theta};\mathcal{D})]$. Since T is a constant, we can maximize $\det[\mathbf{I}(\boldsymbol{\theta};\mathcal{D})]/w^2$ as suggested in the optimality criterion 2.

We will now look for the window lengths that are optimal according to one of these criteria for the exponential, gamma, and Weibull renewal distributions. In the multi-parameter case one could also consider maximizing the trace of the FI matrix.

If the objective is to control cost that is a more general function of m and w, the optimal scheme may be different. See the discussion below.

6.1 The exponential distribution

If the renewal distribution is Expo(β), from (14) it follows that the FI from m iid WCRPs with common window length w and total duration T is $I(\beta;\mathbf{D}) = T\beta^{-3}$. When T fixed, the FI from the data set is not affected by the choice of window length and the optimality criterion 1 simplifies to the constant $1/\beta^3$. Thus for a fixed total duration of the experiment, all combinations of window length and number of subjects have the same amount of FI collected. Then the choice will be made on the basis of the number of available subjects and/or limitations on the maximum window length permitted due to time considerations.

The optimal scheme may be different for other objective functions. Now suppose c (≥ 0) is the cost of recruiting a subject and let $\eta(w)$ be the cost associated with the observation time w per subject. Without loss of generality we assume $\eta(0) = 0$ and $\eta(w)$ is strictly increasing. With the recruitment of m subjects where each is observed for w time units, the total cost would be $C = cm + m\eta(w)$. If this cost is held fixed, $m = C/(c + \eta(w))$, and the FI in the experiment is

$$I(\beta; \mathbf{D}) = m\frac{w}{\beta^3} = \frac{Cw}{(c + \eta(w))\beta^3}.$$

Thus the optimal window length w_0 corresponds to the maximum value of $w/(c + \eta(w))$. If $\eta(w)$ is continuously differentiable and if the solution to $w\eta'(w) - \eta(w) = c$ is unique, then one can determine the optimal design subject to the constraint that m is an integer.

If $c = 0$ and $\eta(w) = w$, we obtain the first scenario we discussed with $C = T$ and all combinations of integer m and w with $mw = T$ provide the same amount of FI. If $c > 0$, while the waiting time cost is still linear, the optimal strategy uses only one subject monitored for T time units. On the other hand if $c > 0$ and $\eta(w) = e^w$ indicating that the cost of monitoring a subject increases exponentially, the optimal design has the unique window length $w_0 > 1$ that is the solution of the equation $w \exp(w) - \exp(w) = c$.

6.2 Gamma renewal distribution

For the Gamma(α, β) renewal distribution, using simulation we estimated the FI from one WCRP in Sect. 4. We obtained $\hat{I}(\alpha, \alpha; \mathcal{D})$, $\hat{I}(\beta, \beta; \mathcal{D})$ and the determinant of the FI matrix, $\det(\hat{\mathbf{I}}(\boldsymbol{\theta}; \mathcal{D}))$ for different (α, β) and w values. Based on these, we can calculate, and plot the following standardized FI measures:

$$\frac{\hat{I}(\alpha, \alpha; \mathcal{D})/w}{I(\alpha; X)/\mu_X}, \quad \frac{\hat{I}(\beta, \beta; \mathcal{D})/w}{I(\beta; X)/\mu_X}, \quad \frac{\det(\hat{\mathbf{I}}(\boldsymbol{\theta}; \mathcal{D}))/w^2}{\det(\mathbf{I}(\boldsymbol{\theta}; X))/\mu_X^2}.$$

It follows from Theorem 4 that these ratios approach 1 in probability as the window length w increases.

These simulation results are plotted in Figs. 5, 6 and 7 using conventions and symbols adopted in Sect. 4.2 and the same combination of (α, β, w) values. As done there, we put three plots in each of these figures and each figure corresponds to a specific value of β. Plot (a) displays $\frac{\hat{I}(\alpha,\alpha;\mathcal{D})/w}{I(\alpha;X)/\mu_X}$ versus w, plot (b) displays $\frac{\hat{I}(\beta,\beta;\mathcal{D})/w}{I(\beta;X)/\mu_X}$ versus w, and plot (c) displays $\frac{\det(\hat{\mathbf{I}}(\boldsymbol{\theta};\mathcal{D}))/w^2}{\det(\mathbf{I}(\boldsymbol{\theta};X))/\mu_X^2}$ versus w. The plots in Figs. 5, 6 and 7 also contain horizontal band bounded by $\pm 10\%$ of the limiting FI. A quick look at the subplots (a) and (b) indicate that the FI on α and β are nonlinear functions of the window length w and the determinant is not a function of just w^2. Subplots (a) and (c) show that the standardized FI can be non-monotonic. Suppose a researcher has some preliminary estimates of α and β which could be based on the estimates of the mean and standard deviations of the renewal distribution. Using these preliminary estimates and one of the above criteria, we want to decide an optimal window length given a fixed total time of the experiment. We have in Figs. 5, 6 and 7 the measures of relevant information we can use to decide on the window length for all combinations

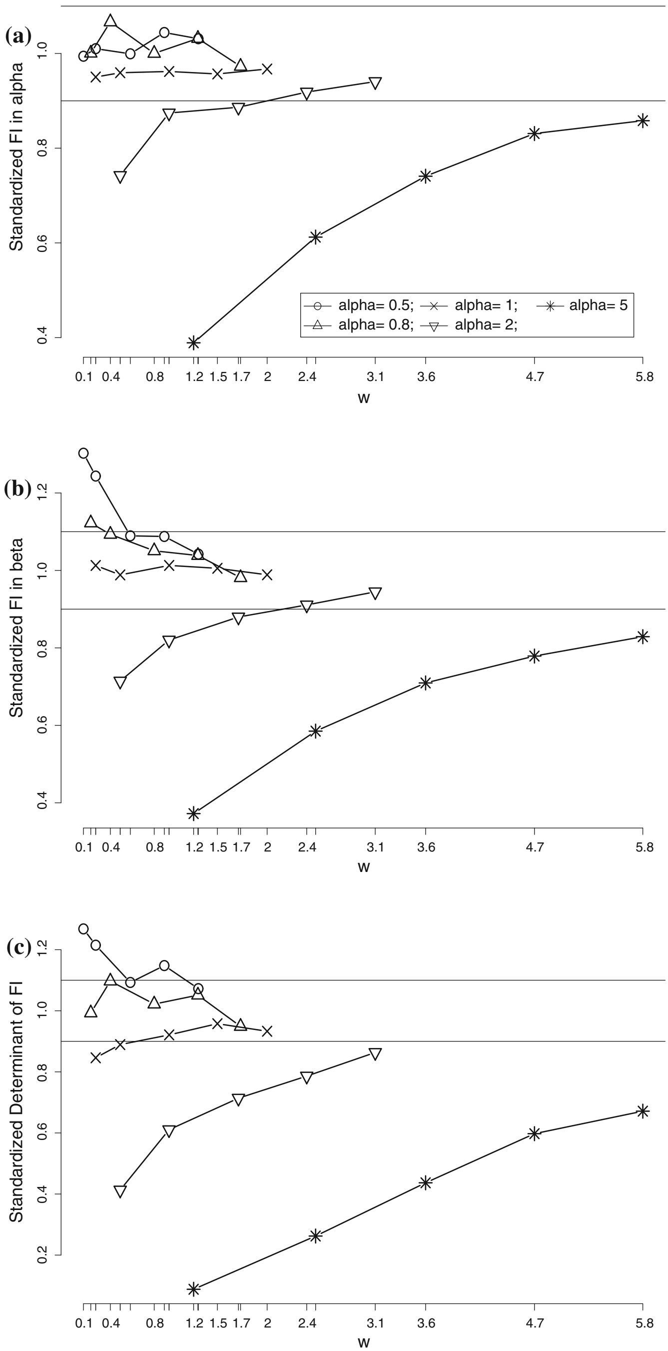

Fig. 5 Standardized FI curves for a WCRP with a gamma renewal distribution with $\beta = 0.5$ and selected values of α and window length w; the *horizontal lines* represent 10% boundaries around the limiting FI or the determinant. Figure legend is given in **a**

