A Simulation Study on Confidence Interval Procedures of Some Mean Cumulative Function Estimators

Jianying Zuo
William Q. Meeker
Huaiqing Wu
Department of Statistics
Iowa State University
Ames, IA 50011

January 25, 2010

Abstract

Recurrence data are collected to study the recurrent events on biological, physical, and other systems. Quantities of interest include the mean cumulative number of events and the mean cumulative cost of events. The mean cumulative function (MCF) can be estimated with nonparametric methods or by fitting parametric models, and many procedures have been suggested to construct the confidence interval (CI) for the MCF. This paper summarizes the results of a large simulation study that was designed to compare five CI procedures for both the nonparametric and parametric estimation. When doing parametric estimation, we assume the power law non-homogeneous Poisson process (NHPP) model. Our results include evaluation of these procedures when they are used for window-observation recurrence data where recurrence histories of some systems are available only in observation windows with gaps in between.

KEY WORDS: Mean cumulative function; MCF; Nonhomogeneous Poisson process; Nonparametric estimation; Recurrence data.

1 INTRODUCTION AND BACKGROUND

1.1 Background and Motivation

Recurrence data arise in many applications, including business processes, medical statistics, and repairable system reliability. Both nonparametric and parametric methods are available to estimate the mean cumulative function (MCF). Although many important questions can be answered by using nonparametric methods, in some applications, parametric models are needed. Nelson (2003) presents graphical and nonparametric statistical methods for recurrence data and describes many applications. Cook and Lawless (2007) provide a comprehensive account of statistical methods to analyze recurrent event data. They describe nonparametric, parametric, and semipamametric methods, give mathematical and statistical background, and also present many applications.

There are several procedures that can be used to construct approximate confidence intervals (CIs) for the MCF. Normal approximation CIs are relatively easy to implement and compute. With modern powerful computers, bootstrap or simulation-based procedures can also be used.

The adequacy of approximate CI procedures depends on the underlying model as well as the number of units and the length of time that each unit is observed. With constraints on resources to observe the process and collect the data, however, it may be difficult to have data with both a large number of observational units and a long time of observation. To help explore the impact of various factors on the performances of CI procedures, we carried out an extensive simulation study to compare five CI procedures. Two of these procedures are normal approximation procedures, and the other three are bootstrap-based procedures. Section 3 gives technical details on how to compute the intervals.

1.2 Window-Observation Data

In some applications, window-observation recurrence data arise because recurrence history of some units are observed in disconnected windows with gaps between the windows. Nelson (2003, page 75) describes an example. Zuo, Meeker and Wu (2008) provide more examples, and extend some nonparametric (NP) and parametric estimation methods to window-observation recurrence data. If there are intervals of time over which the size of the risk set is zero, the NP method can be seriously biased. For such scenarios, Zuo, Meeker and Wu (2008) describe two hybrid MCF estimators that help correct the bias. In this paper, we evaluate CI procedures for both complete and window-observation data.

1.3 Other Previous Related Work

Cook and Lawless (2007) describe different approaches to the modeling of recurrent events such as models based on counts of events, intensity, and time between events. Many publications provide descriptions of analysis and modeling of counting processes, such as Cox and Lewis (1966) and Andersen, Borgan, Gill and Keiding (1993).

There are a number of useful text books that describe bootstrap procedures for computing CIs. For example, Efron and Tibshirani (1993) present basic background of bootstrap methods and many applications to statistical procedures. More specifically, Chapters 12 to 14, and 22 in Efron and Tibshirani (1993) present bootstrap procedures to construct CIs. Hall (1992) interweaves the topics of bootstrap and Edgeworth expansion, and applies Edgeworth expansion methods to characterize the performance of some bootstrap methods.

1.4 Overview

The remainder of this paper is organized as follows. Section 2 describes the model and estimation of the MCF. Section 3 explains the five CI procedures in the simulation study, and Section 4 outlines the details of the simulation study. Section 5 shows the impact of the recurrence rate function on the performances of the CI procedures. Sections 6 to 8 summarize the performances of the CI procedures for the NP estimator, the power law NHPP estimator, and the hybrid estimators, respectively. Concluding remarks and areas for future research are outlined in Section 9. The appendices present some necessary technical details for the NP and NHPP estimators.

2 MODEL AND ESTIMATION

In this section, we describe briefly the recurrence data model and provide formulas for point estimators.

2.1 Notation and Acronyms

We will use the following notation:

- n: the number of units under observation
- t_{endobs}: the pre-specified end-of-observation time for the observational units in the simulation. For the simulation study, recurrences and observation windows up to this time point are recorded in the data

- E(r): expected total number of observed recurrences for all n units over the time range (0, t_{endobs})
- RSSZ: risk-set-size-zero
- RSSONE: risk-set-size-one
- RSSP: risk-set-size-positive
- NP: nonparametric
- NHPP: non-homogeneous Poisson process
- CP: coverage probability

2.2 Model

The models in our study are based on counts of events (i.e., the number of events in some time range of interest, say $[0, t]$). Let $N(t)$ denote the number of events in the time range $[0, t]$. Then one statistic of interest is $\mu(t)$, the expectation of $N(t)$, which is also known as the mean cumulative function (MCF). Our goal is to estimate the MCF.

If the MCF is differentiable, then $\nu(t) = d\mu(t)/dt$ is the recurrence rate and $\nu(t) \times \Delta t$ can be interpreted as the approximate expected number of events to occur during the next short time interval $(t,\ t + \Delta t)$.

The nonparametric approaches do not make assumptions about the form of the recurrence process, while the parametric ones do. Model assumptions for the nonparametric estimation methods are stated in Nelson (2003) and Zuo, Meeker and Wu (2008). Model assumptions for the NHPP parametric estimation methods are given in Rigdon and Basu (2000).

2.3 Estimation for the Nonparametric Model

Detailed descriptions of nonparametric MCF estimation methods are available in Nelson (1988), Lawless and Nadeau (1995), Chapter 16 of Meeker and Escobar (1998), and Chapters 3 to 5 of Nelson (2003). Zuo, Meeker and Wu (2008) extend the nonparametric method to window-observation recurrence data.

Let m denote the number of unique event times. Also, let $t_1, ..., t_m$ be the unique event times. Then the nonparametric estimator of the population MCF is

$$\widehat{MCF}_{NP}(t_j) = \sum_{k=1}^{j} \left[\frac{\sum_{i=1}^{n} \delta_i(t_k) \times d_i(t_k)}{\sum_{i=1}^{n} \delta_i(t_k)} \right] = \sum_{k=1}^{j} \frac{d.(t_k)}{\delta.(t_k)} = \sum_{k=1}^{j} \bar{d}(t_k), \quad j = 1, ..., m, \tag{1}$$

where $d_i(t_k)$ is the number of events recorded at time t_k for unit i, and

$$\delta_i(t_k) = \begin{cases} 1 & \text{if unit } i \text{ is under observation in a time window at time } t_k, \\ 0 & \text{otherwise.} \end{cases}$$

Details on estimators of $\text{Var}[\widehat{MCF}_{NP}(t_j)]$ are available in Nelson (1995), Lawless and Nadeau (1995), and Zuo, Meeker and Wu (2008).

2.4 Estimation for the NHPP Parametric Model

NHPP models and estimation methods for recurrence data are described, for example, in Rigdon and Basu (2000, Chapter 2) and Meeker and Escobar (1998, Chapter 16). Zuo, Meeker and Wu (2008) extend the NHPP estimation methods to window-observation recurrence data. Given the maximum likelihood (ML) estimates of the model parameters $\hat{\boldsymbol{\theta}}$, the ML estimator of the NHPP MCF is

$$\widehat{MCF}_{NHPP}(t) = \int_0^t \nu(x; \hat{\boldsymbol{\theta}})dx. \tag{2}$$

With the estimate of the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, the delta method can be used to estimate $\text{Var}[\widehat{MCF}_{NHPP}(t)]$. Zuo, Meeker and Wu (2008) give more details, using the power law NHPP model as an example.

2.5 Hybrid MCF Estimators for Window-Observation Recurrence Data

Zuo, Meeker and Wu (2008) introduce two hybrid MCF estimators for window-observation recurrence data – the local hybrid estimator and the NHPP hybrid estimator. Such estimators are needed because the existence of RSSZ intervals can cause $\widehat{MCF}_{NP}(t)$ to be seriously biased.

The local hybrid estimator is $\widehat{MCF}_{LH}(t) = \bar{d}.(t) + d^{\ddagger}_{\cdot}(t)$, where $\bar{d}.(t)$ is the nonparametric estimator of the increase in the MCF from RSSP intervals, while $d^{\ddagger}_{\cdot}(t)$ is the estimator of the increase in the MCF from RSSZ intervals, assuming the recurrence rate of the RSSZ interval is the weighted average of the recurrence rates of the two neighboring RSSP intervals.

The NHPP hybrid estimator is $\widehat{MCF}_{NHPPH}(t) = \bar{d}.(t) + d^{\dagger}_{\cdot}(t)$, where $d^{\dagger}_{\cdot}(t)$ is the estimator of the increase in the MCF from RSSZ intervals, assuming the counts of recurrences in the RSSZ intervals follow an NHPP model. The NHPP model is estimated with the data in the RSSP intervals.

2.6 Recommendations on Selection of MCF Estimators

Zuo, Meeker and Wu (2008) present a brief summary from a simulation study that compared the NP estimator, the power law NHPP estimator, the local hybrid estimator, and the NHPP hybrid estimator. The nonparametric approaches, such as the NP estimator and the local hybrid estimator, generally have little or no bias, but might have a large variance. On the other hand, more parametric based approaches generally have small variance, but could be seriously biased if the assumed model form is very different from the true model. Therefore, selection among different MCF estimators becomes a bias-variance tradeoff, and model assumption diagnosis is important in the model selection and estimation process. This paper presents the results of a much more extensive simulation to compare the properties of CI procedures based on the same estimators.

3 CONFIDENCE INTERVAL PROCEDURES

In this section, we outline two normal approximation CI procedures and three bootstrap CI procedures. These five CI procedures are evaluated in our simulation experiments, and applied to the four MCF estimators described in Section 2. We use the following general notation.

- $\widehat{MCF}$: estimate of the MCF from the original data.
- $\widehat{SE}_{\widehat{MCF}}$: standard error of $\widehat{MCF}$ (i.e., estimate of the standard deviation of $\widehat{MCF}$) from the original data.
- $\widehat{MCF}^*$: estimate of the MCF from the bootstrap re-sampled data.
- $\widehat{SE}^*_{\widehat{MCF}}$: standard error of $\widehat{MCF}$ (i.e., estimate of the standard deviation of $\widehat{MCF}$) from the bootstrap re-sampled data.
- t^*: t-like ratio from the bootstrap re-sampled data, computed as $t^* = \left(\widehat{MCF}^* - \widehat{MCF}\right)/\widehat{SE}^*_{\widehat{MCF}}$.

Only the $\widehat{MCF}$ and the $\widehat{SE}_{\widehat{MCF}}$ are needed to construct normal approximation CIs. There are many possible normal approximation CI procedures, depending on the transformation used. The two normal approximation procedures a) and b) below are from Meeker and Escobar (1998, Chapter 16, page 400). In the formulas, $z_{(1-\alpha/2)}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution.

