System Unavailability Analysis Based on Window-Observed Recurrent Event Data

Yili Hong
Department of Statistics, Virginia Tech

Joint work with
M. Li and B. Osborn

NCTS Industrial Statistics Research Group Seminar
Hsinchu, Taiwan
December 14, 2012

Outline

- Background
- Data Overview (Re-Scaled)
- Statistical Modeling and Inference
  - Nonhomogeneous Poisson Process
  - Multiple Event Types
  - Bath-tub Intensity Function
  - Truncated Lognormal Event Duration
- Unavailability Analysis
- Conclusions and Areas for Future Research

Background

- Many service industries, such as airlines, telecommunication, and railways, require a high level of system availability in order to be competitive.

- Service providers often offer unavailability guarantees to their customers.

- That is, the unavailable percentage during a certain time period (e.g., a year) is guaranteed to be below a threshold that is specified in the service contract.

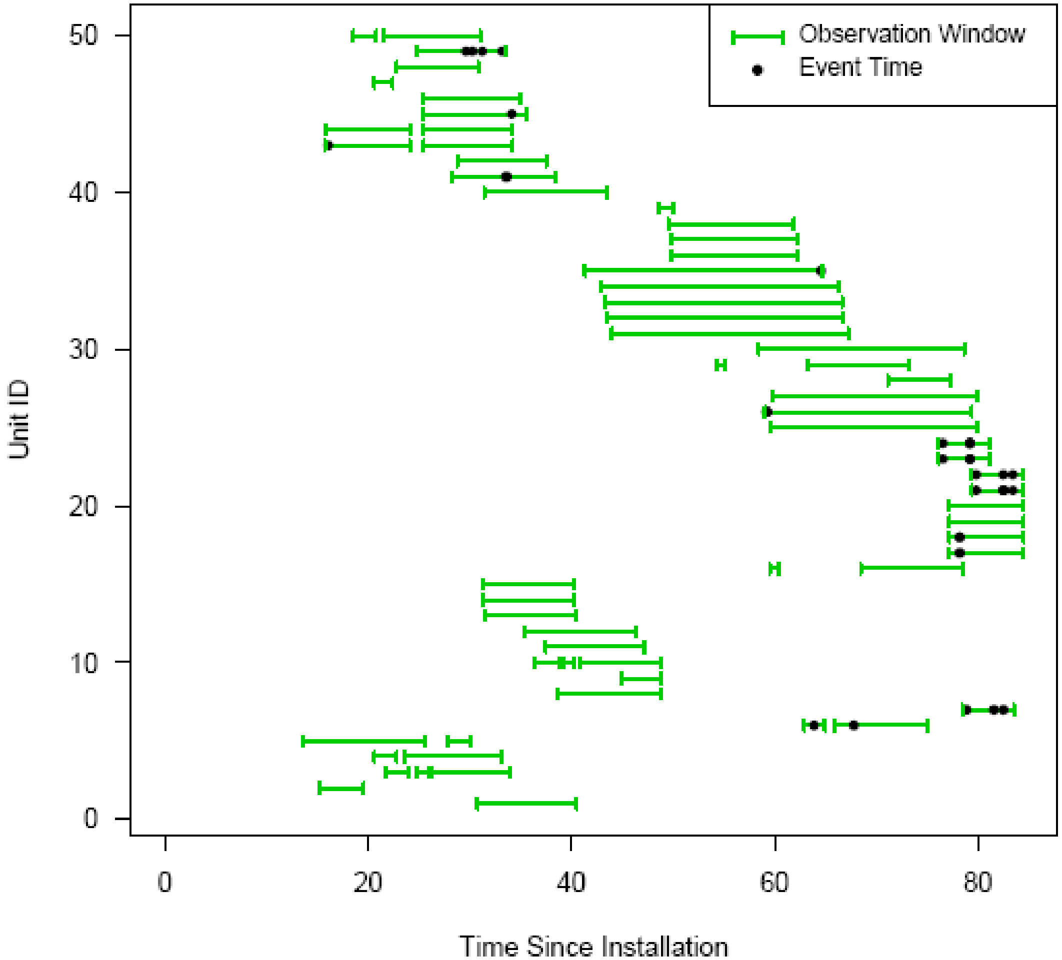
Motivation for System Unavailability Analysis

- The service providers naturally want to offer a low threshold on the unavailability guarantee, to make the business competitive.
- Unfulfillment of a service contract with customers, however, will lead to cost to the service provider.
- Before the service provider decides what to offer, a system unavailability analysis is needed to minimize the risk of unfulfillment.
- Good to be conservative. A conservative metric for system unavailability is desirable.

Application

- Subjects: A fleet of heavy duty industry equipment (System A).
- Goal: Do an unavailability analysis based on historical data.
- System A is a repairable system.
- Good to have: continuous track of all repairs and maintenance during the life of each unit.
- Reality: Only have records for certain time intervals and no information beyond such intervals.

System A Data Overview (Re-Scaled)



- Observation windows
- Event times
- Multiple event types
- Event durations

Analysis Strategy

- Statistics: obtain maximum information from the limited data source.