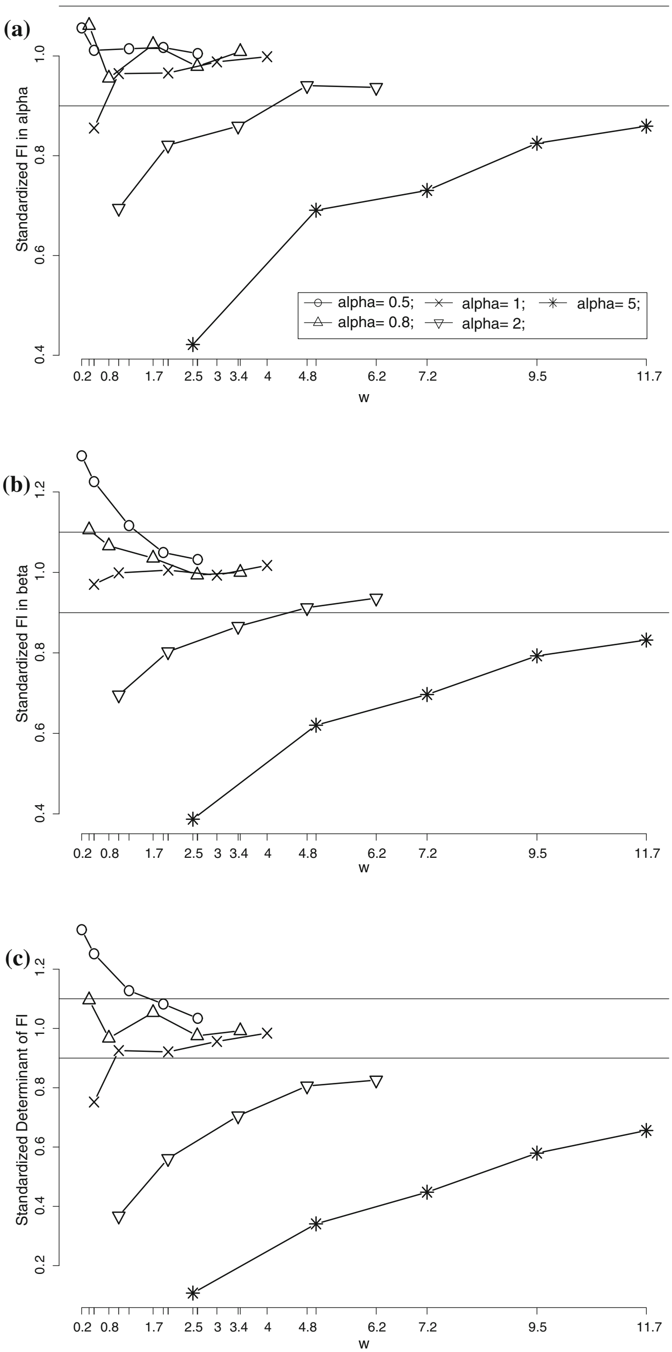

Fig. 6 Standardized FI curves for a WCRP with a gamma renewal distribution with $\beta = 1$ and selected values of α and window length w; the *horizontal lines* represent 10% boundaries around the limiting FI or the determinant. Figure legend is given in **a**

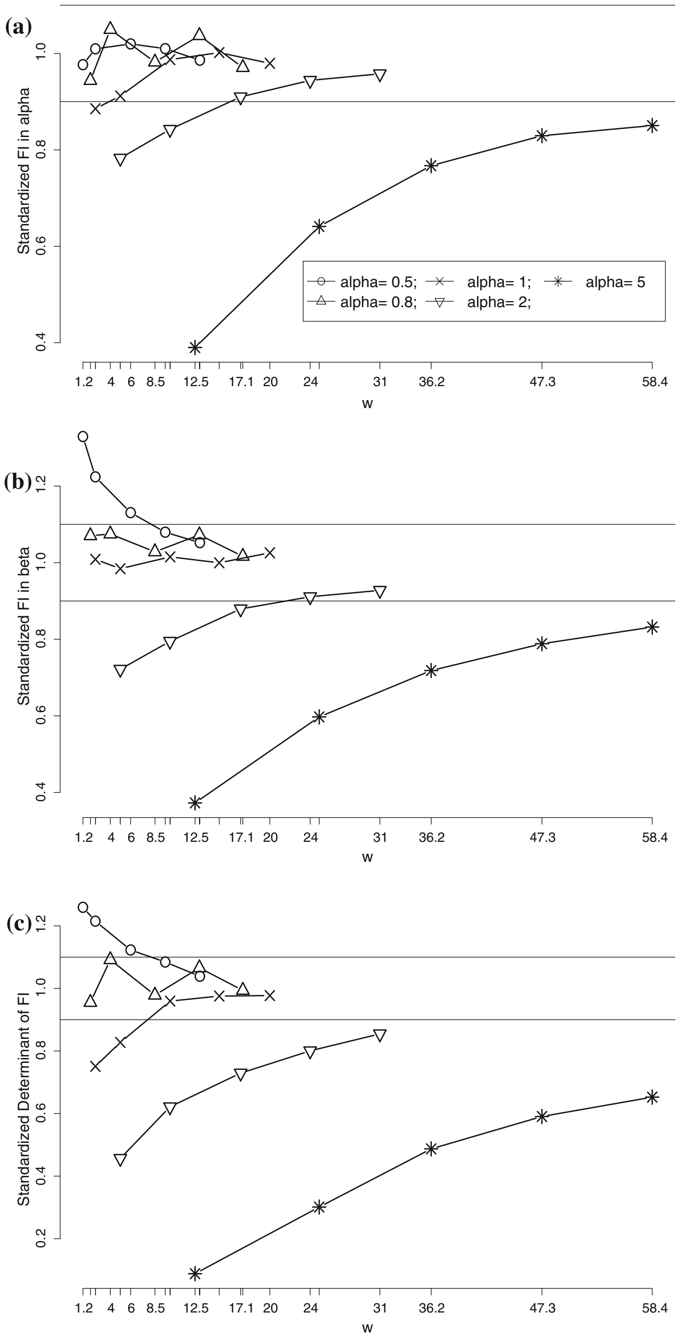

Fig. 7 Standardized FI curves for a WCRP with a gamma renewal distribution with $\beta = 5$ and selected values of α and window length w; the *horizontal lines* represent 10% boundaries around the limiting FI or the determinant. Figure legend is given in **a**