Efron and Tibshirani (1993) describe various bootstrap procedures. The common first step in any bootstrap procedure is to generate bootstrap samples from the original data, and these bootstrap samples are used to calculate $\widehat{MCF}^*$ and $\widehat{SE}^*_{\widehat{MCF}}$ and, for the bootstrap-t procedures, t^*. To construct a bootstrap interval, we repeat the sampling and estimation process a large number of times (say B times), and sort $\widehat{MCF}^*$ or t^* values (depending on the procedure). Let k be the largest integer less than or equal to $(B+1)\alpha/2$, where α is the complement of the desired nominal CP. The three bootstrap CIs can be obtained by c), d), and e) below, where $y_{(k)}$ indicates the k^{th}

ordered value in a sequence where y has been ordered from smallest to largest. Specifically, the five CI procedures that we evaluate in our simulation are

a) *Normal Approximation (NORMA)*: $\widehat{MCF} \pm z_{(1-\alpha/2)}\widehat{SE}_{\widehat{MCF}}$;

b) *Lognormal Approximation (LNORMA)*: $\left[\widehat{MCF}/w, \quad \widehat{MCF} \times w\right]$, where $w = \exp\left[z_{(1-\alpha/2)}\widehat{SE}_{\widehat{MCF}}/\widehat{MCF}\right]$;

c) *Bootstrap Percentile (BootP)*: $\left[\widehat{MCF}^{*}_{(k)}, \quad \widehat{MCF}^{*}_{(B+1-k)}\right]$;

d) *Bootstrap-t (Boott)*: $\left[\widehat{MCF} - t^{*}_{(B+1-k)}\widehat{SE}_{\widehat{MCF}}, \quad \widehat{MCF} - t^{*}_{(k)}\widehat{SE}_{\widehat{MCF}}\right]$, where $t^{*} = \frac{\widehat{MCF}^{*} - \widehat{MCF}}{\widehat{SE}^{*}_{\widehat{MCF}}}$;

e) *Bootstrap-t Based on Log Transformation (LBoott)*: $\left[\frac{\widehat{MCF}}{\exp\left(t^{*}_{(B+1-k)}\widehat{SE}_{\widehat{MCF}}/\widehat{MCF}\right)}, \quad \frac{\widehat{MCF}}{\exp\left(t^{*}_{(k)}\widehat{SE}_{\widehat{MCF}}/\widehat{MCF}\right)}\right]$,

where $t^{*} = \left(\log(\widehat{MCF}^{*}) - \log(\widehat{MCF})\right)/\widehat{SE}^{*}_{\log(\widehat{MCF})}$ and $\widehat{SE}^{*}_{\log(\widehat{MCF})} = \widehat{SE}^{*}_{\widehat{MCF}}/\widehat{MCF}^{*}$.

For simplification, we use the acronyms in the parentheses to represent the five CI procedures.

4 SIMULATION EXPERIMENTAL DESIGN

4.1 Factors and Factor Levels

In previous simulation experiments to study confidence interval properties for censored lifetime data (e.g., Jeng and Meeker 2000), it was shown that the adequacy of asymptotic approximations tend to depend on the number of failures, rather than the sample size. Thus, an important experimental factor in our simulation experiment is the expected number of events, $\mathrm{E}(r)$, with four factor levels at 10, 20, 50, and 100. $\mathrm{E}(r)$ is, however, affected by the following factors:

- **The pattern of the observation windows for the units in the data set (Window Schemes)**. Three window schemes, corresponding to data sets analyzed in Zuo, Meeker and Wu (2008), are used in our study, and they are

  1. *Complete* data: All units in the data are observed continuously in the same single window $[0, t_{endobs}]$.

  2. *Window1* data: There are some gaps between the observation windows for each unit. Length of the observation window follows a uniform distribution between 0.08 and 0.12, while length of the gap follows a uniform distribution between 0.12 to 0.28. Whether a unit begins with a window or a gap follows a Bernoulli (0.5) distribution.

3. *Window2* data: Similar to *Window1* data, except that the length for gap i is simulated from a uniform distribution between $0.04 \times 2^{(i-1)}$ and $0.08 \times 2^{(i-1)}$. Therefore, for units in a *Window2* data set, as time gets larger, the probability of observing a given recurrence gets smaller.

We did simulations using all three window schemes. However, the results of Window1 data are similar to those of Complete data, mainly because the percentage of times with RSSZ and RSSONE is zero or very small. Therefore, we focus on the results from the Complete data and Window2 data.

- **Number of Units**: n, with five levels at the values 10, 20, 50, 100, and 200 units being observed.

- **Form of the MCF of the recurrence process**. We use the power law NHPP model with the recurrence function as

$$\nu(t; \beta, \eta) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta - 1}, \quad \beta > 0, \eta > 0. \tag{3}$$

  Without loss of generality, we use $\eta = 1$. For the shape parameter β, there are four levels at values 0.8 (decreasing recurrence rate), 1 (constant recurrence rate), 2 (moderately increasing recurrence rate), and 3 (rapidly increasing recurrence rate).

Note that when **Window Schemes**, **Number of Units**, and **Form of the MCF** are given, t_{endobs} depends only on the value of $\mathrm{E}(r)$.

4.2 Simulation Algorithm

Given a set of factor levels from our simulation design, the following procedure was carried out to estimate the CPs of the five CI procedures.

1. Generate simulated data based on the inputs for the window scheme, n, β, and t_{endobs}.

2. If the window scheme is Window1 or Window2, use the simulated data to compute the $\widehat{MCF}$ at t_{endobs} for the four different estimators – the NP estimator, the power law NHPP estimator, the local hybrid estimator, and the power law NHPP hybrid estimator. For the Complete data case, we compute only the NP estimator and the power law NHPP estimator. We also compute the corresponding standard errors.

3. Compute the NORMA and the LNORMA CIs described in Section 3.

4. Generate bootstrap samples, and construct the BootP, the Boott, and the LBoott CIs described in Section 3. In our simulation study, we used simple random sampling with replacement (SRSWR) to draw the whole history of a unit and thus created bootstrap re-sampled data that had the same number of units in the original data. We used $B = 2000$ bootstrap samples.

5. Check whether the five CIs obtained in Steps 3 and 4 capture the true MCF at t_{endobs}, assign to coverage indicator a value of 1 if the true MCF is in the CI and a value of 0 otherwise.

6. Repeat Steps 1 to 5 a large number of times (5000 in our simulation study) and then calculate the average of coverage indicators as the estimate of the CP for each of the five CI procedures.

When the number of observed recurrences is small (e.g., 4 or fewer), MCF estimates are poor and there can be estimation problems (e.g., $\boldsymbol{\theta}$ is not estimable if the dimension of $\boldsymbol{\theta}$ is larger than the number of observed recurrences for $\widehat{MCF}_{NHPP}$, or some components of the variance estimator of $\widehat{MCF}_{NP}$ are not estimable if fewer than two units are being observed for the time with observed recurrences). Thus we used only simulated data sets with 5 or more distinct recurrence times, as well as bootstrap re-sampled data with 3 or more distinct recurrence times, to estimate CP of the MCF estimators. That is, our simulation results are conditional on $\sum_{i=1}^{n} X_i \geq 5$, where $\sum_{i=1}^{n} X_i$ is the total number of observed recurrences among all n units. As described in Appendix A.1, the values of $\Pr\left(\sum_{i=1}^{n} X_i \leq 4\right)$ are 0.0293, 1.69×10^{-5}, 5.45×10^{-17}, and 1.61×10^{-37} respectively for $\mathrm{E}(r) = 10$, 20, 50, and 100. Therefore, when $\mathrm{E}(r) = 20$, 50, and 100, the conditional probabilities are close to 1.

5 THE EFFECT THAT THE RECURRENCE RATE SHAPE HAS ON CI PROCEDURES

In our simulation study, we assume that the true model is the power law NHPP model. For a power law NHPP model, the value of β determines whether the recurrence rate is increasing (for $\beta > 1$), constant (for $\beta = 1$), or decreasing (for $\beta < 1$). One question of interest is how the shape of the recurrence rate across time affects the performances of the five CI procedures. In particular, we will study whether the estimated CP values are approximately the same or have some pattern across the different values of β in this section, and we will discuss the comparisons among the five CI procedures in Sections 6 to 8.

In order to graphically show the impact of n and $\mathrm{E}(r)$, there are six plots for each of the estimators in Figures 1 to 6, arranged in three rows and two columns. In rows 1 to 3, n is 10, 20, and 50, respectively. $\mathrm{E}(r)$ is 10 on the left and 20 on the right.

5.1 Results from Complete Data

First we describe the simulation results based on Complete data, where each unit in the data is observed from time zero to t_{endobs}. Because there are no observation gaps for the Complete data, we only need to compare CI procedures for the NP method and the power law NHPP method. Figures 1 and 2 are for the NP estimator and the power law NHPP estimator respectively, and some noticeable patterns for both estimators are:

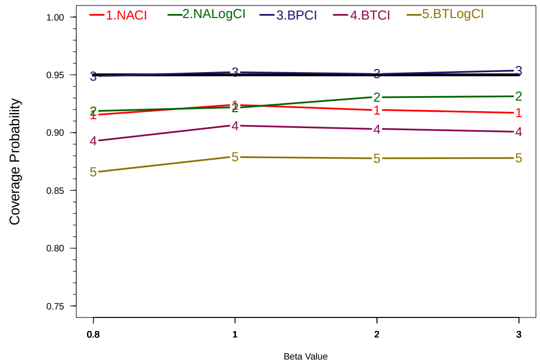
- CP values are about parallel across the four values of β for each of the five CI procedures, and a parallel pattern exists for all n and $\mathrm{E}(r)$ values.
- At each n level, when $\mathrm{E}(r)$ increases from 10 to 20, differences among the five CI procedures get smaller and the CP values are more closely clustered around the nominal value 0.95. This trend continues with higher $\mathrm{E}(r)$ values, and thus plots with $\mathrm{E}(r) \geq 50$ are not shown.
- For $\mathrm{E}(r) = 20$, differences among the five CI procedures get smaller when n increases. However, when $\mathrm{E}(r) = 10$, the maximum distance among the CP values of the five CI procedures does not get smaller when n increases. This indicates that $\mathrm{E}(r)$ must be in the order of 20 for the large sample approximations to be adequate, even if n is large.

The explanation for the observed parallel patterns in Figures 1 and 2 is that $\widehat{MCF}_{NP}(t_{endobs}) = \widehat{MCF}_{NHPP}(t_{endobs}) = \sum_{i=1}^{n} X_i/n$, where X_i is the number of observed recurrences for unit i and $\sum_{i=1}^{n} X_i$ follows a Poisson distribution with $\lambda = \mathrm{E}(r)$, as shown in Appendix A. Therefore, $\widehat{MCF}_{NP}(t_{endobs})$ and $\widehat{MCF}_{NHPP}(t_{endobs})$ are proportional to a Poisson random variable with a mean $\lambda = \mathrm{E}(r)$, not depending on the value of β. Appendix A also shows that the distributions of $\widehat{\mathrm{Var}}[\widehat{MCF}_{NP}(t_{endobs})]$ and $\widehat{\mathrm{Var}}[\widehat{MCF}_{NHPP}(t_{endobs})]$ depend only on $\mathrm{E}(r)$ and n as well. Therefore, how the recurrences are distributed across time (more clustered at the beginning of life for $\beta < 1$, equally likely across time for $\beta = 1$, and with higher density as time increases for $\beta > 1$) does not have a strong effect on the performances of the CI procedures and MCF estimators, as long as the expected number of recurrences for the time of estimation is the same.