- Modeling: extend NHPP for repairable system + distribution of event duration

- Unavailability: Define a conservative metric for system unavailability and use simulation to estimate it

Statistical Modeling and Inference

- ❖ Nonhomogeneous Poisson Process
- ❖ Multiple Event Types
- ❖ Bath-tub Intensity Function
- ❖ Truncated Lognormal Event Durations

Nonhomogeneous Poisson Process (NHPP)

- For a repairable system, the NHPP is usually used:
  - Poisson Process has constant intensity
  - With non-constant intensity function
  - The intensity is a function of time

- The intensity function

$$\lambda(t;\boldsymbol{\theta}) = \lim_{\Delta \to 0} \frac{\Pr[N(t+\Delta) - N(t) = 1]}{\Delta}$$

- Cumulative intensity function

$$\Lambda(t;\boldsymbol{\theta}) = \int_0^t \lambda(s;\boldsymbol{\theta})\,ds$$

Power Law Relationship

❖ Event intensity:

$$\lambda_{\mathrm{pl}}(t;\beta,\eta) = \left(\frac{\beta}{\eta}\right)\left(\frac{t}{\eta}\right)^{\beta-1}$$

❖ Flexible:

$$\beta\begin{cases} <1 & \text{: Decreasing intensity (infant mortality)} \\ =1 & \text{: Constant intensity (failures by random)} \\ >1 & \text{: Increasing intensity (wear out failure)} \end{cases}$$

❖ Cumulative event intensity:

$$\Lambda_{\mathrm{pl}}(t;\beta,\eta) = \int_0^t \lambda_{\mathrm{pl}}(s;\beta,\eta)\,ds = \left(\frac{t}{\eta}\right)^{\beta}.$$

Multiple Event Types

❖ There are multiple types of events to cause a unit out of services.

❖ For example, for a car, engine failure, transmission failure, and flat tire are three types of event.

❖ The overall intensity for the system is:

$$\lambda(t;\boldsymbol{\theta}) = \sum_{k=1}^{K} \lambda_k(t;\boldsymbol{\theta})$$

❖ In System-A data set, there are two failure modes, i.e., k=2, and

$$\lambda(t;\boldsymbol{\theta}) = \sum_{k=1}^{2} \lambda_k(t;\boldsymbol{\theta})$$

The Bath-tub Intensity Function

❖ To fit a model with bathtub shape intensity function, we use a mixture of the PL intensity functions.

❖For a repairable system subject to three different process:

- the decreasing intensity $0 < \beta < 1$
- the increasing intensity $\beta > 1$
- the constant intensity $\beta = 1$

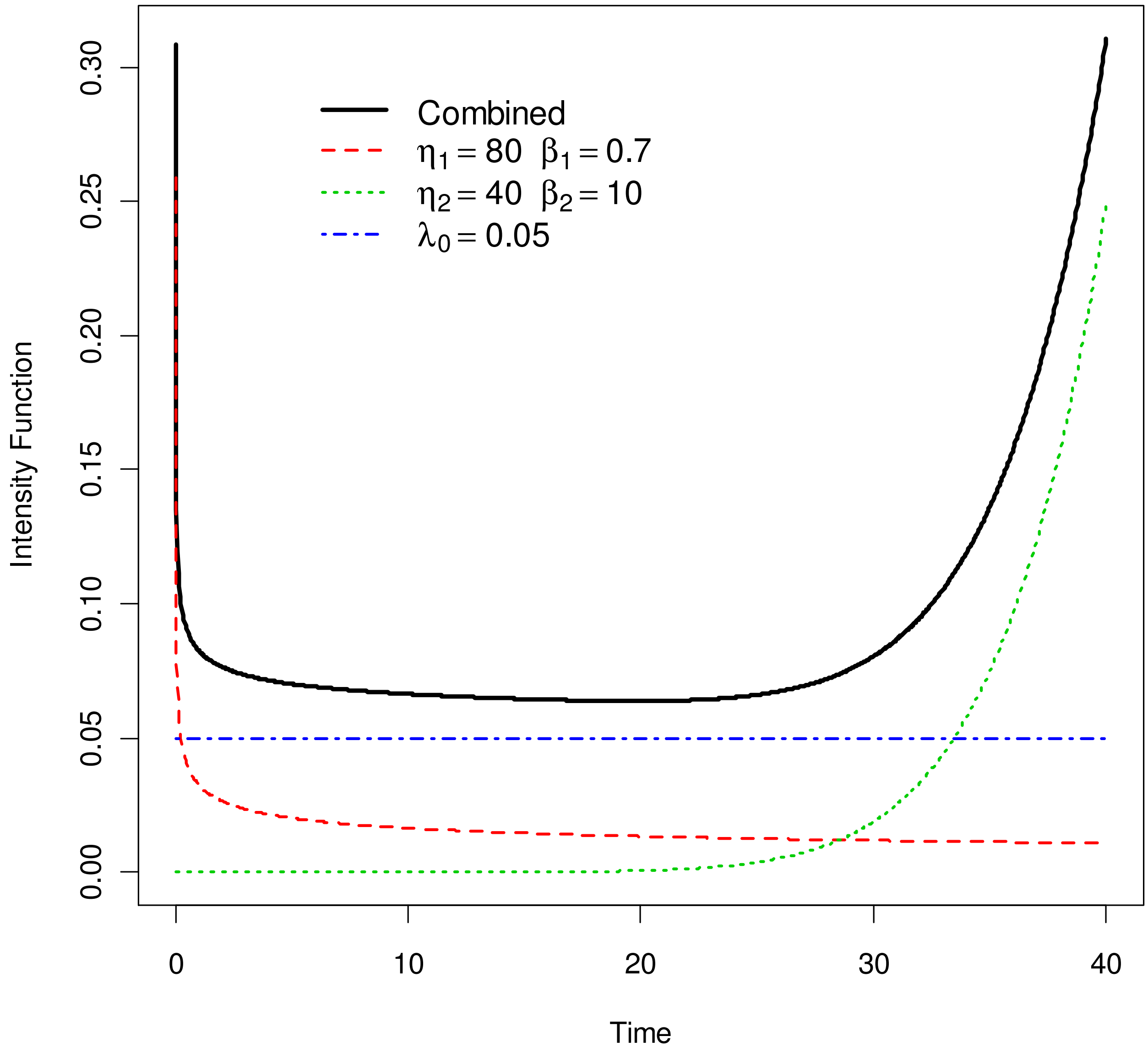
❖ Thus the combined event intensity is:

$$\lambda(t) = \lambda_0 + \frac{\beta_1}{\eta_1}\left(\frac{t}{\eta_1}\right)^{\beta_1 - 1} + \frac{\beta_2}{\eta_2}\left(\frac{t}{\eta_2}\right)^{\beta_2 - 1}$$

❖ The cumulative intensity function is given by:

$$\Lambda(t) = \lambda_0 t + \left(\frac{t}{\eta_1}\right)^{\beta_1} + \left(\frac{t}{\eta_2}\right)^{\beta_2}.$$

Example of Bath-tub Intensity Function



Maximum Likelihood Estimation

❖ Likelihood Function:

$$L(\boldsymbol{\theta} \mid DATA) = \prod_{i=1}^{n} \left\{ \prod_{k=1}^{K} \prod_{j=1}^{n_i} \left[\lambda_k (t_{ij}; \boldsymbol{\theta}_k)^{\mathbf{1}(\delta_{ij}=k)} \right] \times \exp\left(-\sum_{k=1}^{K} \sum_{l=1}^{m_i} [\Lambda_k(\tau_{il}^R; \boldsymbol{\theta}_k) - \Lambda_k(\tau_{il}^L; \boldsymbol{\theta}_k)] \right) \right\}$$

❖ The ML estimates are obtained by finding $\boldsymbol{\theta}$ that maximizes the likelihood function

❖ Local Information Matrix:

$$I_{\hat{\boldsymbol{\theta}}} = \left[-\frac{\partial^2 \log[L(\boldsymbol{\theta} \mid DATA)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

❖ The large-sample approximate variance–covariance matrix is $\Sigma_{\hat{\boldsymbol{\theta}}} = I_{\hat{\boldsymbol{\theta}}}^{-1}$, if the model assumptions hold well.

Robust Inference

- In reality, there might be some departures from the assumptions of the NHPP.
- The sandwich-type robust variance estimator is used to assure valid statistical inferences.
- The log-likelihood function is $l(\boldsymbol{\theta}) = \sum_{i=1}^{n} l_i(\boldsymbol{\theta})$ where

$$l_i(\boldsymbol{\theta}) = \sum_{k=1}^{K} \sum_{j=1}^{n_i} 1_{(\delta_{ij}=k)} \log\left[\lambda_k(t_{ij}; \boldsymbol{\theta}_k)\right] - \sum_{k=1}^{K} \sum_{l=1}^{m_i} \left[\Lambda_k(\tau_{il}^{R}; \boldsymbol{\theta}_k) - \Lambda_k(\tau_{il}^{L}; \boldsymbol{\theta}_k)\right]$$