of $\alpha = (0.5, 0.8, 1, 2, 5)$ and $\beta = (0.5, 1, 5)$. For each value of (α, β), we look at the three corresponding curves on the subplots. For instance, when $\alpha = 0.5$ and $\beta = 0.5$, we look at Fig. 5 for the curves with circle ($\circ$) symbols on them in each of the three plots. From subplot (a), the standardized FI on α varies within the 0.9–1.1 range. The curve in plot (b) representing the standardized FI about β decreases as window length increases, suggesting that a smaller window length is preferred (in contrast, if $\alpha > 1$, longer window is preferred for inference on β). And the curve cross the 1.1 line at $\mu_X + \sigma_X = 0.25 + \sqrt{0.125} = 0.60$. The curve in plot (c) representing the standardized measure based on the determinant of the FI first decreases and then increases. The curve attains a local maximum at $w = \mu_X + 2 \times \sigma_X = 0.25 + 2 \times 0.354 = 0.958$. Based on all three plots, we recommend a window length between 0.60 and 0.958 (that is between $\mu_X + \sigma_X$ and $\mu_X + 2\sigma_X$). Generally the recommendation depends on the parameter or parameters of interest and researcher's preliminary estimates.

Figures 5, 6 and 7 show that some of the standardized information curves are monotonic. In such cases, we need to consider practical issues. Even if the curves show a decreasing trend, we should not consider extremely short window lengths. It would mean more subjects on test and the assumption of iid RPs across a large number of subjects or processes would be difficult to defend. On the other hand, if the curves show an increasing trend, we should not be increasing the window length in an unlimited fashion. This is because, even for a single process, the assumption of stationarity implied by the RP model may be untenable for a long duration under observation (especially in disease relapse models). Thus, in case a curve is monotonic, we recommend the window length at which the functions under consideration first leaves or first enters the region (0.9, 1.1), or in another words, the FI achieved is within 10% of the limiting standardized FI. For instance, when $\alpha = 2$ and $\beta = 0.5$, in view of plots (a) and (b) of Fig. 5, a window length between $\mu_X + \sigma_X$ and $\mu_X + 3\sigma_X$ is a good choice.

There are situations where the standardized curves are non-monotonic. For example, when β is away from 1 and $\alpha < 1$, the maximum FI per unit time is attained in plots (a) and (c) when $w = \mu_X + \sigma_X$. So, a window of this length is recommended in such cases.

Finally, these figures also show that when $\alpha = 1$ (which corresponds to the Exponential renewal distribution) the simulated values of the standardized FI on β remains very close to 1 even for very small w values.

6.3 Weibull distribution

For the Weibull(r, β) renewal distribution, as done for the gamma parent, we used simulation and computed

$$\frac{\hat{I}(r, r; D)/w}{I(r; X)/\mu_X}, \quad \frac{\hat{I}(\beta, \beta; D)/w}{I(\beta; X)/\mu_X}, \quad \frac{\det(\hat{\mathbf{I}}(\boldsymbol{\theta}; D))/w^2}{\det(\mathbf{I}(\boldsymbol{\theta}; X))/\mu_X^2}.$$

for $r = 0.5, 0.8, 1.0, 1.2, 1.5$, and $\beta = 0.5, 1, 5.0$ and for selected w. The results are similar to the gamma case, now the parameter r playing the role of α. For illustration, we present the results for $\beta = 1$ in Fig. 8 using the conventions adopted for the

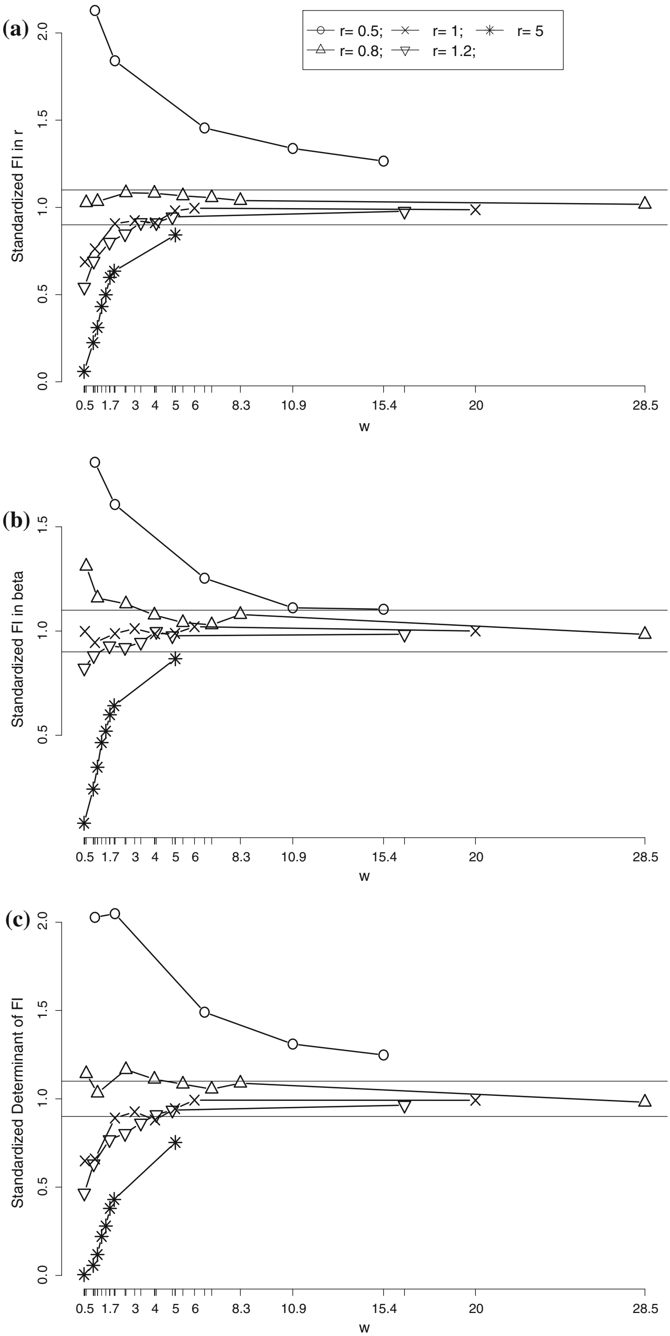

Fig. 8 Standardized FI curves for a WCRP with a Weibull renewal distribution with $\beta = 1$ for selected values of r and window length w; the *horizontal lines* represent 10% boundaries around the limiting FI or the determinant. Figure legend is given in **a**

gamma case. As done there, we put three plots in each of these figures and each figure corresponds to a specific value of β. Subplot (a) plots $\frac{\hat{I}(r,r;D)/w}{I(r;X)/\mu_X}$ versus w, (b) plots $\frac{\hat{I}(\beta,\beta;D)/w}{I(\beta;X)/\mu_X}$ versus w, (c) plots $\frac{\det(\hat{\mathbf{I}}(\boldsymbol{\theta};D))/w^2}{\det(\mathbf{I}(\boldsymbol{\theta};X))/\mu_X^2}$ versus w.

There are five curves on each subplot corresponding to five different values of r. On each curve, there are five window lengths, same as those chosen in Sect. 5, at which we calculate the functions noted on the ordinate axis. From Theorem 4, it follows that the functions plotted in these graphs approach 1 as the window length w increases.