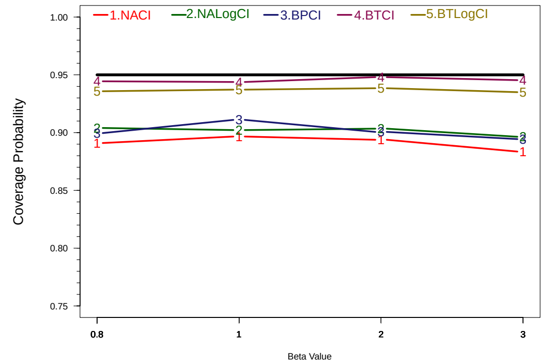
5.2 Results from Window2 Data

Figures 3 to 6 show the plots of simulation results using Window2 data, for the four estimators, organized as in Figure 1. The main observations from the four sets of plots are:

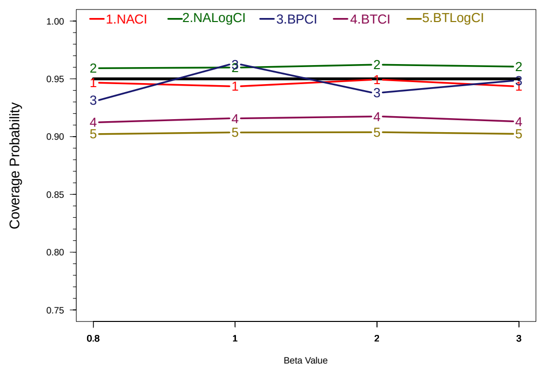
- For the NHPP estimator in Figure 4, there is again a parallel pattern across the values of β for all six plots.



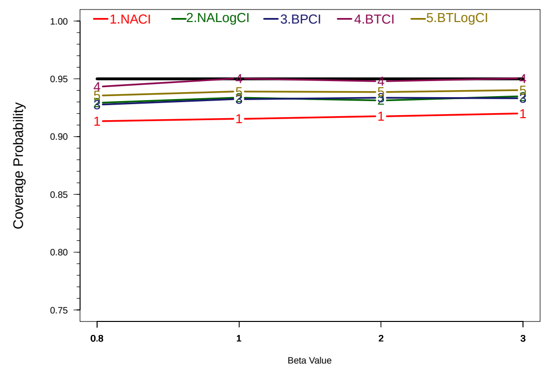
(a) $n = 10$ and $\mathrm{E}(r) = 10$



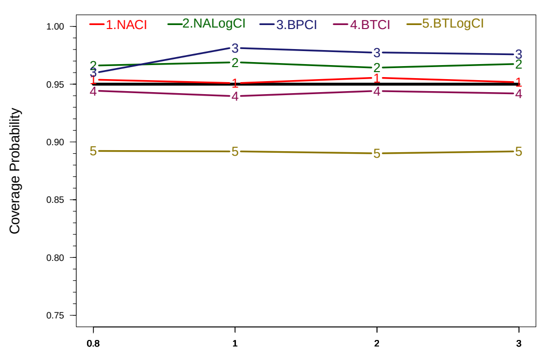
(b) $n = 10$ and $\mathrm{E}(r) = 20$



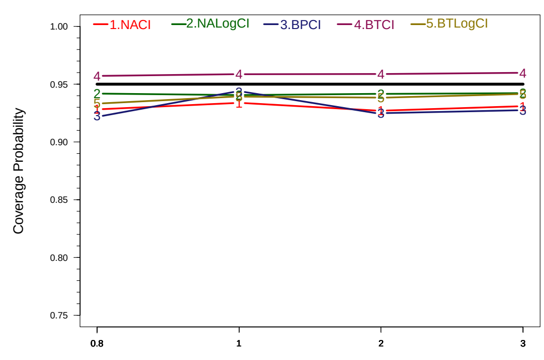
(c) $n = 20$ and $\mathrm{E}(r) = 10$



(d) $n = 20$ and $\mathrm{E}(r) = 20$

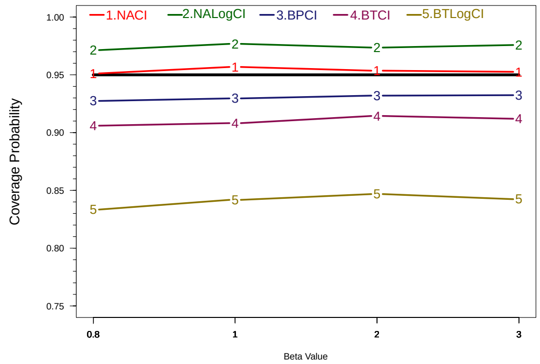


(e) $n = 50$ and $\mathrm{E}(r) = 10$

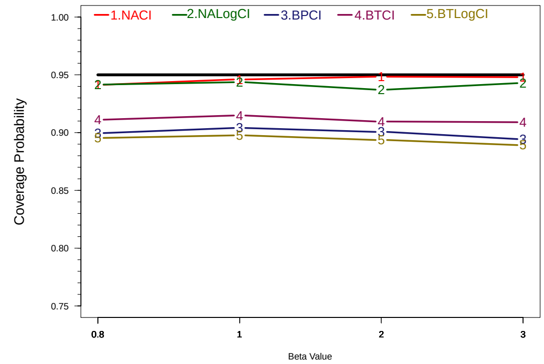


(f) $n = 50$ and $\mathrm{E}(r) = 20$

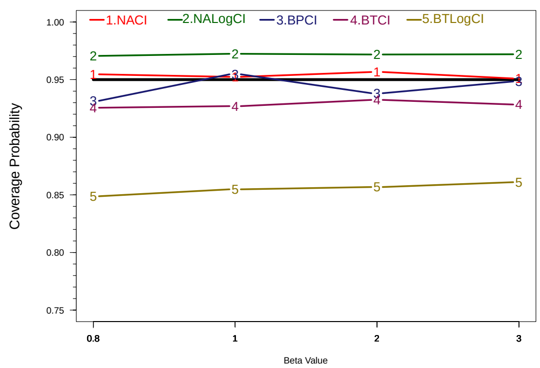
Figure 1: Comparison of 4 β Values: NP Estimator for the Complete Data



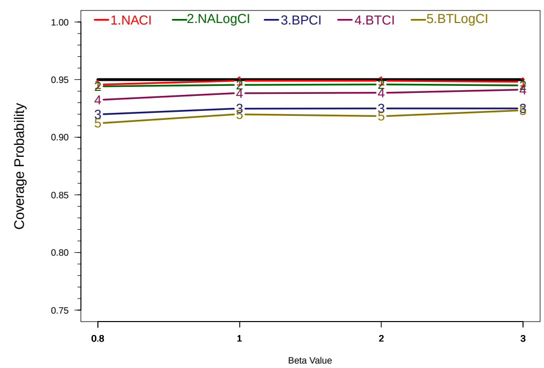
(a) $n = 10$ and $\mathrm{E}(r) = 10$



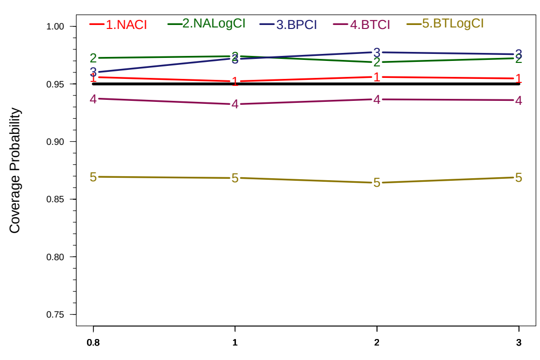
(b) $n = 10$ and $\mathrm{E}(r) = 20$



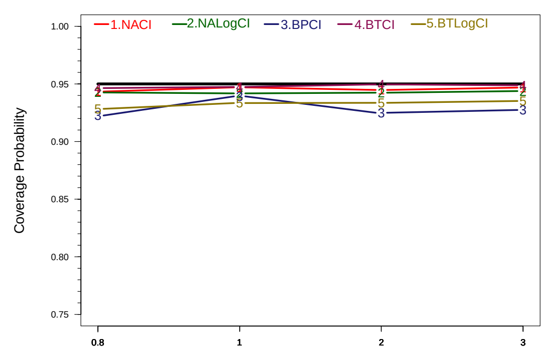
(c) $n = 20$ and $\mathrm{E}(r) = 10$



(d) $n = 20$ and $\mathrm{E}(r) = 20$

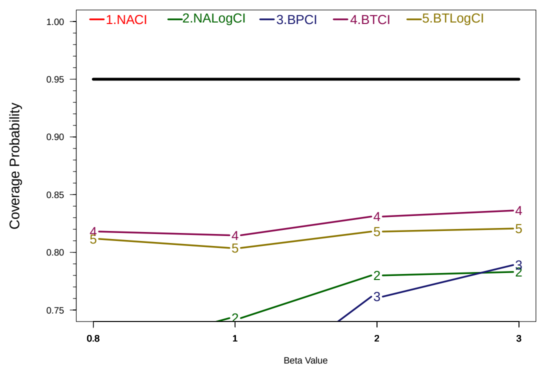


(e) $n = 50$ and $\mathrm{E}(r) = 10$

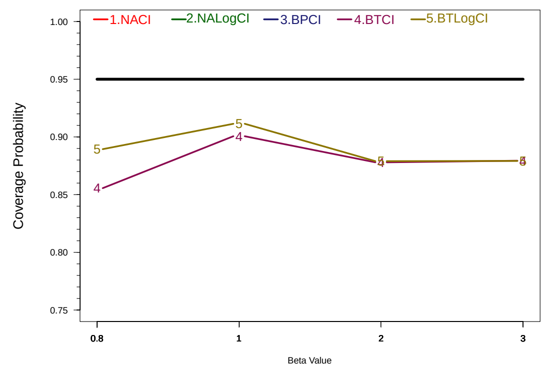


(f) $n = 50$ and $\mathrm{E}(r) = 20$

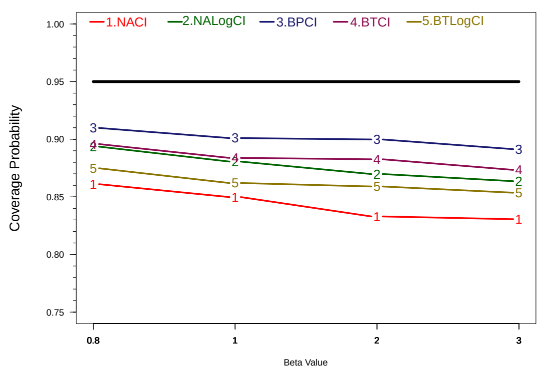
Figure 2: Comparison of 4 β Values: NHPP Estimator for the Complete Data