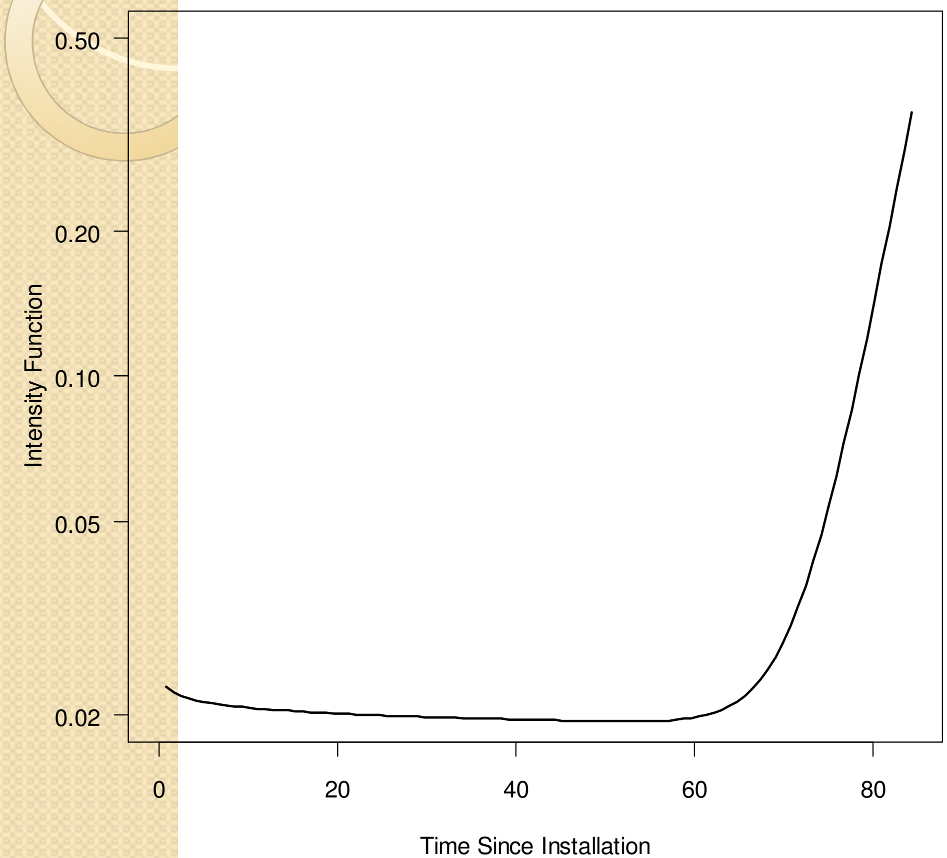
- Let $\mathrm{l}(\boldsymbol{\theta}) = [l_1(\boldsymbol{\theta}), \cdots, l_n(\boldsymbol{\theta})]'$, $U_n(\boldsymbol{\theta}) = \dfrac{\partial \mathrm{l}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, $J_n(\boldsymbol{\theta}) = U_n(\boldsymbol{\theta}) U_n(\boldsymbol{\theta})'$.
- The asymptotic theory provides

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \rightarrow \mathrm{N}\left[0, I(\boldsymbol{\theta}^*)^{-1} J(\boldsymbol{\theta}^*) I(\boldsymbol{\theta}^*)^{-1}\right], \text{ as } n \rightarrow \infty,$$

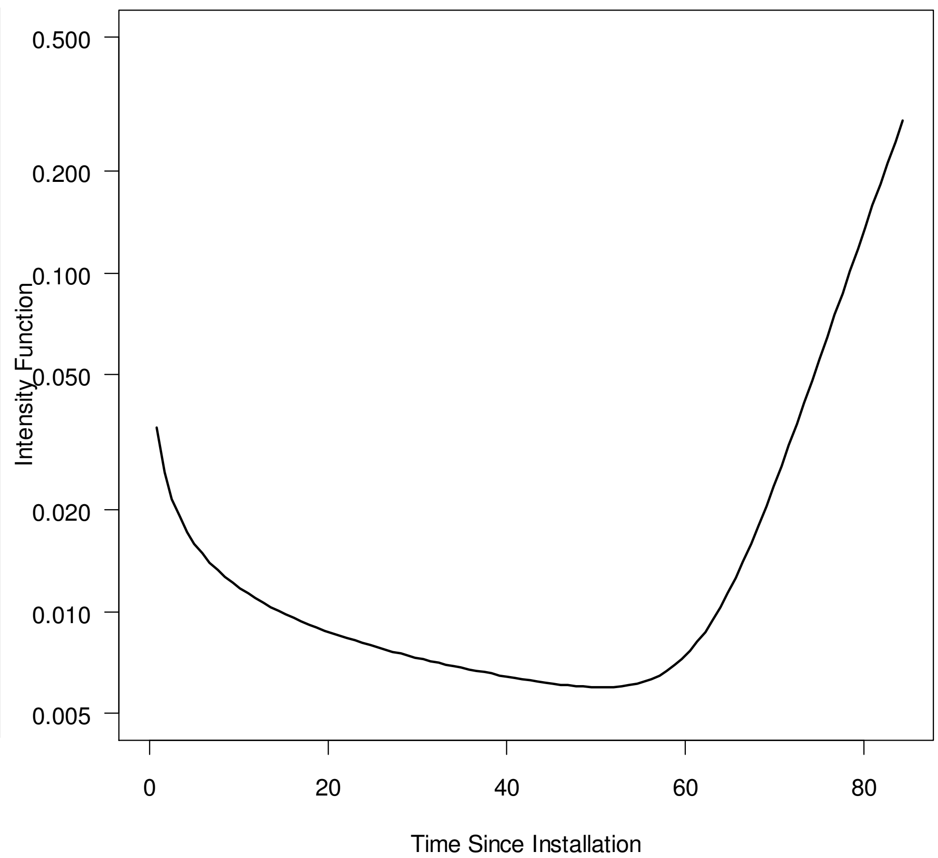
where

$$I(\boldsymbol{\theta}^*) = \lim_{n\to\infty} \frac{1}{n} I_n(\boldsymbol{\theta}^*), \quad \text{and} \quad J(\boldsymbol{\theta}^*) = \lim_{n\to\infty} \frac{1}{n} J_n(\boldsymbol{\theta}^*$$