The findings are similar to the gamma case and details are given in Zhao (2006, Sect. 6.3). For both the gamma and Weibull distributions the shape parameters α, r determine the hazard rate properties. In particular, α, r exceeding (being less than) 1 means the renewal distribution has increasing (decreasing) hazard rate. Of course, $\alpha = r = 1$ corresponds to the exponential distribution with a constant hazard rate. From our limited study, it appears that for the scale parameter (β) the standardized FI curves are monotonic when these shape parameters are considerably away from 1, monotonically decreasing (increasing) for low (high) values of the shape parameter. Such a monotonic property holds for the FI about the shape parameter both for the Weibull and Gamma for the large shape parameter values, but only for the Weibull for small shape parameter values. Changes in the values of the scale parameter seem to affect the shapes of the standardized FI curves for shape parameter values under 1.

7 Discussion

There are very few results on Fisher information in WCRPs in the literature and our study provides a comprehensive investigation of this area. If we have some knowledge of the renewal distribution and the parameter values before the start of our study, we can use the FI to determine the optimal observation window length for designing experiments. Simulation study can be done for any parameter values with any renewal distribution to search for an optimal window length.

Our WCRP models are applicable to recurrence data sets that do not start from time zero, a common occurrence in clinical trials involving recurring events. For example, in longitudinal studies of lupus and relapsing remitting multiple sclerosis patients, flare-ups occur periodically and patients have shown these symptoms for a long period of time prior to enrollment in the study. That is, the data set usually starts with an FRT and the work here is useful for parametric inference on such data sets. Using a small data set of this type arising from a longitudinal study of lupus flares, we found that lognormal distribution provides a good fit for the renewal distribution.

To apply our models to recurrence data, certain assumptions need to be met. First we need to check if the distributions of time between events are iid; second we need to check if we have chosen the correct type of renewal distribution. Rigdon and Basu (2000, p. 87) comment on the selection of models: "... show graphs of [the number of renewals] $N(t_i)$ versus t_i. A linear relationship ... indicates that the system remained stable over the time that data were collected. In this case a RP, or possibly a homogeneous Poisson process, may be appropriate model if the times between failure are independent."

In studies with many WCRPs (corresponding to different patients in a clinical trial), homogeneity of the parameter values and fitness of the renewal distribution to the data should be checked. In such settings either a fixed or random effects models can be considered. Validation of and inference under such models are currently under investigation. Some preliminary work was reported in Zhao (2006).

Appendix A

A.1 Proof of Theorem 1

When $N = 0$, there is no renewal inside the observation window, and we have a right-censored FRT of length w. The likelihood function of the data set is

$$L_0(\mathcal{D}; \boldsymbol{\theta}) = P(N = 0) = P(Y > w) = \int_w^\infty g(y)\mathrm{d}y = 1 - G(w).$$

When $N = 1$, we observe one renewal inside the window and the data set contains $Y = y$ and $Z = z = w - y$, where Z is the right censored renewal variable. The likelihood function of the data set is

$$L_1(\mathcal{D}; \boldsymbol{\theta}) = g(y)[1 - F(z)] \quad \text{where } z = w - y;\ y, z > 0.$$

When $N = n \geq 2$, we have at least two renewals inside the window and the data set contains $Y = y$, $X = (x_1, \ldots, x_{n-1})$, and $Z = z$ with the constraint $z = w - y - \sum_{i=1}^{n-1} x_i$. The likelihood function of the data set is

$$L_n(\mathcal{D}; \boldsymbol{\theta}) = g(y) f(x_1) \ldots f(x_{n-1})[1 - F(z)].$$

Let $\{N = n\}$ be the event that n renewals occur inside the observation window. For one WCRP, events $\{N = n\}$ for $n = 0, 1, \ldots$ are mutually exclusive and exhaust all possibilities. The expression represented by the infinite series on the right side of (2) contains a single non-zero term representing the likelihood function corresponding to the event $\{N = n\}$ for a given $\mathcal{D}$.

A.2 Proof of Theorem 2

The proof directly follows from the definition of Fisher information. From the likelihood function given in Theorem 1, the log-likelihood function of the data set from one WCRP can be written as

$$\begin{aligned}\ell(\mathcal{D}; \boldsymbol{\theta}) &= \log L(\mathcal{D}; \boldsymbol{\theta}) \\ &= \delta_0 \log L_0(\mathcal{D}; \boldsymbol{\theta}) + \delta_1 \log L_1(\mathcal{D}; \boldsymbol{\theta}) + \sum_{n=2}^{\infty} \delta_n \log L_n(\mathcal{D}; \boldsymbol{\theta}).\end{aligned}$$

Upon taking the first derivative of the log-likelihood function with respect to the parameter θ_k, we obtain

$$\begin{aligned}\frac{\partial}{\partial\theta_k}\ell(\mathcal{D};\boldsymbol{\theta}) &= \frac{\partial}{\partial\theta_k}\log L(\mathcal{D};\boldsymbol{\theta})\\ &= \delta_0\frac{\partial}{\partial\theta_k}\log L_0(\mathcal{D};\boldsymbol{\theta})+\delta_1\frac{\partial}{\partial\theta_k}\log L_1(\mathcal{D};\boldsymbol{\theta})+\sum_{n=2}^{\infty}\delta_n\frac{\partial}{\partial\theta_k}\log L_n(\mathcal{D};\boldsymbol{\theta}),\end{aligned}$$

where

$$\begin{aligned}\frac{\partial}{\partial\theta_k}\log L_0(\mathcal{D};\boldsymbol{\theta}) &= \frac{\partial}{\partial\theta_k}\log\left[1-G(w)\right] = \frac{-\frac{\partial}{\partial\theta_k}G(w)}{1-G(w)},\\ \frac{\partial}{\partial\theta_k}\log L_1(\mathcal{D};\boldsymbol{\theta}) &= \frac{\partial}{\partial\theta_k}\left[\log g(y)+\log\left(1-F(z)\right)\right] = \frac{\frac{\partial}{\partial\theta_k}g(y)}{g(y)}-\frac{\frac{\partial}{\partial\theta_k}F(z)}{1-F(z)},\\ \frac{\partial}{\partial\theta_k}\log L_n(\mathcal{D};\boldsymbol{\theta}) &= \frac{\partial}{\partial\theta_k}\left[\log g(y)+\left(\sum_{i=1}^{n-1}\log f(x_i)\right)+\log(1-F(z))\right]\\ &= \frac{\frac{\partial}{\partial\theta_k}g(y)}{g(y)}+\left(\sum_{i=1}^{n-1}\frac{\frac{\partial}{\partial\theta_k}f(x_i)}{f(x_i)}\right)-\frac{\frac{\partial}{\partial\theta_k}F(z)}{1-F(z)}.\end{aligned}$$