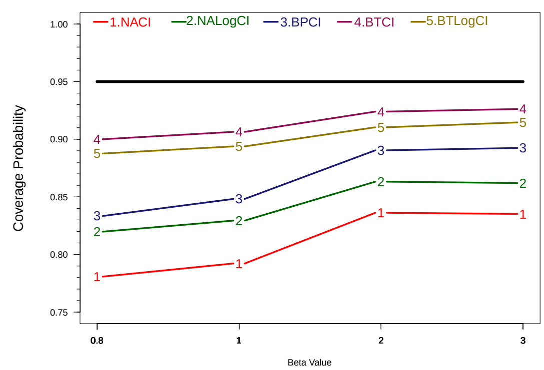
(a) $n = 10$ and $\mathrm{E}(r) = 10$



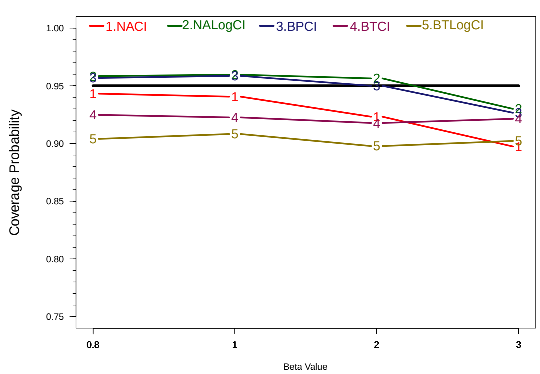
(b) $n = 10$ and $\mathrm{E}(r) = 20$



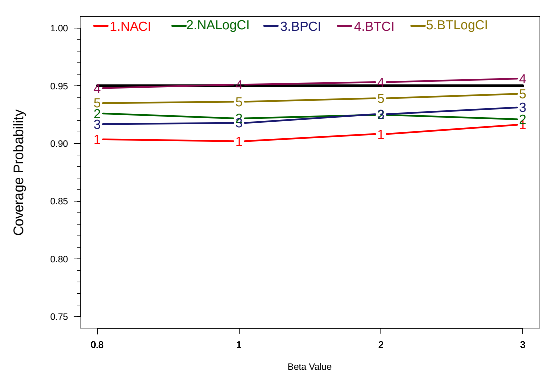
(c) $n = 20$ and $\mathrm{E}(r) = 10$



(d) $n = 20$ and $\mathrm{E}(r) = 20$

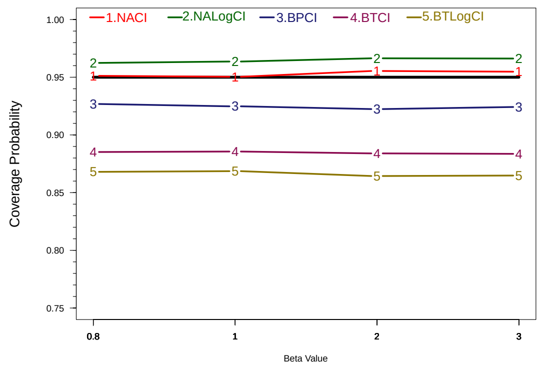


(e) $n = 50$ and $\mathrm{E}(r) = 10$

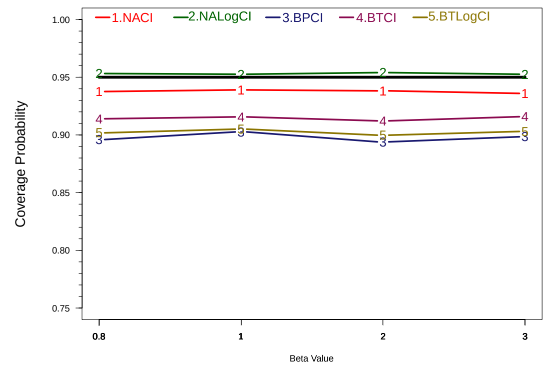


(f) $n = 50$ and $\mathrm{E}(r) = 20$

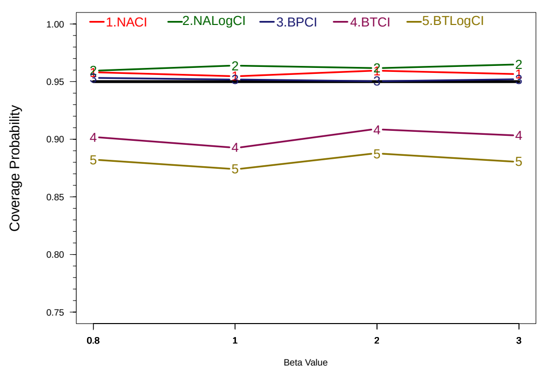
Figure 3: Comparison of 4 β Values: NP Estimator for the Window2 Data



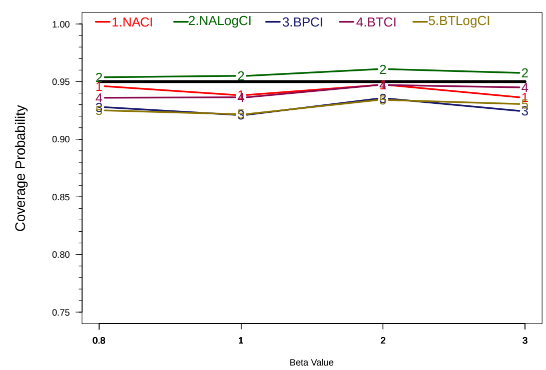
(a) $n = 10$ and $\mathrm{E}(r) = 10$



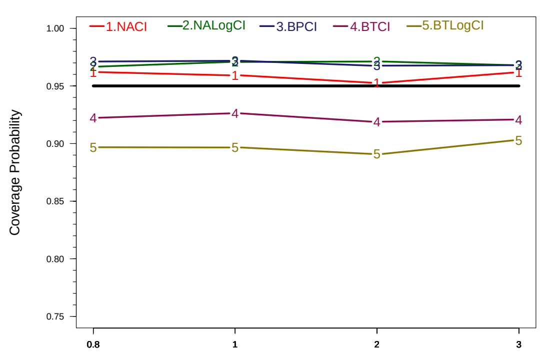
(b) $n = 10$ and $\mathrm{E}(r) = 20$



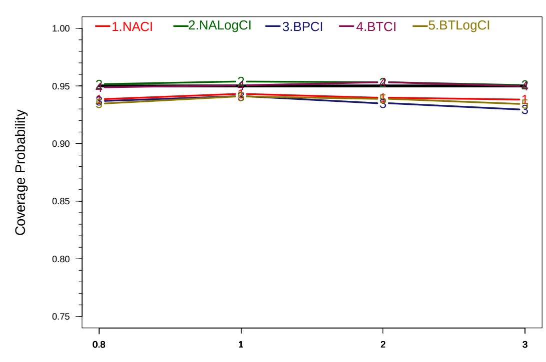
(c) $n = 20$ and $\mathrm{E}(r) = 10$



(d) $n = 20$ and $\mathrm{E}(r) = 20$

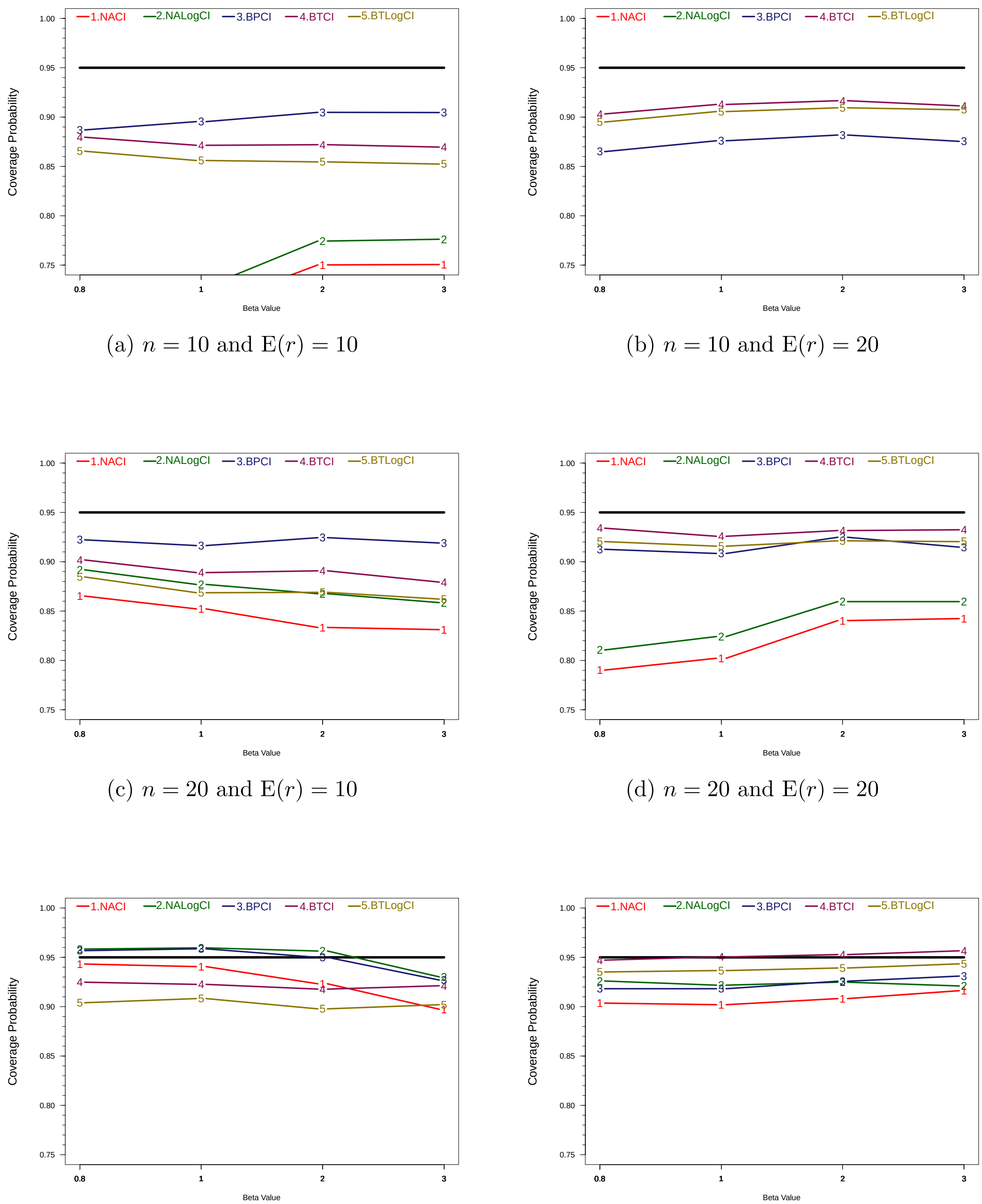


(e) $n = 50$ and $\mathrm{E}(r) = 10$



(f) $n = 50$ and $\mathrm{E}(r) = 20$

Figure 4: Comparison of 4 β Values: NHPP Estimator for the Window2 Data



(a) $n = 10$ and $\mathrm{E}(r) = 10$

(b) $n = 10$ and $\mathrm{E}(r) = 20$

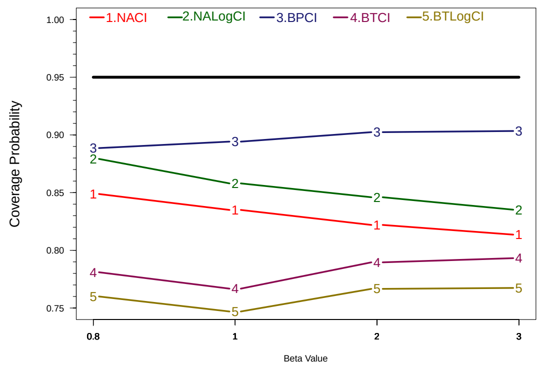
(c) $n = 20$ and $\mathrm{E}(r) = 10$

(d) $n = 20$ and $\mathrm{E}(r) = 20$

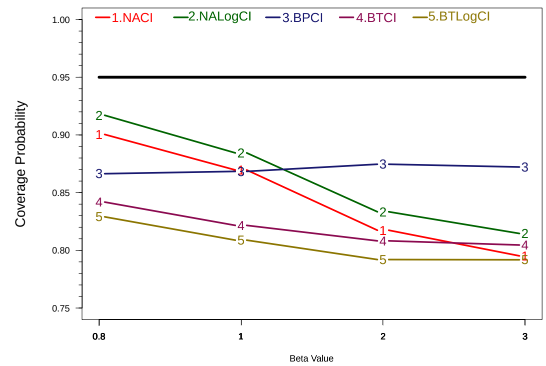
(e) $n = 50$ and $\mathrm{E}(r) = 10$

(f) $n = 50$ and $\mathrm{E}(r) = 20$

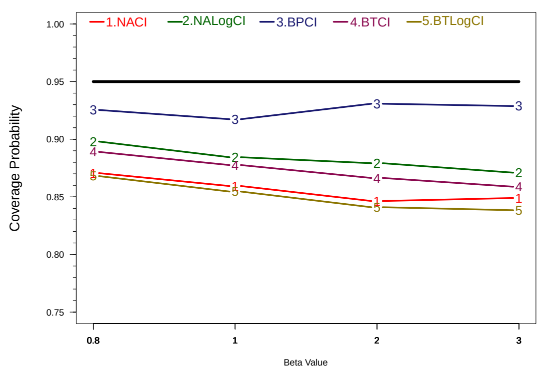
Figure 5: Comparison of 4 β Values: Local Hybrid Estimator for the Window2 Data