Estimated Intensity Functions for System A

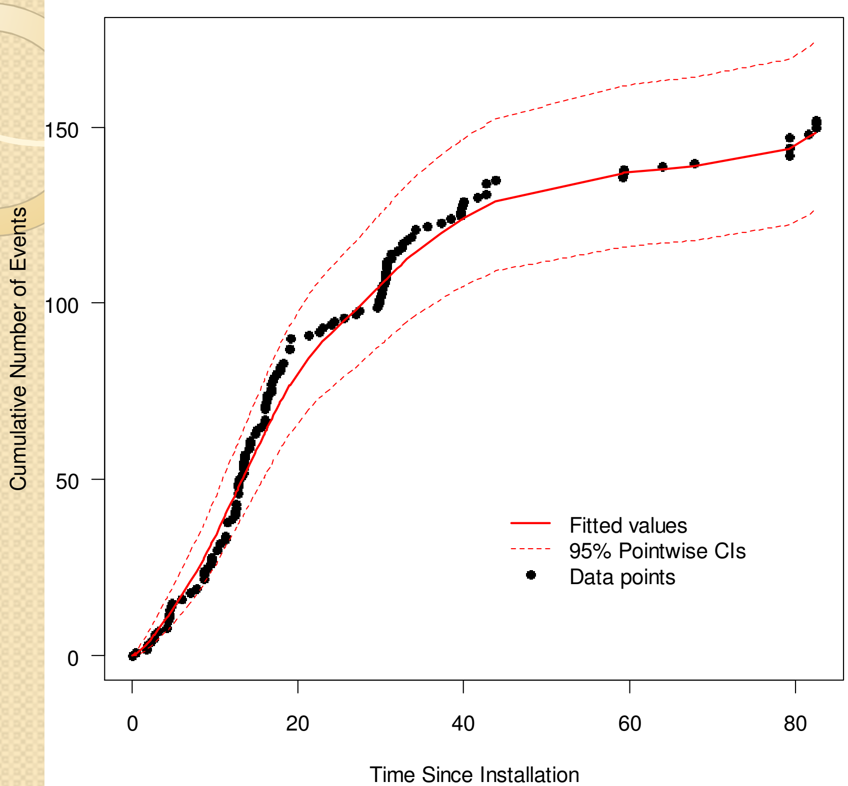


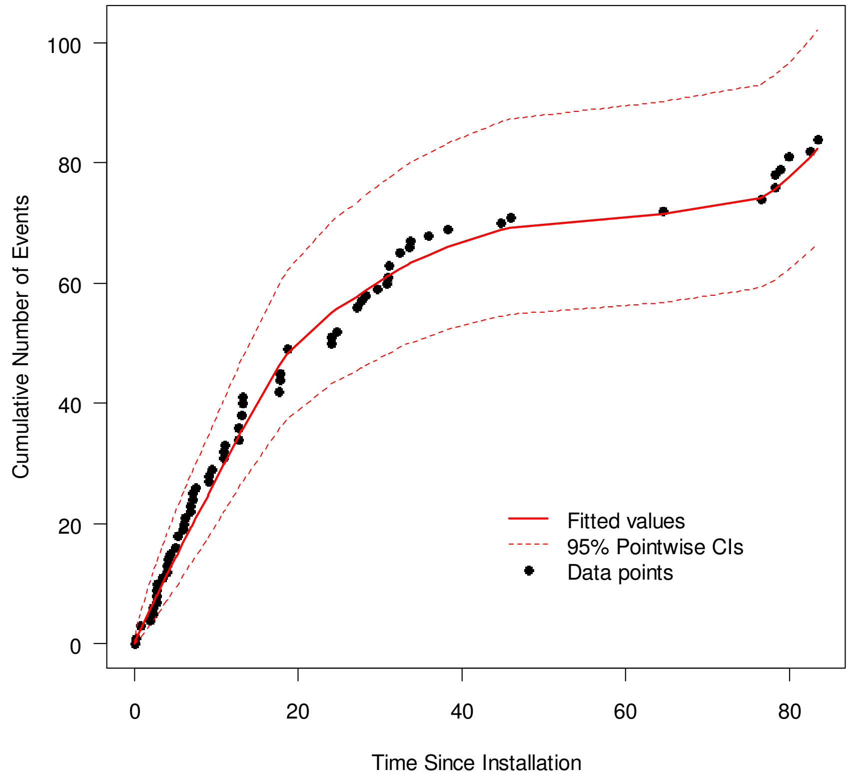
FM1



FM2

Access Model Fitting

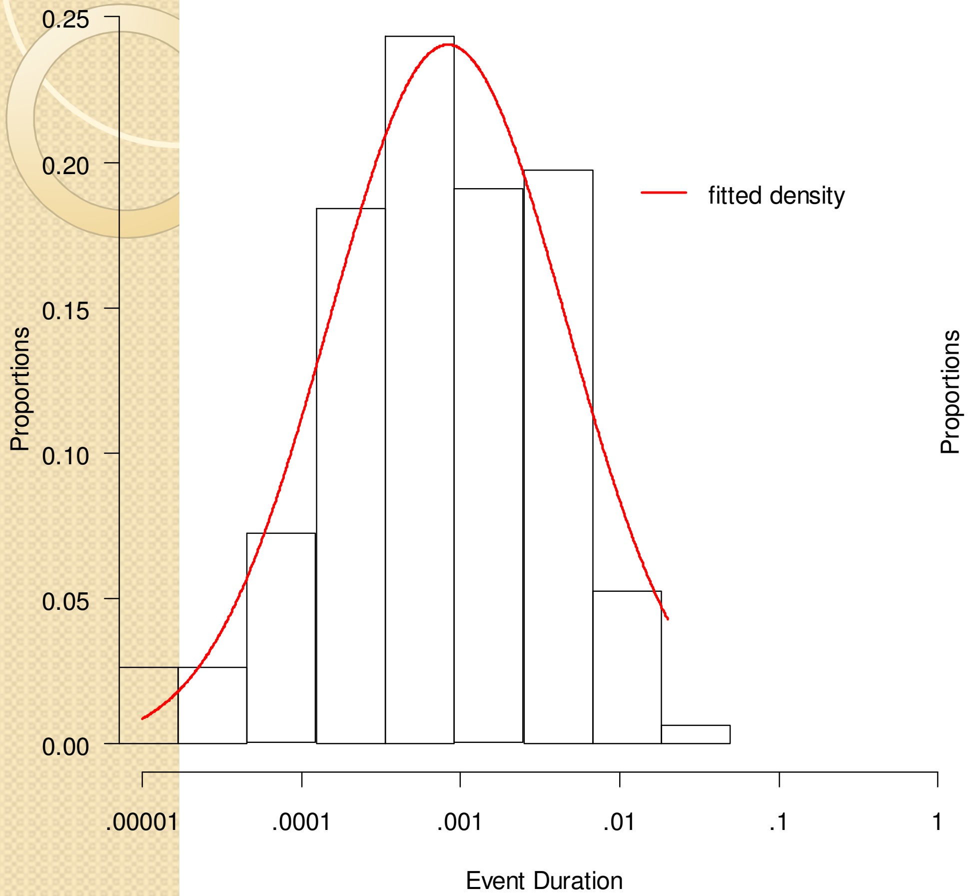


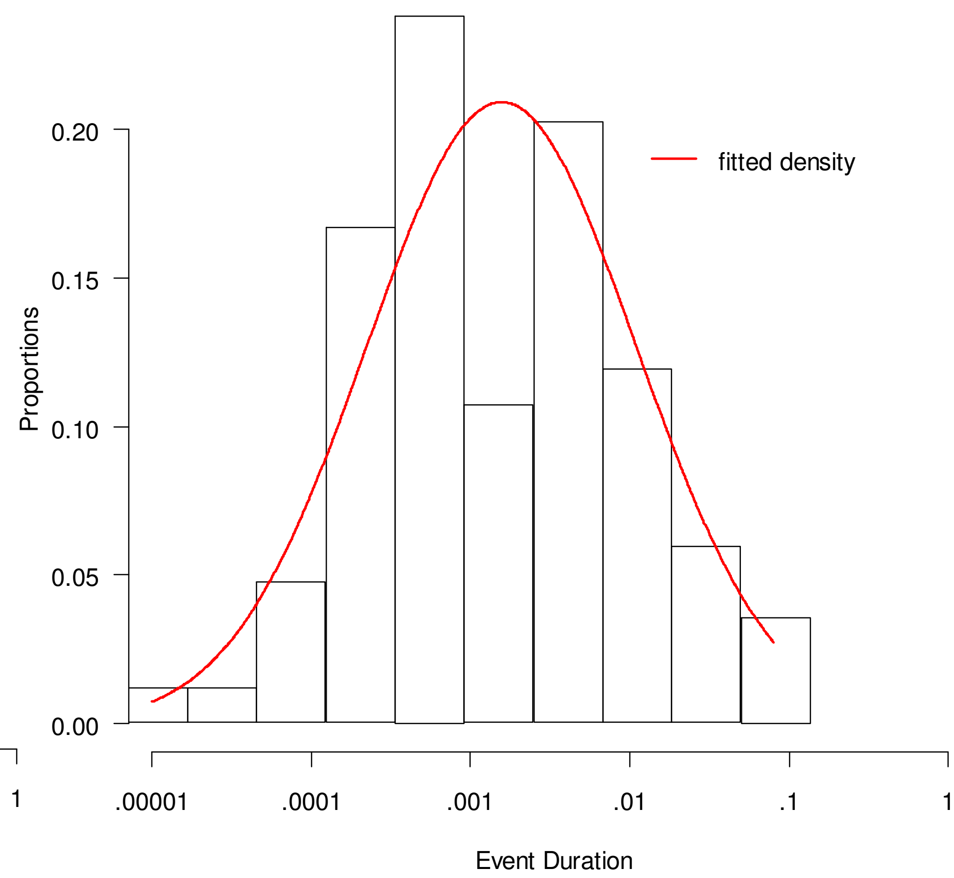


❖ Observation-window-adjusted observed versus estimated expected cumulative number of events

$$E_k(t;\boldsymbol{\theta}_k) = \sum_{i=1}^{n}\sum_{l=1}^{m_i} \mathbf{1}_{\left(\tau_{il}^{L} \le t\right)} \left[\Lambda_k(\min\{t, \tau_{il}^{R}\};\boldsymbol{\theta}_k) - \Lambda_k(\tau_{il}^{L};\boldsymbol{\theta}_k)\right].$$