We know that $\delta_i\delta_j=0$ for $i\neq j$ and $\delta_i\delta_i=\delta_i$ for $i=0,1,\ldots$. Hence the product of $\frac{\partial}{\partial\theta_k}\ell(\mathcal{D};\boldsymbol{\theta})$ and $\frac{\partial}{\partial\theta_l}\ell(\mathcal{D};\boldsymbol{\theta})$ can be expressed as

$$\begin{aligned}&\sum_{n=0}^{\infty}\delta_n\left(\frac{\partial}{\partial\theta_k}\log L_n(\mathcal{D};\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_l}\log L_n(\mathcal{D};\boldsymbol{\theta})\right)\\ &\quad=\delta_0\frac{\frac{\partial}{\partial\theta_k}G(w)\frac{\partial}{\partial\theta_l}G(w)}{(1-G(w))^2}+\delta_1\left(\frac{\frac{\partial}{\partial\theta_k}g(y)}{g(y)}-\frac{\frac{\partial}{\partial\theta_k}F(z)}{1-F(z)}\right)\left(\frac{\frac{\partial}{\partial\theta_l}g(y)}{g(y)}-\frac{\frac{\partial}{\partial\theta_l}F(z)}{1-F(z)}\right)\\ &\qquad+\sum_{n=2}^{\infty}\delta_n\left[\left\{\frac{\frac{\partial}{\partial\theta_k}g(y)}{g(y)}+\left(\sum_{i=1}^{n-1}\frac{\frac{\partial}{\partial\theta_k}f(x_i)}{f(x_i)}\right)-\frac{\frac{\partial}{\partial\theta_k}F(z)}{1-F(z)}\right\}\right.\\ &\qquad\left.\times\left\{\frac{\frac{\partial}{\partial\theta_l}g(y)}{g(y)}+\left(\sum_{i=1}^{n-1}\frac{\frac{\partial}{\partial\theta_l}f(x_i)}{f(x_i)}\right)-\frac{\frac{\partial}{\partial\theta_l}F(z)}{1-F(z)}\right\}\right].\end{aligned}\tag{20}$$

By definition, the (k,l)th entry of the FI matrix $I(\theta_k,\theta_l;\mathcal{D})$ is the expectation $E\left[\frac{\partial}{\partial\theta_k}\ell(\mathcal{D};\boldsymbol{\theta})\frac{\partial}{\partial\theta_l}\ell(\mathcal{D};\boldsymbol{\theta})\right]$ and, in view of (20) is given by

$$E\left\{\delta_0\frac{\frac{\partial}{\partial\theta_k}G(w)\frac{\partial}{\partial\theta_l}G(w)}{(1-G(w))^2}\right\}+E\left\{\delta_1\left(\frac{\frac{\partial}{\partial\theta_k}g(Y)}{g(Y)}-\frac{\frac{\partial}{\partial\theta_k}F(Z)}{1-F(Z)}\right)\left(\frac{\frac{\partial}{\partial\theta_l}g(Y)}{g(Y)}-\frac{\frac{\partial}{\partial\theta_l}F(Z)}{1-F(Z)}\right)\right\}$$

$$
\begin{aligned}
&+\sum_{n=2}^{\infty} E\left\{\delta_n\left[\left(\frac{\frac{\partial}{\partial\theta_k}g(Y)}{g(Y)}+\left(\sum_{i=1}^{n-1}\frac{\frac{\partial}{\partial\theta_k}f(X_i)}{f(X_i)}\right)-\frac{\frac{\partial}{\partial\theta_k}F(Z)}{1-F(Z)}\right)\right.\right.\\
&\left.\left.\times\left(\frac{\frac{\partial}{\partial\theta_l}g(Y)}{g(Y)}+\left(\sum_{i=1}^{n-1}\frac{\frac{\partial}{\partial\theta_l}f(X_i)}{f(X_i)}\right)-\frac{\frac{\partial}{\partial\theta_l}F(Z)}{1-F(Z)}\right)\right]\right\}\\
&=E\left\{\delta_0\frac{\frac{\partial}{\partial\theta_k}G(w)\frac{\partial}{\partial\theta_l}G(w)}{(1-G(w))^2}\right\}+E\left\{\delta_1A_1(\theta_k,\theta_l;\mathcal{D})\right\}+\sum_{n=2}^{\infty}E\left\{\delta_nA_n(\theta_k,\theta_l;\mathcal{D})\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
E\left\{\delta_0\frac{\frac{\partial}{\partial\theta_k}G(w)\frac{\partial}{\partial\theta_l}G(w)}{(1-G(w))^2}\right\}&=\frac{\frac{\partial}{\partial\theta_k}G(w)\frac{\partial}{\partial\theta_l}G(w)}{(1-G(w))^2}\times E(\delta_0)\\
&=\frac{\frac{\partial}{\partial\theta_k}G(w)\frac{\partial}{\partial\theta_l}G(w)}{(1-G(w))^2}\times(1-G(w))\\
&=\frac{\frac{\partial}{\partial\theta_k}G(w)\frac{\partial}{\partial\theta_l}G(w)}{1-G(w)}\\
&=A_0(\theta_k,\theta_l;\mathcal{D}).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
E\left\{\delta_1A_1(\theta_k,\theta_l;\mathcal{D})\right\}&=\int_{y=0}^{w}A_1(\theta_k,\theta_l;\mathcal{D})L_1(\mathcal{D};\boldsymbol{\theta})\mathrm{d}y\\
&=\int_{y=0}^{w}A_1(\theta_k,\theta_l;\mathcal{D})g(y)[1-F(z)]\mathrm{d}y.
\end{aligned}
$$

This holds because $\delta_1\delta_n=0$ for $n\neq 1$ and hence only the term for $n=1$ remains in the sum. Similarly,

$$
\begin{aligned}
&E\big\{\delta_nA_n(\theta_k,\theta_l;\mathcal{D})\big\}\\
&=\int_{y=0}^{w}\int_{x_1=0}^{w-y}\cdots\int_{x_{n-1}=0}^{w-y-\sum_1^{n-2}x_i}A_n(\theta_k,\theta_l;\mathcal{D})L_n(\mathcal{D};\boldsymbol{\theta})\mathrm{d}x_{n-1}\cdots\mathrm{d}x_1\,\mathrm{d}y,
\end{aligned}
$$

since the other terms in the sum are all zeroes.

A.3 Proof of Theorem 3

Let N_w denote the number of renewals in an observation window of length w. The FI on θ in a data set $\mathcal{D}$ from one WCRP is

$$
I(\theta;\mathcal{D})=-E\left(\frac{\partial^2}{\partial\theta^2}\log L(\mathcal{D};\theta)\right)
$$

$$
\begin{aligned}
&= \left[-\frac{\partial^2}{\partial\theta^2}\log(1-G(w))\right](1-G(w)) \\
&\quad + \int_0^w \left[-\frac{\partial^2}{\partial\theta^2}\log g(y) - \frac{\partial^2}{\partial\theta^2}\log(1-F(z))\right] g(y)[1-F(z)]\mathrm{d}y \\
&\quad + \sum_{n=2}^{\infty}\int\cdots\int_{\sum_{i=1}^{n-1}x_i+y\le w} \\
&\quad \times \left[-\frac{\partial^2\log g(y)}{\partial^2\theta} - \sum_{i=1}^{n-1}\frac{\partial^2\log f(x_i)}{\partial^2\theta} - \frac{\partial^2[1-F(z)]}{\partial^2\theta}\right] \\
&\quad \times g(y)f(x_1)\cdots f(x_{n-1})[1-F(z)]\mathrm{d}x_{n-1}\cdots\mathrm{d}x_1\,\mathrm{d}y \\
&= H_1 + H_2 + H_3,
\end{aligned}
$$