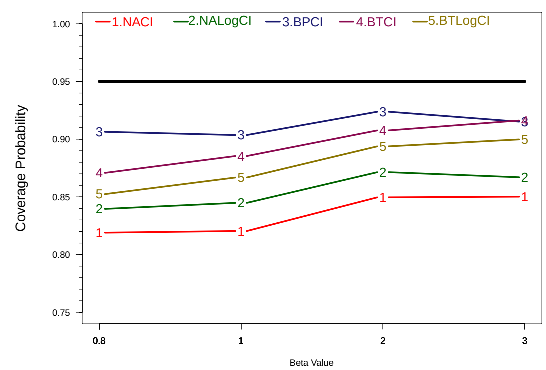
(a) $n = 10$ and $\mathrm{E}(r) = 10$



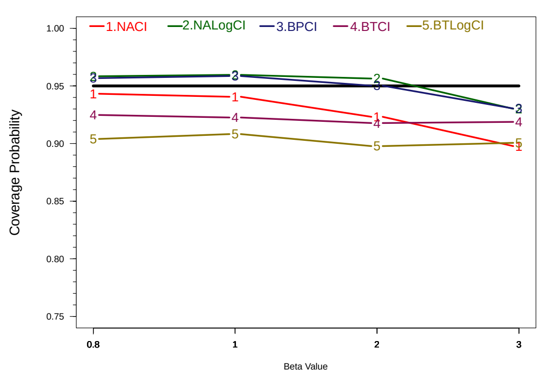
(b) $n = 10$ and $\mathrm{E}(r) = 20$



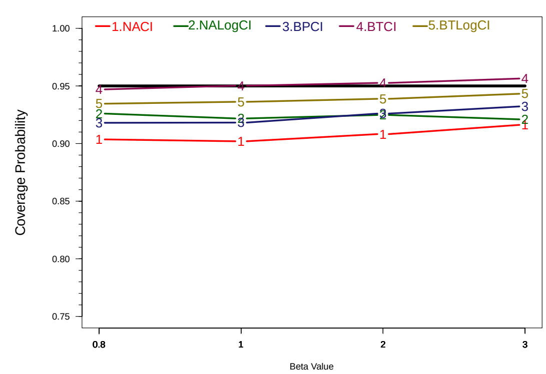
(c) $n = 20$ and $\mathrm{E}(r) = 10$



(d) $n = 20$ and $\mathrm{E}(r) = 20$



(e) $n = 50$ and $\mathrm{E}(r) = 10$



(f) $n = 50$ and $\mathrm{E}(r) = 20$

Figure 6: Comparison of 4 β Values: NHPP Hybrid Estimator for the Window2 Data

Table 1: RSSZ and RSSONE for Window2 Data with $n = 10$, E(r) = 10, and Number of Simulations = 5000

	Average			Percentage of time relative to t_{endobs}		
β	t_{endobs}	RSSZ	RSSONE	RSSZ	RSSONE	RSSZ+RSSONE
0.8	6.83	1.29	2.01	18.9%	29.5%	48.3%
1	4.30	0.60	1.07	14.0%	24.9%	38.8%
2	1.95	0.15	0.30	7.5%	15.4%	22.9%
3	1.53	0.09	0.20	5.6%	12.9%	18.5%

- For the NP estimator and the two hybrid estimators, in Figures 3, 5 and 6 respectively, CP values are close to parallel across the β values when $n = 20$ and E(r) = 10 or when $n = 50$. When $n = 10$, or $n = 20$ and E(r) = 20, there is noticeable increasing or decreasing trend of the CP values as β changes, indicating that the asymptotic approximations are far from adequate.

One important difference between the Complete data and the Window2 data is that it is possible to have a high percentage of time with RSSZ and RSSONE for the Window2 data. For the Complete data, there is no RSSZ or RSSONE. As observed in Figures 1 and 2, the performances of the CI procedures do not depend on the β values when there is no RSSZ or RSSONE, for the Complete data case. For the Window2 data, the percentage of time with RSSZ and RSSONE is zero or negligible when $n = 50$, and is somewhat close to 5% for all four β values when $n = 20$ and E(r) = 10. This percentage is, however, relatively large and different among the four β values when $n = 10$, as well as $n = 20$ and E(r) = 20. Therefore, the existence of RSSZ and RSSONE is the main reason that CP depends strongly on β in Figures 3, 5 and 6. For example, Table 1 shows, for E(r) = 10 and $n = 10$, the percentage of time with RSSZ and RSSONE increases as β decreases.

Among the four estimators, the NHPP estimator is robust to having a high percentage of time with RSSZ, as observed by the parallel patterns in Figure 4. This nice property is, however, based on the condition that the assumed NHPP model adequately describes the true process.

Based on the results of the Complete data and Window2 data in this section, the performances of the CI procedures for each of the four MCF estimators depend more on the existence of RSSZ and RSSONE rather than on the value of β. On the other hand, we carried out simulations on all four β values, but did not find special patterns that depend solely on the value of β. Therefore, in the subsequent sections on the CI performances for each of the MCF estimators, we will focus on the simulation results for $\beta = 1$, and discuss the impact of n, $E(r)$, and the existence of RSSZ and RSSONE in more detail.

6 PERFORMANCES OF THE CI PROCEDURES BASED ON THE NP ESTIMATOR

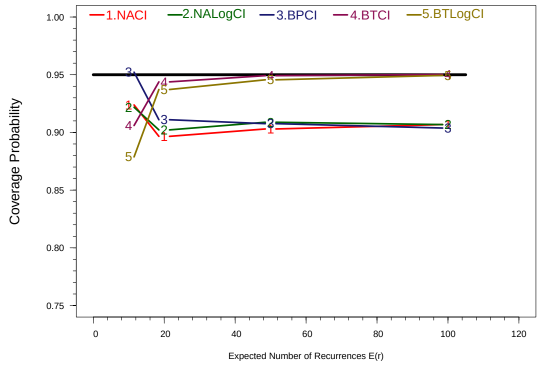
This section summarizes the simulation results for the NP estimator, and makes some recommendations. Figure 7 shows six plots for the NP estimator, with factor $E(r)$ as the x-axis in each plot. From top down, n increases from 10 to 20 and then to 50, and plots on the left are for Complete data while those on the right are for Window2 data.

6.1 Comparison of Results from Complete Data

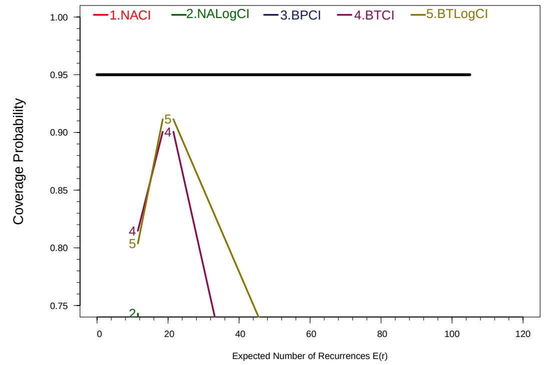
The main observations from the three plots for the Complete data results, shown in Figures 7 (a), (c) and (e), are:

- For each n value, CP values stabilize when $E(r) \geq 20$, and over the parallel part, the Boott and LBoott procedures consistently have CP values that are very close to the nominal value.
- When $E(r) \geq 20$, with the increase of n, the performances of the NORMA procedure, the LNORMA procedure, and the BootP procedure improve, with CP values getting closer to the nominal value.
- When $E(r) = 10$, increasing the value of n does little to help decrease the differences among the five CI procedures, and which CI procedures to use depends on the value of n. When $n = 10$, the BootP procedure has the best CP, while the other four CI procedures have CP values that are less than the nominal 0.95 value. When $n \geq 20$, the NORMA procedure generates the best CP, followed by the LNORMA procedure (with a CP that is larger than the nominal 0.95). For all the n values, the LBoott procedure severely underestimates the CP, an indication that when $E(r)$ is small and thus there is smaller variation in the bootstrap re-samples, the log-transformation is a less appropriate choice.

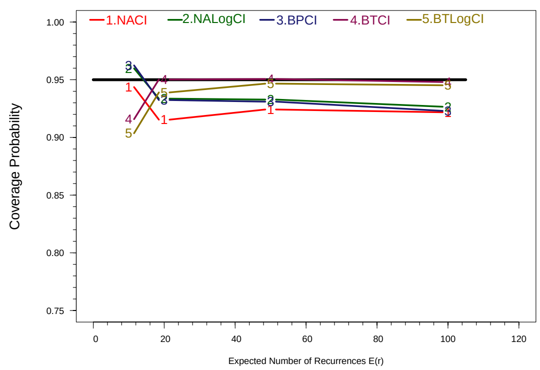
For the two normal approximation procedures, Appendix A.2 shows that, when $n = 10$, the CP values are close to 0.9 for E(r) = 10, 20, 50, and 100, even though the nominal CP is 0.95. Therefore, it is somewhat surprising to observe CP at about 0.92 when E(r) = 10 in Figure 7 (a). The reason for this behavior is that the CP values in our simulation study are conditional on $\sum_{i=1}^{n} X_i \geq 5$. The impact of this condition is very small for E(r) ≥ 20; therefore the conditional CP values are close to 0.9 as well when E(r) ≥ 20. However, when E(r) = 10, $\Pr\left(\sum_{i=1}^{n} X_i \leq 4\right) = 0.0293$ is not negligible. There are 12 distinct scenarios that have $\sum_{i=1}^{n} X_i \leq 4$ when $n = 10$, such as none of the 10 units have a recurrence, as well as 1 unit has 4 recurrences and 9 units have no recurrences.



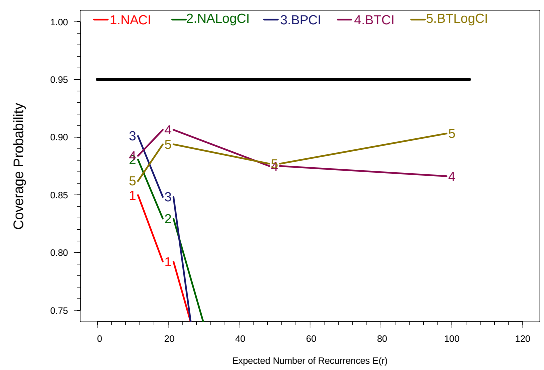
(a) Complete Data with $n = 10$



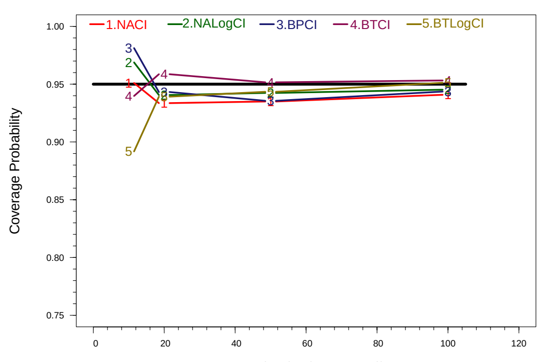
(b) Window2 Data with $n = 10$



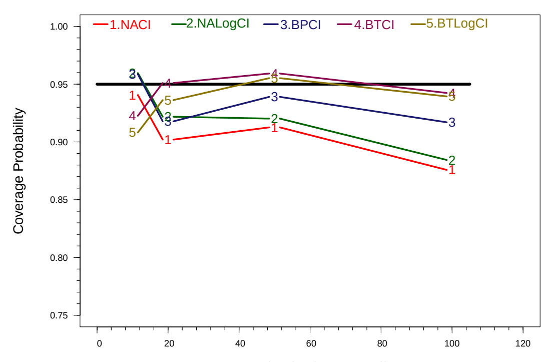
(c) Complete Data with $n = 20$



(d) Window2 Data with $n = 20$



(e) Complete Data with $n = 50$



(f) Window2 Data with $n = 50$

Figure 7: CP Plots for NP Estimator with $\beta = 1$

Among these 12 scenarios, there is only one case for which the CI captures the true MCF – one unit has 4 recurrences and 9 units have no recurrences, and the corresponding probability for this case is 1.89×10^{-5}. As a result, the conditional CP given $\sum_{i=1}^{n} X_i \leq 4$ is $1.89 \times 10^{-5}/0.0293 = 6.45 \times 10^{-4}$. With Pr(MCF $\in$ CI) = 0.89415 from Table 3 in Appendix A.2, we have

$$\Pr(\text{MCF} \in \text{CI}|\sum_{i=1}^{n} X_i \geq 5) = \frac{\Pr(\text{MCF} \in \text{CI}) - \Pr(\text{MCF} \in \text{CI}|\sum_{i=1}^{n} X_i \leq 4)}{1 - \Pr(\sum_{i=1}^{n} X_i \leq 4)} = 0.92.$$

This explains why in Figure 7 (a), the CP for the NORMA and LNORMA procedures are about 0.92 when $\text{E}(r) = 10$, while about 0.90 when $\text{E}(r) \geq 20$.