Event Duration Modeling





❖ Truncated Log-Normal Distribution

Truncated Lognormal Distribution

❖ There is often an upper limit for the event duration

❖ The event durations for FM1 and FM2 of System A are typically less than 0.02 TU and 0.08 TU

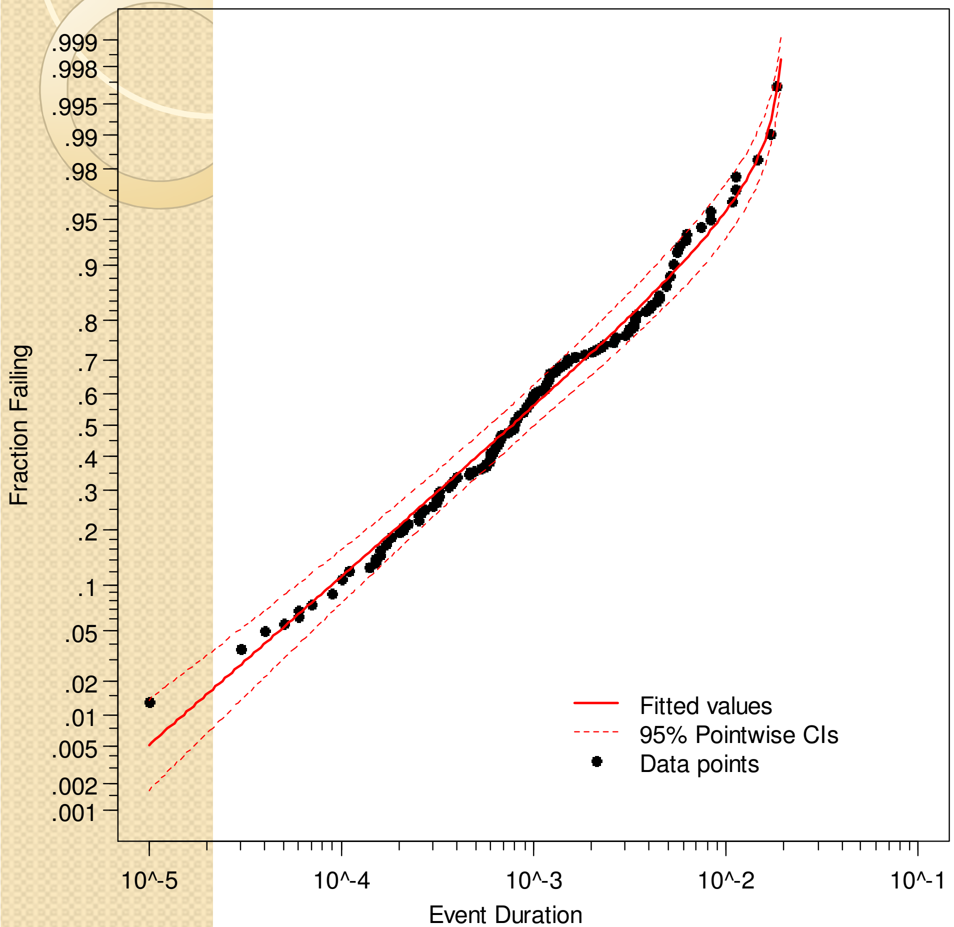
❖ Truncated lognormal is used to model the event duration

❖ The cdf and pdf for truncated lognormal distribution are

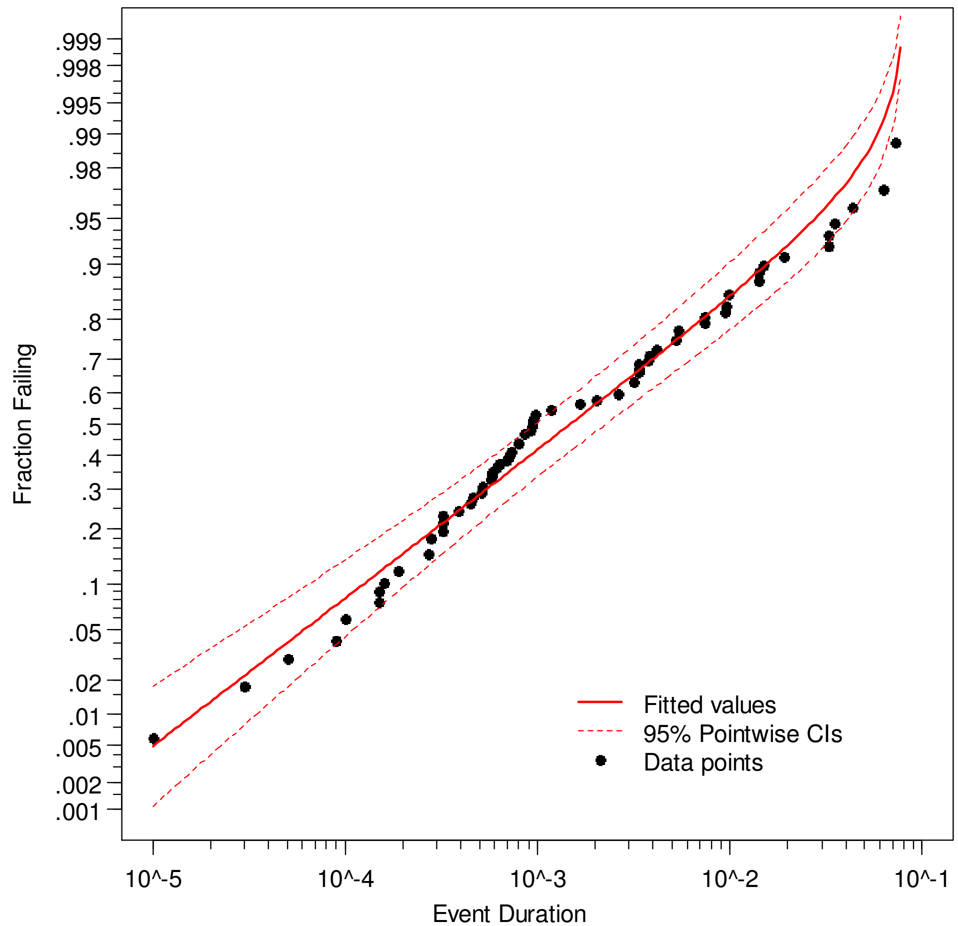
$$F_{\text{TLN}}(x;\psi) = \frac{F(x;\psi)}{F(x_R;\psi)} \quad \text{and} \quad f_{\text{TLN}}(x;\psi) = \frac{f(x;\psi)}{F(x_R;\psi)}, \quad 0 < x \le x_R,$$