where $z = w - y - \sum_{i=1}^{n-1} x_i$ and

$$
\begin{aligned}
H_1 &= \left[-\frac{\partial^2}{\partial\theta^2}\log(1-G(w))\right](1-G(w)) \\
&\quad + \int_{y=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log g(y)\right] g(y)(1-F(z))\mathrm{d}y && (21) \\
&\quad + \sum_{n=2}^{\infty}\int_{y=0}^{w}\int_{x_1=0}^{w-y}\cdots\int_{x_{n-1}=0}^{w-y-\sum_{i=1}^{n-2}x_i}\left[-\frac{\partial^2}{\partial\theta^2}\log g(y)\right] \\
&\quad \times g(y)f(x_1)\cdots f(x_{n-1})[1-F(z)]\mathrm{d}x_{n-1}\cdots\mathrm{d}x_1\,\mathrm{d}y, && (22) \\
H_2 &= \sum_{n=2}^{\infty}\int_{y=0}^{w}\int_{x_1=0}^{w-y}\cdots\int_{x_{n-1}=0}^{w-y-\sum_{i=1}^{n-2}x_i}\left[-\sum_{i=1}^{n-1}\frac{\partial^2}{\partial\theta^2}\log f(x_i)\right] \\
&\quad \times g(y)f(x_1)\cdots f(x_{n-1})[1-F(z)]\mathrm{d}x_{n-1}\cdots\mathrm{d}x_1\,\mathrm{d}y,\ \text{and} \\
H_3 &= \int_{y=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log(1-F(z))\right] g(y)[1-F(z)]\mathrm{d}y && (23) \\
&\quad + \sum_{n=2}^{\infty}\int_{y=0}^{w}\int_{x_1=0}^{w-y}\cdots\int_{x_{n-1}=0}^{w-y-\sum_{i=1}^{n-2}x_i}\left[-\frac{\partial^2}{\partial\theta^2}\log(1-F(z))\right] \\
&\quad \times g(y)f(x_1)\cdots f(x_{n-1})[1-F(z)]\mathrm{d}x_{n-1}\cdots\mathrm{d}x_1\,\mathrm{d}y. && (24)
\end{aligned}
$$

We will prove that $\frac{H_1}{w}$ and $\frac{H_3}{w}$ go to zero when w goes to infinity and that $\frac{H_2}{w}$ goes to $\frac{I(\theta;X)}{\mu_X}$.

First consider H_1. The expression in (21) becomes

$$
\int_{y=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log g(y)\right] g(y)P(N_w = 1|Y = y)\mathrm{d}y,
$$

and (22) becomes

$$\begin{aligned}&\sum_{n=2}^{\infty}\int_{y=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log g(y)\right]g(y)\int_{x_1=0}^{w-y}\cdots\int_{x_{n-1}=0}^{w-y-\sum_{i=1}^{n-2}x_i}\\&\quad\times f(x_1)\cdots f(x_{n-1})[1-F(z)]\mathrm{d}x_{n-1}\cdots\mathrm{d}x_1\,\mathrm{d}y\\&\quad=\sum_{n=2}^{\infty}\int_{y=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log g(y)\right]g(y)P(N_w=n|Y=y)\mathrm{d}y.\end{aligned}$$

Thus, H_1 becomes

$$\begin{aligned}&\left[-\frac{\partial^2}{\partial\theta^2}\log(1-G(w))\right](1-G(w))\\&\quad+\sum_{n=1}^{\infty}\int_{y=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log g(y)\right]g(y)P(N=n|Y=y)\mathrm{d}y\\&=\left[-\frac{\partial^2}{\partial\theta^2}\log(1-G(w))\right](1-G(w))+\int_{y=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log g(y)\right]g(y)\mathrm{d}y,\end{aligned}$$

since whenever $y < w$, $\sum_{n=1}^{\infty} P(N=n|Y=y)=1$. Note that

$$\begin{aligned}\left[-\frac{\partial^2}{\partial\theta^2}\log(1-G(w))\right](1-G(w))&=-\left(\frac{\partial}{\partial\theta}\frac{-\frac{\partial}{\partial\theta}G(w)}{1-G(w)}\right)(1-G(w))\\&=\frac{\partial^2}{\partial\theta^2}G(w)+\frac{\left(\frac{\partial}{\partial\theta}G(w)\right)^2}{1-G(w)},\end{aligned}$$

where

$$\frac{\partial^2}{\partial\theta^2}G(w)=\frac{\partial^2}{\partial\theta^2}\int_0^w g(y)\mathrm{d}y\to\frac{\partial^2}{\partial\theta^2}1=0\ \text{ as }\ w\to\infty,$$

and

$$\frac{\partial}{\partial\theta}G(w)=\frac{\partial}{\partial\theta}\int_0^w g(y)\mathrm{d}y\to\frac{\partial}{\partial\theta}1=0\ \text{ as }\ w\to\infty.$$

Since $1-G(w)\to 0$ as $w\to\infty$, we apply L'Hopital's rule to determine the following limit.

$$\begin{aligned}\lim_{w\to\infty}\frac{\left(\frac{\partial}{\partial\theta}G(w)\right)^2}{1-G(w)}&=\lim_{w\to\infty}\frac{2\left(\frac{\partial}{\partial\theta}G(w)\right)\left(\frac{\partial^2}{\partial w\partial\theta}G(w)\right)}{-\frac{\partial}{\partial w}G(w)}\\&=\lim_{w\to\infty}\frac{2\left(\frac{\partial}{\partial\theta}G(w)\right)\left(\frac{\partial}{\partial\theta}g(w)\right)}{-g(w)}\end{aligned}$$

$$= \lim_{w\to\infty} -2\left(\frac{\partial}{\partial\theta}G(w)\right)\frac{\partial}{\partial\theta}\log g(w) \to 0$$

by assumptions (RC1) and (RC2). Thus $H_1 < \infty$ for any w.

Next, we consider H_3. The first term (23), is

$$\begin{aligned}
&\int_{z=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log(1-F(z))\right](1-F(z))g(w-z)\mathrm{d}z \\
&\quad = \int_{z=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log(1-F(z))\right](1-F(z))P(N_w=1|Z=z)\mathrm{d}z.
\end{aligned}$$

The terms in (24) can be expressed as

$$\begin{aligned}
&\sum_{n=2}^{\infty}\int_{z=0}^{w}\int_{x_1=0}^{w-z}\cdots\int_{x_{n-1}=0}^{w-z-\sum_{i=1}^{n-2}x_i}\left[-\frac{\partial^2}{\partial\theta^2}\log(1-F(z))\right](1-F(z)) \\
&\quad \times f(x_1)\cdots f(x_{n-1})g(w-\sum_{i=1}^{n-1}x_i-z)\mathrm{d}x_{n-1}\cdots\mathrm{d}x_1\,\mathrm{d}z \\
&= \sum_{n=2}^{\infty}\int_{z=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log(1-F(z))\right](1-F(z))\int_{x_1=0}^{w-z}\cdots\int_{x_{n-1}}^{w-z-\sum_{i=1}^{n-2}x_i} \\
&\quad \times f(x_1)\cdots f(x_{n-1})g(w-\sum_{i=1}^{n-1}x_i-z)\mathrm{d}x_{n-1}\cdots\mathrm{d}x_1\,\mathrm{d}z \\
&= \sum_{n=2}^{\infty}\int_{z=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log(1-F(z))\right](1-F(z))P(N_w=n|Z=z)\mathrm{d}z \\
&= \int_{z=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log(1-F(z))\right](1-F(z))\sum_{n=2}^{\infty}P(N_w=n|Z=z)\mathrm{d}z.
\end{aligned}$$