6.2 Comparison of Results from Window2 Data

Compared to the Complete data, outcomes for the Window2 data are more complicated, because the amount of time with RSSZ and RSSONE depends on the simulation experiment factor levels. Table 2 shows the average length in time of RSSZ and RSSONE, as well as the corresponding ratios to t_{endobs}, for $n = 10$, 20, and 50. Even though the NP estimator is biased when there are times with RSSZ, we still summarize below all the results that we have observed, because some CI procedures are relatively robust to the existence of RSSZ and RSSONE, and sometimes the NP estimator might be the only applicable option to use. Recommendations, including the scenario that the NP estimator should not be used, are outlined in Section 6.3.

- Consider Figure 7 (b), when $n = 10$. All five CI procedures perform poorly when the amount of time with RSSZ and RSSONE is large, especially the NORMA, the LNORMA, and the BootP procedures. These three CI procedures generate CP values that are below 0.75 even when $\text{E}(r) = 10$, and the CP values deteriorate fast to 0 when $\text{E}(r) \geq 50$, and thus no lines are shown for these three procedures in the plot. By comparison, the Boott and the LBoott CI procedures have better performance, and their CP values, even though still well below the nominal value at 0.95, are much higher, especially when $\text{E}(r) \leq 20$. Table 2 shows that, when $n = 10$, the percentage of time with RSSZ ranges from 14% to 74.3%, and as $\text{E}(r)$ increases, the time with RSSZ and RSSONE becomes more dominant.

- Consider Figure 7 (d), when $n = 20$. The Boott and the LBoott CI procedures show strong robustness to the existence of RSSZ and RSSONE, with all CP values above 0.85. The NORMA, the LNORMA, and the BootP procedures still perform poorly when $\text{E}(r) \geq 20$, but the BootP procedure has CP value that is closest to the nominal value when $\text{E}(r) = 10$. Table 2 shows that, when $n = 20$, the percentage of time with RSSZ ranges from 1.2% to 35.4%.

Table 2: RSSZ and RSSONE for Window2 Data with $\beta = 1$ and Number of Simulations = 5000

		Average			Percentage of time relative to t_{endobs}		
n	E(r)	t_{endobs}	RSSZ	RSSONE	RSSZ	RSSONE	RSSZ+RSSONE
10	10	4.30	0.60	1.07	14.0%	24.9%	38.8%
10	20	14.58	4.31	4.98	29.6%	34.2%	63.7%
10	50	81.42	46.65	24.42	57.3%	30.0%	87.3%
10	100	312.91	232.63	65.19	74.3%	20.8%	95.2%
20	10	1.40	0.02	0.05	1.2%	3.8%	5.0%
20	20	4.30	0.14	0.42	3.1%	9.6%	12.8%
20	50	21.97	3.37	5.92	15.3%	26.9%	42.3%
20	100	81.42	28.80	27.65	35.4%	34.0%	69.3%
50	10	0.38	0.00	0.00	0.0%	0.0%	0.0%
50	20	1.00	0.00	0.00	0.0%	0.1%	0.1%
50	50	4.30	0.00	0.02	0.1%	0.4%	0.5%
50	100	14.58	0.09	0.44	0.6%	3.0%	3.6%

- Consider Figure 7 (f), when $n = 50$. All CI procedures have performances that are much closer to those of the Complete data. This agrees with what is shown in Table 2 that the amount of time with RSSZ is zero or negligible when $n = 50$. The CP values, however, show some drop from E(r) = 50 to 100, and one contributing factor is the existences of RSSONE and intervals with relatively small size of the risk set. As shown in Table 2, the percentage of time with RSSONE is 3% when E(r) = 100. As for the comparisons among the five CI procedures, when E(r) = 10, the NORMA, the LNORMA, and the BootP procedures have CP that are closer to the nominal value, while the Boott and the LBoott CI procedures have CP values that are farther away and lower than the nominal value. When E(r) $\geq$ 20, however, the Boott and the LBoott CI procedures have CP values that are very close to the nominal value, while the other three procedures are not as good.

6.3 Recommendations for the NP Estimator

Based on the simulation results, we recommend the following for the NP estimator. For Complete data, or window data with very small percentage of time as RSSZ and RSSONE, we recommend:

1. When E(r) = 10 and $n \geq 20$ ($n \geq 50$ for Window2 data), one should use the NORMA procedure, because of simplicity in calculation and because it has a CP that is close to the nominal value.

2. When both E(r) and n are small, the BootP procedure provides a CP that is closest to the nominal value.

3. When E(r) ≥ 20, one can use either the Boott or the LBoott procedure to have a CP that is close to the nominal value.

For window data with non-negligible amounts of time with RSSZ and RSSONE, we recommend:

1. When the amount of time with RSSZ and RSSONE is more than 70% of t_{endobs}, the NP estimator should not be used, because it is seriously biased, and none of the CI procedures behave well.

2. When the amount of time with RSSZ and RSSONE is non-negligible but less dominating:

   a When E(r) is 20 or more, the Boott and the LBoott procedures should be used because of their robustness to the existence of RSSZ and RSSONE and because their CP values are close to the nominal value.

   b When E(r) is small, and the amount of time with RSSZ and RSSONE is about 5%, as in the case of $n = 20$, the BootP procedure is recommended, because its CP values are relatively close to the nominal values.

   c When E(r) is small, yet the amount of time with RSSZ and RSSONE is relatively large, as about 40% for the case of $n = 10$, the Boott and the LBoott procedures are recommended because their CP values are relatively close to the nominal values.

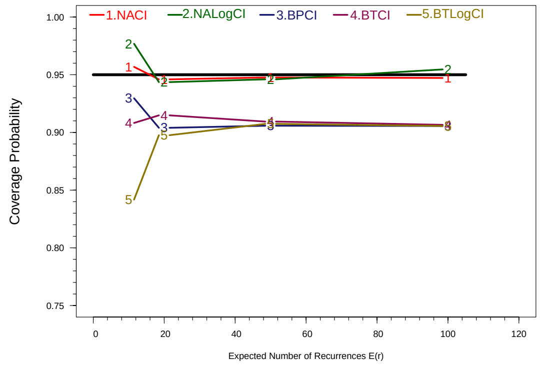
7 PERFORMANCES OF THE CI PROCEDURES BASED ON THE NHPP ESTIMATOR

This section summarizes the simulation results for the power law NHPP estimator, and makes some recommendations. Figure 8 shows six plots for the NHPP estimator, similar to those in Figure 7 for the NP estimator.

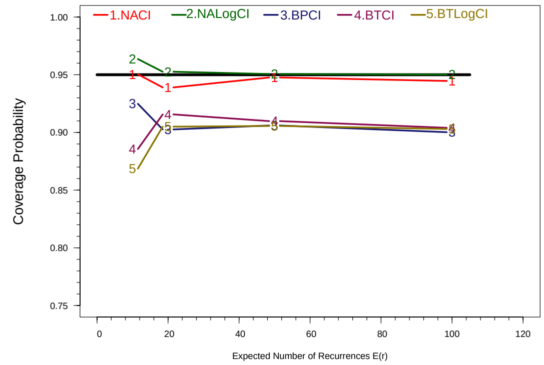
7.1 Comparison of Results from Complete Data and Window2 Data

From the plots in Figure 8, we have the following primary observations for the NHPP estimator:

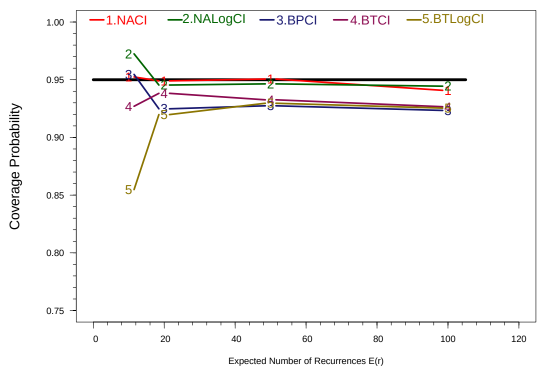
- The two plots on the same row, left for the Complete data and right for the Window2 data, are very similar. This indicates that the performances of the CI procedures for the NHPP estimator is not strongly affected by the existence of RSSZ intervals.



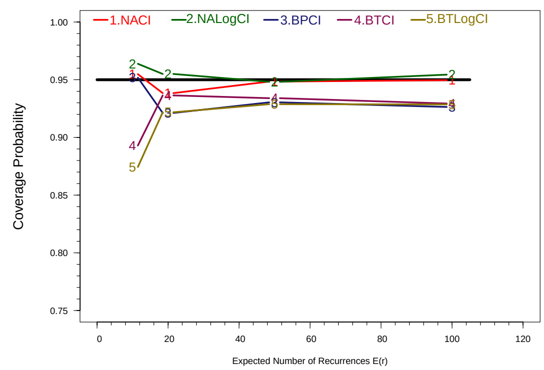
(a) Complete Data with $n = 10$



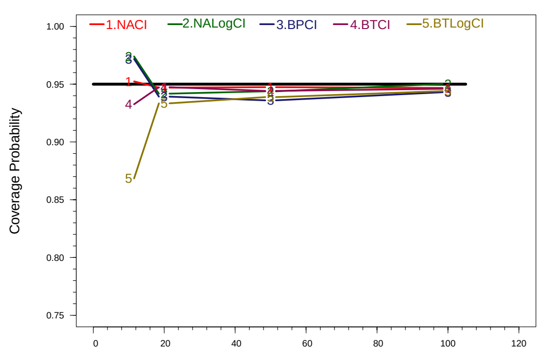
(b) Window2 Data with $n = 10$



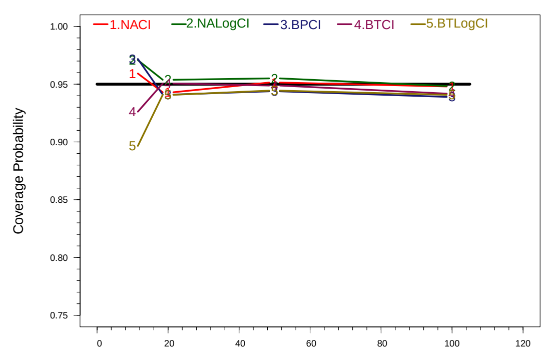
(c) Complete Data with $n = 20$



(d) Window2 Data with $n = 20$



(e) Complete Data with $n = 50$



(f) Window2 Data with $n = 50$

Figure 8: CP Plots for the Power Law NHPP Estimator with $\beta = 1$

- For each n value, CP values level off when $E(r) \geq 20$.
- When $E(r) \geq 20$, the two normal-approximation-based procedures have CP values that are close to the nominal value. The three bootstrap methods, on the other side, are not as good when n is small because of the discreteness in the re-sampling procedure. However, when n increases, the performances of the three bootstrap methods improve, and they are close to the two normal-approximation-based procedures when $n = 50$.
- When $E(r) = 10$, the NORMA procedure has a CP that is consistently close to the nominal value. The LNORMA procedure has a CP that is consistently larger than the nominal 0.95, while the Boott procedure has a CP that is smaller than the nominal value. The LBoott procedure has a CP that is much smaller than the nominal value. When n increases, however, there is only a small improvement in CI performances for the LNORMA and LBoott procedures, and this differs from the observation when $E(r) \geq 20$ that the CP values become closer to the nominal value of 0.95 as n increases.