So H_3 becomes

$$\begin{aligned}
&\sum_{n=1}^{\infty}\int_{z=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log(1-F(z))\right](1-F(z))P(N_w=n|Z=z)\mathrm{d}z \\
&\quad = \int_{z=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log(1-F(z))\right](1-F(z))\sum_{n=1}^{\infty}P(N_w=n|Z=z)\mathrm{d}z \\
&\quad = \int_{z=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log(1-F(z))\right](1-F(z))\mathrm{d}z.
\end{aligned}$$

From the regularity condition (RC2) in the theorem we have

$$\int_0^\infty \left(-\frac{\partial^2}{\partial\theta^2}\log g(y)\right) g(y)\mathrm{d}y < \infty$$
$$\iff \int_0^\infty \left(-\frac{\partial^2}{\partial\theta^2}\log\frac{1-F(y)}{\mu_X}\right)\frac{1-F(y)}{\mu_X}\mathrm{d}y < \infty$$
$$\iff \frac{\int_0^\infty \left(-\frac{\partial^2}{\partial\theta^2}\log(1-F(y))\right)(1-F(y))\mathrm{d}y + \int_0^\infty \left(\frac{\partial^2}{\partial\theta^2}\log\mu_X\right)(1-F(y))\mathrm{d}y}{\mu_X} < \infty$$
$$\iff \frac{\int_0^\infty \left(-\frac{\partial^2}{\partial\theta^2}\log(1-F(y))\right)(1-F(y))\mathrm{d}y + \left(\frac{\partial^2}{\partial\theta^2}\log\mu_X\right)\mu_X}{\mu_X} < \infty$$
$$\iff \int_0^\infty \left(-\frac{\partial^2}{\partial\theta^2}\log(1-F(y))\right)(1-F(y))\mathrm{d}y < \infty;$$

i.e., $H_3 < \infty$.

The expression H_2 is

$$\sum_{n=2}^{\infty}\sum_{j=1}^{n-1}\int_{y=0}^{w}\int_{x_1=0}^{w-y}\cdots\int_{x_{n-1}=0}^{w-y-\sum_{i=1}^{n-2}x_i}\left[-\frac{\partial^2}{\partial\theta^2}\log f(x_j)\right]$$
$$\times\, g(y)f(x_1)\cdots f(x_{n-1})[1-F(z)]\mathrm{d}x_{n-1}\cdots\mathrm{d}x_1\,\mathrm{d}y$$
$$=\sum_{n=2}^{\infty}\sum_{j=1}^{n-1}\int_{x_j=0}^{w}\int_{y=0}^{w-x_j}\int_{x_1=0}^{w-y-x_j}\cdots\int_{x_{n-1}=0}^{w-y-\sum_{i=1}^{n-2}x_i}\left[-\frac{\partial^2}{\partial\theta^2}\log f(x_j)\right]$$
$$\times f(x_j)g(y)f(x_1)\cdots f(x_{n-1})[1-F(z)]\mathrm{d}x_{n-1}\cdots\mathrm{d}x_1\,\mathrm{d}y\,\mathrm{d}x_j$$

and H_2 can be expressed as

$$=\sum_{n=2}^{\infty}\sum_{j=1}^{n-1}\int_{x_j=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log f(x_j)\right]f(x_j)P(N_w=n|X_j=x_j)\mathrm{d}x_j$$
$$=\sum_{n=2}^{\infty}(n-1)\int_{x=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log f(x)\right]f(x)P(N_w=n|X=x)\mathrm{d}x$$
$$=\int_{x=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log f(x)\right]f(x)\sum_{n=2}^{\infty}(n-1)P(N_w=n|X=x)\mathrm{d}x$$
$$=\int_{x=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log f(x)\right]f(x)$$
$$\times\left(\sum_{n=2}^{\infty}nP(N_w=n|X=x)-\sum_{n=2}^{\infty}P(N_w=n|X=x)\right)\mathrm{d}x$$
$$=\int_{x=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log f(x)\right]f(x)\left(E(N_w|X=x)-1\right)\mathrm{d}x.$$

We have

$$\frac{H_2}{w} = \frac{\int_{x=0}^{w} \left[-\frac{\partial^2}{\partial\theta^2}\log f(x)\right] f(x)\,(E(N_w|X=x)-1)\,\mathrm{d}x}{w}$$
$$= \frac{\int_{x=0}^{w} \left[-\frac{\partial^2}{\partial\theta^2}\log f(x)\right] f(x)\,(E(N_w|X=x))\,\mathrm{d}x}{w}$$
$$- \frac{\int_{x=0}^{w} \left[-\frac{\partial^2}{\partial\theta^2}\log f(x)\right] f(x)\mathrm{d}x}{w}$$
$$= \int_{x=0}^{w} \left[-\frac{\partial^2}{\partial\theta^2}\log f(x)\right] f(x)\frac{E(N_w|X=x)}{w}\mathrm{d}x$$
$$- \frac{\int_{x=0}^{w} \left[-\frac{\partial^2}{\partial\theta^2}\log f(x)\right] f(x)\mathrm{d}x}{w}.$$

Since

$$\lim_{w\to\infty} \frac{\int_{x=0}^{w} \left[-\frac{\partial^2}{\partial\theta^2}\log f(x)\right] f(x)\mathrm{d}x}{w} = 0,$$

if we have

$$\lim_{w\to\infty} \int_{x=0}^{w} \left[-\frac{\partial^2}{\partial\theta^2}\log f(x)\right] f(x)\frac{E(N_w|X=x)}{w}\mathrm{d}x = \frac{I(\theta;X)}{\mu_X}, \tag{25}$$

we will have

$$\lim_{w\to\infty} \frac{H_2}{w} = \frac{I(\theta;X)}{\mu_X} + 0 = \frac{I(\theta;X)}{\mu_X},$$

and overall,

$$\frac{I(\theta;\mathcal{D})}{w} = \frac{H_1+H_2+H_3}{w} \to 0 + \frac{I(\theta;X)}{\mu_X} + 0 = \frac{I(\theta;X)}{\mu_X}.$$

Now we prove (25) is true.

Since

$$\lim_{w\to\infty} \frac{E(N_w)}{w} = \frac{1}{\mu_X},$$

$\forall\delta > 0$, $\exists$ a w_1 such that

$$\frac{E(N_w)}{w} < \frac{1}{\mu_X} + \delta \quad \text{for all} \quad w > w_1.$$

By condition (RC1), for all $\epsilon > 0$ we can find a w_2 large enough such that

$$\int_w^\infty \left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right) f(x)\mathrm{d}x < \epsilon' = \min\left\{\frac{\epsilon}{3}\mu_X, \frac{\epsilon}{3(\frac{1}{\mu_X+\delta})}, \frac{\epsilon}{3}\right\}$$

for all $w > w_2$. Now

$$\begin{aligned} E(N_w|X=x) &= E(\text{number of renewals in a window of length } w|X=x) \\ &= E(\text{number of renewals in a window of length } (w-x)) + 1 \\ &= E(N_{w-x}) + 1. \end{aligned}$$

Then for $w_2 < w_0 < w$,

$$\int_0^w \left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right) f(x)\frac{E(N_w|X=x)}{w}\mathrm{d}x$$

$$= \int_0^{w_0} \left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right) f(x)\frac{E(N_{w-x})}{w}\mathrm{d}x \quad (26)$$

$$+ \int_{w_0}^{w} \left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right) f(x)\frac{E(N_{w-x})}{w}\mathrm{d}x \quad (27)$$

$$+ \frac{1}{w}\int_0^{w} \left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right) f(x)\mathrm{d}x. \quad (28)$$

Denote the expressions in (26)–(28) as K_1, K_2 and K_3, respectively. It is obvious that

$$\begin{aligned} \int_0^{w_0} \left(-\frac{\partial^2\log f(x)}{\partial\theta^2}\right) f(x)\frac{E(N_{w-w_0})}{w}\mathrm{d}x &\le K_1 \\ &\le \int_0^{w_0} \left(-\frac{\partial^2\log f(x)}{\partial\theta^2}\right) f(x)\frac{E(N_w)}{w}\mathrm{d}x. \end{aligned}$$

Since

$$\lim_{w\to\infty}\frac{E(N_{w-w_0})}{w} = \frac{1}{\mu_X} \quad \text{and} \quad \lim_{w\to\infty}\frac{E(N_w)}{w} = \frac{1}{\mu_X},$$

we have

$$\lim_{w\to\infty} K_1 = \int_0^{w_0} \left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right) f(x)\frac{1}{\mu_X}\mathrm{d}x.$$

So, for all $\epsilon > 0$, that exists a w_3 such that

$$\left|K_1 - \int_0^{w_0} \left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right) f(x)\frac{1}{\mu_X}\mathrm{d}x\right| \le \frac{\epsilon}{6} \quad \text{for all } w > w_3.$$

Also,

$$\begin{aligned}|K_2| &\le \left|\int_{w_0}^{w}\left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right)f(x)\frac{E(N_w)}{w}\mathrm{d}x\right|\\ &\le \left(\frac{1}{\mu_X}+\delta\right)\left|\int_{w_0}^{w}\left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right)f(x)\mathrm{d}x\right| \text{ for } w>w_1\\ &\le \left(\frac{1}{\mu_X}+\delta\right)\left|\int_{w_0}^{\infty}\left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right)f(x)\mathrm{d}x\right|\\ &\le \left(\frac{1}{\mu_X}+\delta\right)\frac{\epsilon}{3\left(\frac{1}{\mu_X}+\delta\right)} \text{ for } w_0>w_2\\ &= \frac{\epsilon}{3}.\end{aligned}$$

Since $|K_3|$ is bounded by $I(\beta;X)/w$, it approaches 0 as $w\to\infty$ and hence there exists a w_4 such that $|(28)|\le\epsilon/6$ for all $w>w_4$. So,

$$\begin{aligned}&\left|\int_{x=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log f(x)\right]f(x)\frac{E(N_w|X=x)}{w}\mathrm{d}x-\frac{I(X;\theta)}{\mu_X}\right|\\ &=\left|\int_{0}^{w_0}\left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right)f(x)\frac{E(N_{w-x})}{w}\mathrm{d}x\right.\\ &\quad+\int_{w_0}^{w}\left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right)f(x)\frac{E(N_{w-x})}{w}\mathrm{d}x+\frac{1}{w}\int_{0}^{w}\left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right)f(x)\mathrm{d}x\\ &\quad\left.-\int_{0}^{w_0}\left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right)f(x)\frac{1}{\mu_X}\mathrm{d}x-\int_{w_0}^{\infty}\left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right)f(x)\frac{1}{\mu_X}\mathrm{d}x\right|\\ &\le\left|\int_{0}^{w_0}\left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right)f(x)\frac{E(N_{w-x})}{w}\mathrm{d}x-\int_{0}^{w_0}\left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right)f(x)\frac{1}{\mu_X}\mathrm{d}x\right|\\ &\quad+\left|\int_{w_0}^{w}\left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right)f(x)\frac{E(N_{w-x})}{w}\mathrm{d}x\right|+\left|\int_{0}^{w}\left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right)f(x)\mathrm{d}x\right|\\ &\quad+\left|\int_{w_0}^{\infty}\left(-\frac{\partial^2}{\partial\theta^2}\log f(x)\right)f(x)\frac{1}{\mu_X}\mathrm{d}x\right|\\ &\le\frac{\epsilon}{6}+\frac{\epsilon}{3}+\frac{\epsilon}{6}+\frac{\epsilon}{3} \text{ for } w>\max\{w_1,w_2,w_3,w_4\} \text{ and } w_0>w_2.\end{aligned}$$

Since we can make ϵ arbitrarily small, we have

$$\lim_{w\to\infty}\int_{x=0}^{w}\left[-\frac{\partial^2}{\partial\theta^2}\log f(x)\right]f(x)\frac{E(N_w|X=x)}{w}\mathrm{d}x=\frac{I(X;\theta)}{\mu_X}.$$

Acknowledgments We thank the anonymous referee for insightful comments.

References

Alvarez, E. E. (2005). Estimation in stationary markov renewal processes, with application to earthquake forecasting in Turkey. *Methodology and Computing in Applied Probability, 7*(1), 119–130.
Alvarez, E. E. (2006). Maximum likelihood estimation in alternating renewal processes under window censoring. *Stochastic Models, 22*(1), 55–76.
Anderson, T. W. (1984). *An introduction to multivariate statistical analysis* (2nd ed.). Hoboken: Wiley.
Basu, A. P., Rigdon, S. E. (1986). Examples of parametric empirical Bayes methods for the estimation of failure processes for repairable systems. In A. P. Basu (Ed.), *Reliability and quality control* (pp. 47–55). Amsterdam: North-Holland.
Cox, D. R. (1962). *Renewal theory*. London/Hoboken: Methuen and Co. Ltd./Wiley.
Denby, L., Vardi, Y. (1985). A short-cut method for estimation in renewal processes. *Technometrics, 27*(4), 361–373.
Karlin, S., Taylor, H. M. (1975). *A first course in stochastic processes* (2nd ed.). NY: Academic Press.
Lehmann, E. L., Casella, G. (1998). *Theory of point estimation* (2nd ed.). NY: Springer.
Nelson, W. B. (2003). *Recurrent events data analysis for product repairs, disease recurrences, and other applications*. Philadelphia: Society for Industrial and Applied Mathematics (SIAM).
Rigdon, S. E., Basu, A. P. (2000). *Statistical methods for the reliability of repairable systems*. NY: Wiley.
Soon, G., Woodroofe, M. (1996). Nonparametric estimation and consistency for renewal processes. *Journal of Statistical Planning and Inference, 53*(2), 171–195.
Vardi, Y. (1982). Nonparametric estimation in the presence of length bias. *The Annals of Statistics, 10*(2), 616–620.
Zhao, Y. (2006). *Parametric inference from window censored renewal process data*. PhD dissertation, The Ohio State University, Columbus, Ohio.