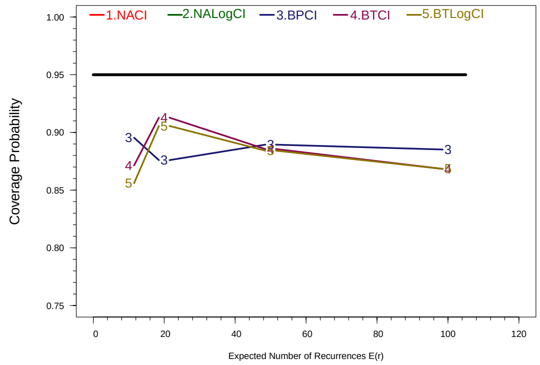
The simulation results suggest that the adequacy of the CI procedures for the NHPP estimator depends mainly on E(r), and is relatively independent of n and tends to be robust to the existence of RSSZ and RSSONE. When n is small, however, say 20 or smaller, the bootstrap procedures can have CP values that are importantly less than the nominal value.

7.2 Recommendations for the NHPP Estimator

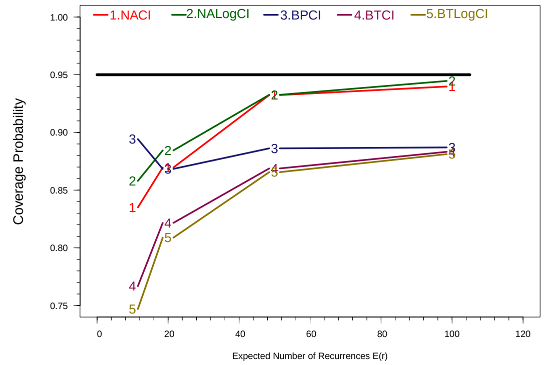
When the assumed NHPP model provides an adequate description of the underlying stochastic process, the NORMA procedure would be a good choice, because of simplicity in calculation and because the CP is close to the nominal value.

8 PERFORMANCES OF THE CI PROCEDURES BASED ON THE HYBRID ESTIMATORS FROM WINDOW2 DATA

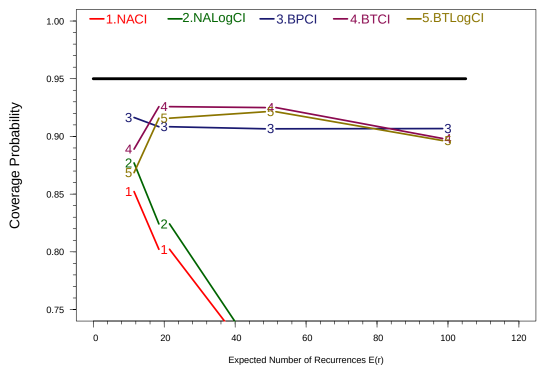
This section summarizes the simulation results for the local hybrid estimator and the power law NHPP hybrid estimator for the Window2 data, and makes some recommendations. Figure 9 shows six plots for these two hybrid estimators, with factor $E(r)$ as the x-axis in each plot. From the top down, n increases from 10 to 20 and then to 50, and the plots on the left are for the local hybrid estimator while those on the right are for the power law NHPP hybrid estimator.



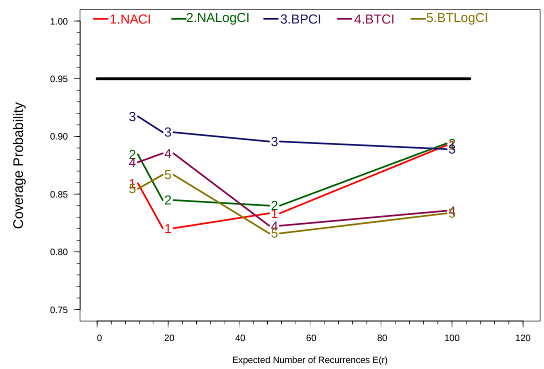
(a) Local Hybrid Estimator with $n = 10$



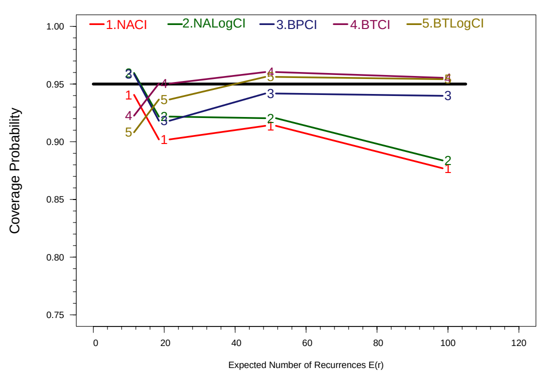
(b) NHPP Hybrid Estimator with $n = 10$



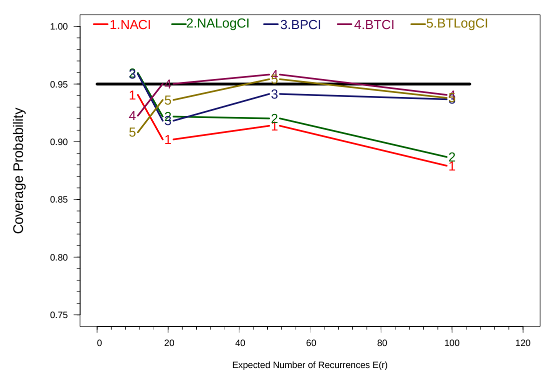
(c) Local Hybrid Estimator with $n = 20$



(d) NHPP Hybrid Estimator with $n = 20$



(e) Local Hybrid Estimator with $n = 50$



(f) NHPP Hybrid Estimator with $n = 50$

Figure 9: CP Plots for the Hybrid Estimators for the Window2 Data with $\beta = 1$

8.1 Comparison of Results – the Local Hybrid Estimator and the Power Law NHPP Hybrid Estimator

As shown in Section 2.5, the local hybrid estimator and the power law NHPP hybrid estimator have the same $\bar{d}.(t)$, the nonparametric estimator of the increase in the MCF from the RSSP intervals. Thus, when there is no RSSZ (or the amount of time with RSSZ is small), the two hybrid estimators are identical (or close) to the NP estimator. For example, there is hardly any differences observed in Figure 9 (e) for the local hybrid estimator and Figure 9 (f) for the NHPP hybrid estimator, because when $n = 50$, Table 2 shows that the percentage of time with RSSZ is less than 1%. Therefore, meaningful comparisons come from the simulation settings where the time with RSSZ as a percentage of t_{endobs} is not negligible.

When $n = 10$, Table 2 shows that the average amount of time of RSSZ plus RSSONE as a percentage of t_{endobs} increases from 38.8% to 95.2% when $\mathrm{E}(r)$ increases from 10 to 100. Correspondingly, the differences among the two hybrid estimators also increase, as shown in Figure 9 (a) and (b). For the local hybrid estimator, all five CI procedures, especially the three bootstrap-based procedures, have better performances than those of the NP estimator, which are partially observable by comparing Figure 9 (a) and Figure 7 (b). The CP values for the NORMA and LNORMA procedures are all below 0.75 and fall to 0.49 when $\mathrm{E}(r) = 100$, and thus are not shown in Figure 9 (a). One major difficulty for the normal-approximation-based procedures is that the normal distribution assumption is hard to satisfy when the size of the risk set is small. Compared to the local hybrid estimator, the five CI procedures for the power law NHPP hybrid estimator have much better performances: the CP values increase and get closer to the nominal value as $\mathrm{E}(r)$ increases. The increasing dominance of RSSZ and RSSONE does not impair the performances of the CI procedures. The main reason is that the increases in the MCF from the RSSZ intervals are estimated from the assumed power law NHPP model, which is the correct model form that is used to generate the simulation data, and which, as observed in Section 7, is robust to high percentage of time with RSSZ and RSSONE.

When holding $\mathrm{E}(r)$ at the same values and increasing n from 10 to 20, the proportion of time with RSSZ, as well as time with RSSZ and RSSONE, gets smaller. Thus we see better performances of the CI procedures with this increase in the values of n. For the local hybrid estimator, despite the different levels of improvement, the two normal-approximation-based procedures are still far from being desirable, as shown in Figure 9 (c), and the CP values fall to 0.57 at $\mathrm{E}(r) = 100$. For the power law NHPP hybrid estimator, however, the BootP procedure is the only procedure that has CP values closer to the nominal value for all four $\mathrm{E}(r)$ values, while the other four CI procedures show different amounts of deterioration at $\mathrm{E}(r) = 50$ and $\mathrm{E}(r) = 100$, as observed by comparing Figure 9 (b) and (d). One contributing factor for this behavior is the relatively large proportion of time with RSSONE. Even though the percentage of time with RSSZ and RSSONE gets smaller when n increases from 10 to 20, Table 2 shows that the percentage of time with RSSONE only drops slightly from 30% to 26.9% at E(r) = 50, and even inceases from 20.8% to 34% at E(r) = 100. Covariances among the recurrence times over periods of time with RSSONE are not estimable, and also covariances between RSSONE intervals and other time intervals are not estimable. Therefore, the existence of a relatively large percentage of time with RSSONE has a negative effect on the estimation of the variances of the MCF estimators. The BootP procedure only depends on the point estimates of the MCF, and larger values of n ensure more randomized bootstrap re-samples, and thus its performance improved.

8.2 Recommendations for the Hybrid Estimators

Among our simulation experiment factor levels, when $n \geq 50$, the percentage of time with RSSZ is zero or close to zero. In these cases, the hybrid estimators are the same as or close to the NP estimator. Therefore, our observations and the recommendations below for the hybrid MCF estimators mainly apply to data with relatively small n (e.g., ≤ 20) where there is a non-negligible amount of time with RSSZ.

1. When the percentage of time with RSSZ and RSSONE is large, the differences in CI performances between the local hybrid estimator and those of the NHPP hybrid estimator are large. When there is no strong indication of model deviation from the NHPP model, the NHPP hybrid estimator is preferred, because the variance estimate tends to be smaller in the bias-variance tradeoff; otherwise, the local hybrid estimator is preferred.

2. For the NHPP hybrid estimator, assuming that the NHPP model provides a relatively good description of the underlying point process,
   - When the percentage of time with RSSZ is very large (e.g., 57.3% when $E(r) = 50$ and $n = 10$ or 74.3% when $E(r) = 100$ and $n = 10$), the two normal-approximation-based procedures are preferred, because their CP values are closer to the nominal value, and they are simple to construct. For these cases, the NHPP hybrid estimator is close to the NHPP estimator, and as shown in Section 7.2, the NORMA procedure is recommended for the NHPP estimator.
   - When the percentage of time with RSSZ is moderate (e.g., $\leq$ 15.3% for the cases of $E(r) <= 50$ and $n = 20$, as well as $E(r) = 10$ and $n = 10$), the BootP procedure is preferred, because its CP is the closest to the nominal value. The relative robustness

to the existence of RSSONE is one contributing factor for the comparatively better performance of the BootP procedure.

3. For the local hybrid estimator,
   - When E(r) is small, for example around 10, the BootP procedure is preferred, because its CP is the closest to the nominal value.
   - When E(r) $\geq$ 20, the three bootstrap-based procedures generate similar CP results and these results are substantially better than those of the two normal-approximation-based procedures. Among the three bootstrap-based procedures, the BootP procedure might be preferred because it is, by comparison, simpler to construct.

9 CONCLUDING REMARKS AND AREAS FOR FURTHER RESEARCH

In our simulation study, we have shown the performances of five of the commonly used CI procedures. All five CI procedures perform well when both the number of units n and the expected number of observed recurrences $E(r)$ are reasonably large, and there is no time with RSSZ and RSSONE or the percentage of time with RSSZ and RSSONE is small. However, when n is small, or $E(r)$ is small, or the percentage of time with RSSZ and RSSONE is not negligible, choices among the MCF estimators and the CI procedures can lead to very different results. We have made some recommendations based on the simulation results. There are, however, some important areas for further research.

- The five CI procedures in our simulation study are relatively simple and easy to construct. With improvements in the computing power, more complicated bootstrap CI procedures could be included in the comparison study. The BC_a method and the ABC method described in Efron and Tibshirani (1993) are both second-order accurate and transformation respecting, and their finite-sample properties would be of interest.

- In addition to normal-approximation and bootstrap-based CI procedures, likelihood-based methods could be used to construct CIs for the NHPP model. Similarly, empirical likelihood methods could be used for the NP estimators to construct CIs. These likelihood and empirical likelihood methods could also be extended to the hybrid estimators. In particular, methods described in Owen (2001) could be extended for these cases.

- It is sometimes possible to obtain better estimation of the MCF by including the explanatory variables in the model. Some such models are described in Lawless and Nadeau (1995), and

Cook and Lawless (2007). Confidence interval procedures for such models could also be studied.

ACKNOWLEDGMENTS

This material is based upon work supported, in part, by the National Science Foundation under Grant No. 0421916.

APPENDIX

A PROPERTIES OF THE COMPLETE DATA SIMULATED FROM THE POWER LAW NHPP MODEL

A.1 Distribution of the Number of Recurrences

Let X_1, X_2, ..., X_n be the number of observed recurrences for the n units from a Complete data set (i.e., no gaps), simulated from a power law NHPP model with a given value of $\mathrm{E}(r)$. Then X_1, X_2, ..., X_n are independent and identically distributed (*iid*) from a Poisson $(\mathrm{E}(r)/n)$ distribution. From this, the probability mass function (pmf) of $(X_1, X_2, ..., X_n)$ is

$$\Pr\left(X_1 = x_1, X_2 = x_2, ..., X_n = x_n\right) = \prod_{i=1}^{n} \left\{ \exp\left(-\mathrm{E}(r)/n\right) \times \frac{[\mathrm{E}(r)/n]^{x_i}}{x_i!} \right\}. \tag{4}$$

Because the sum of independent Poisson random variables is also a Poisson random variable, $\sum_{i=1}^{n} X_i$ follows a Poisson distribution with $\lambda = \sum_{i=1}^{n} \lambda_i = \mathrm{E}(r)$, where λ_i is the parameter for X_i. Therefore, the probability of a simulated data set with x recurrences is

$$\Pr\left(\sum_{i=1}^{n} X_i = x\right) = \exp\left(-\mathrm{E}(r)\right) \times \frac{[\mathrm{E}(r)]^x}{x!}, \tag{5}$$

which is a function of $\mathrm{E}(r)$ and x only. For the four values of $\mathrm{E}(r)$ in our simulation study, 10, 20, 50, and 100, the values of $\Pr\left(\sum_{i=1}^{n} X_i \leq 4\right)$ are 0.0293, 1.69×10^{-5}, 5.45×10^{-17}, and 1.61×10^{-37}, respectively.

Note that (4) and (5) depend only on $\mathrm{E}(r)$ and n, and thus results in this subsection apply not only to the power law NHPP model, but also to NHPP model with other forms of the recurrence functions.

A.2 NP MCF Estimator

One interesting property of the NP method for the Complete data case is that both $\widehat{MCF}_{NP}(t_{endobs})$ and $\widehat{\mathrm{Var}}[\widehat{MCF}_{NP}(t_{endobs})]$, which are needed to construct the normal-approximation-based CIs, are functions of X_1, X_2, ..., X_n and n, and do not depend on the observed times of the recurrences. Then, for a given set of $(x_1, x_2, ..., x_n)$, we can construct the NORMA and the LNORMA CIs and find out whether the CIs capture the true MCF at $\mathrm{E}(r)/n$. By (4), we can also calculate $\Pr(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$. Therefore, if we can identify that part of the sample space of $(x_1, x_2, ..., x_n)$ for which the constructed CIs capture the true MCF, and then the sum of the corresponding probabilities is the actual CP of the CI procedure. Because x_i can take the value of any non-negative integer, for $i = 1, ..., n$, the complete sample space for $(x_1, x_2, ..., x_n)$ is countable, but infinite. As a result, it is not possible to identify the complete set of $(x_1, x_2, ..., x_n)$ that we want to identify to calculate the actual CP. However, a close approximation of the actual CP can be obtained by using the following approach.

1. Select a finite number of combinations of $(x_1, x_2, ..., x_n)$ from the complete sample space such that the sum of the corresponding probabilities is close to 1.

2. For each set of values of $(x_1, x_2, ..., x_n)$ in the selected sample space, calculate $\widehat{MCF}_{NP}(t_{endobs})$ and $\widehat{\mathrm{Var}}[\widehat{MCF}_{NP}(t_{endobs})]$, and construct a CI. If the CI captures the true MCF, then add the corresponding probability of the set to CP.

The complement of the selected sample space provides a bound on the error for the actual CP. Because the probability of getting large values of x_i is very small, the sample space we chose is $[0, k] \times [0, k] \times ... \times [0, k]$, where k is an integer large enough to keep the error bound small.

We now derive the expressions for $\widehat{MCF}_{NP}(t_{endobs})$ and $\widehat{\mathrm{Var}}[\widehat{MCF}_{NP}(t_{endobs})]$. For Complete data, having a constant risk set size of n, (1) leads to $\widehat{MCF}_{NP}(t_{endobs}) = \sum_{i=1}^{n} X_i/n$. Lawless and Nadeau (1995) present the following formula (here with our notation), for data with all units having the same end-of-observation times,

$$\widehat{\mathrm{Var}}[\widehat{MCF}_{NP}(t_{endobs})] = \frac{1}{n^2} \sum_{i=1}^{n} \left[X_i - \frac{\sum_{i=1}^{n} X_i}{n} \right]^2 .$$

Table 3 shows results for the approximate CP, which is the sum of the probabilities for the sets of $(x_1, x_2, ..., x_n)$ that have the calculated CI capturing the true MCF, as well as the error bound, for the NP estimator with $n = 10$. Two CI procedures are used, the NORMA and the LNORMA. The approximate probability of not capturing the true MCF equal $(1 - \text{Prob. in CI} - \text{error bound})$. Note that, the nominal coverage probability is 0.95, yet the probabilities with true MCF in the CI

Table 3: Probability in CI and Error Bound for Complete Data at $n = 10$

	NP Estimator			Power Law NHPP Estimator		
E$(r)/n$	NA	LOGNA	Error Bound	NA	LOGNA	Error Bound
1	0.89415	0.91038	1.11×10^{-6}	0.92573	0.96262	7.98×10^{-8}
2	0.89250	0.90384	2.07×10^{-6}	0.94751	0.94428	4.83×10^{-9}
5	0.89983	0.90550	1.40×10^{-5}	0.94878	0.94440	1.57×10^{-10}
10	0.90224	0.90510	7.65×10^{-6}	0.94503	0.94912	7.08×10^{-11}

for the two procedures are well below this nominal value. This shows that the normal approximation procedures do not work well for the NP estimator when the sample size is small, and increasing the expected number of observed recurrences for each unit does not help improve the CP. The LNORMA CI procedure performs better than the NORMA procedure, because the number of recurrences is non-negative.

A.3 The Power Law NHPP Estimator

Rigdon and Basu (2000, Section 5.4) present the likelihood function of n independent systems from the same power law NHPP model, and point out that when all n systems are observed from time zero to the same end-of-observation time t_{endobs}, the Complete data scenario in our simulations, there is an explicit solution for estimating the two model parameters,

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} \sum_{j=1}^{X_i} \log(t_{endobs}/t_{ij})} \tag{6}$$

$$\hat{\eta} = \frac{n^{1/\hat{\beta}} t_{endobs}}{(\sum_{i=1}^{n} X_i)^{1/\hat{\beta}}}, \tag{7}$$

where t_{ij} denotes the j^{th} recurrence time for the i^{th} system. With the power law recurrence rate function (3), (2) can be simplified to

$$\widehat{MCF}_{NHPP}(t_{endobs}) = (t_{endobs}/\hat{\eta})^{\hat{\beta}} = \sum_{i=1}^{n} X_i/n.$$

This shows that $\widehat{MCF}_{NHPP}(t_{endobs}) = \widehat{MCF}_{NP}(t_{endobs})$ for the Complete data case and these estimators only depend on the number of observed recurrences and the number of observational units.

It can also be shown that,

$$\widehat{\text{Var}}[\widehat{MCF}_{NHPP}(t_{endobs})] = \sum_{i=1}^{n} X_i/n^2 = \widehat{MCF}_{NHPP}(t_{endobs})/n.$$

This result is not surprising, because for the Poisson distribution, the variance and the mean are the same, and here we have n units to estimate the variance. However, this simple form is only available for the Complete data case when all units are observed to t_{endobs}. The main steps to derive this result are listed below.

1. Derive the Hessian matrix by taking second derivatives of the likelihood function with β and η.
2. Obtain variance-covariance matrix of $\hat{\beta}$ and $\hat{\eta}$ by evaluating the inverse negative Hessian matrix at the MLEs $\hat{\beta}$ and $\hat{\eta}$.
3. Apply the delta method to get $\widehat{\mathrm{Var}}[\widehat{MCF}_{NHPP}(t_{endobs})]$.

We used the same approach described in A.2 for the NP estimator, and obtained the approximate CP values and the error bound for the power law NHPP estimator in Table 3. Because both $\widehat{MCF}_{NHPP}(t_{endobs})$ and $\widehat{\mathrm{Var}}[\widehat{MCF}_{NHPP}(t_{endobs})]$ depend only on the total number of observed recurrences and the number of units in the data, we used (5) instead of (4) to simplify the calculation. Compared to the NP estimator, the NORMA and the LNORMA procedures for the power law NHPP estimator have CPs that are very close to the nominal value at 0.95.

REFERENCES

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, New York: Springer-Verlag.

Cook, R. J., and Lawless, J. F. (2007), *The Statistical Analysis of Recurrent Events*, New York: Springer-Verlag.

Cox, D. R., and Lewis, P. A. W. (1966), *The Statistical Analysis of Series of Events*, New York: Wiley.

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

Jeng, S. L., and Meeker, W. Q. (2000), "Comparisons of Approximate Confidence Interval Procedures for Type I Censored Data," *Technometrics*, 42, 135-148.

Lawless, J. F., and Nadeau, C. (1995), "Some Simple Robust Methods for the Analysis of Recurrent Events," *Technometrics*, 37, 158-168.

Meeker, W. Q., and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, New York: Wiley.

Nelson, W. B. (1988), "Graphical Analysis of System Repair Data," *Journal of Quality Technology*, 20, 24-35.

Nelson, W. B. (1995), "Confidence Limits for Recurrence Data: Applied to Cost or Number of Product Repairs," *Technometrics*, 37, 147-157.

Nelson, W. B. (2003), *Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications*, Philadelphia: ASA-SIAM.

Owen, A. B. (2001), *Empirical Likelihood*, New York: Chapman & Hall/CRC.

Rigdon, S. E., and Basu, A. P. (2000), *Statistical Methods for the Reliability of Repairable Systems*, New York: Wiley.

Zuo, J., Meeker, W. Q., and Wu, H. (2008), "Analysis of Window-Observation Recurrence Data," *Technometrics*, 50, 128-143.