❖ Parameters are estimated by ML method and sandwich type variance estimator is used for inference.

Assess Model Fitting
Truncated Lognormal on Lognormal Probability Plot



FM1



FM2

Unavailability Analysis

Unavailable Duration

❖ The unavailable duration within a certain time interval is of interest

❖ The total unavailable time in the TU interval ω is

$$Y_\omega = \sum_{k=1}^{K} \mathbf{1}_{(Z_{k\omega}>0)} \sum_{l=1}^{Z_{k\omega}} X_{kl}$$

❖ $Z_{k\omega} = N_k(\omega) - N_k(\omega-1)$ is the number of events of type *k* which occur during interval $(\omega-1,\ \omega]$

❖ $Z_{k\omega}$ follows the Poisson distribution with intensity parameter $\xi_{k\omega} = \Lambda_k(\omega) - \Lambda_k(\omega-1)$.

❖ X_k is the duration for an event with type *k* .

Unavailability Metric

❖ The traditional metric for system unavailability is

$$u_m(\omega) = \frac{\mathbf{E}(\text{downtime})}{\text{Length of the time interval}} = \frac{\mathbf{E}(Y_\omega)}{1} = \mathbf{E}(Y_\omega).$$

❖ Only considers the mean of unavailable time

❖ The distributional variability is not considered

❖ A conservative metric is needed to quantify the variability in the system unavailable duration

❖ Important to consider the possibility of extreme events

❖ A *Quantile*-unavailability (*Q*-unavailability) metric is defined as

$$u_p(\omega) = F_\omega^{-1}(p)$$

$F_\omega(y)$ is the cdf of Y_ω

❖ It is the *p* quantile of the distribution, e.g., *p*=0.95, 0.99

Unavailability Metric Estimation Algorithm

1. Simulate $Z^*_{k\omega}$ from Poisson$(\hat{\xi}_{k\omega})$, $k = 1, \ldots, K$.
2. If $\sum_{k=1}^{K} Z^*_{k\omega} > 0$, go to step 3.

   If $\sum_{k=1}^{K} Z^*_{k\omega} = 0$, set $Y_\omega^{\ *} = 0$ and go to step 5.
3. For those $Z^*_{k\omega} > 0$, simulate $Z^*_{k\omega}$ number of X^*_{kl} from a truncated lognormal distribution .
4. Compute $Y^*_\omega = \sum_{k=1}^{K} \sum_{l=1}^{Z^*_{k\omega}} X^*_{kl}$.
5. Repeat the above steps B times to obtain $Y^{*b}_\omega, b = 1, 2, \ldots, B$.
6. The estimate of availability metric, $\hat{u}_p(\omega)$, is computed as the empirical p quantile of Y_ω by $Y_\omega^{*([pB])}$.

Confidence Interval Procedure

1. Simulate $\hat{\boldsymbol{\theta}}^*$ from $\mathrm{N}(\hat{\boldsymbol{\theta}}, \Sigma_{\hat{\boldsymbol{\theta}}})$ and $\hat{\boldsymbol{\psi}}^*$ from $\mathrm{N}(\hat{\boldsymbol{\psi}}, \Sigma_{\hat{\boldsymbol{\psi}}})$.

2. Compute $\hat{\xi}^*_{k\omega}$ with $\hat{\boldsymbol{\theta}}$ replaced by $\hat{\boldsymbol{\theta}}^*$

3. Compute $\hat{u}^*_p(\omega)$ with $\hat{\xi}_{k\omega}$ replaced by $\hat{\xi}^*_{k\omega}$ and $\hat{\boldsymbol{\psi}}$ replaced by $\hat{\boldsymbol{\psi}}^*$.

4. Repeat the above steps B times to obtain $\hat{u}^{*b}_p(\omega)$, $b = 1, 2, \ldots, B$.

5. Compute the one-side upper CI by using $\hat{u}_p^{*([(1-\alpha)B])}(\omega)$.

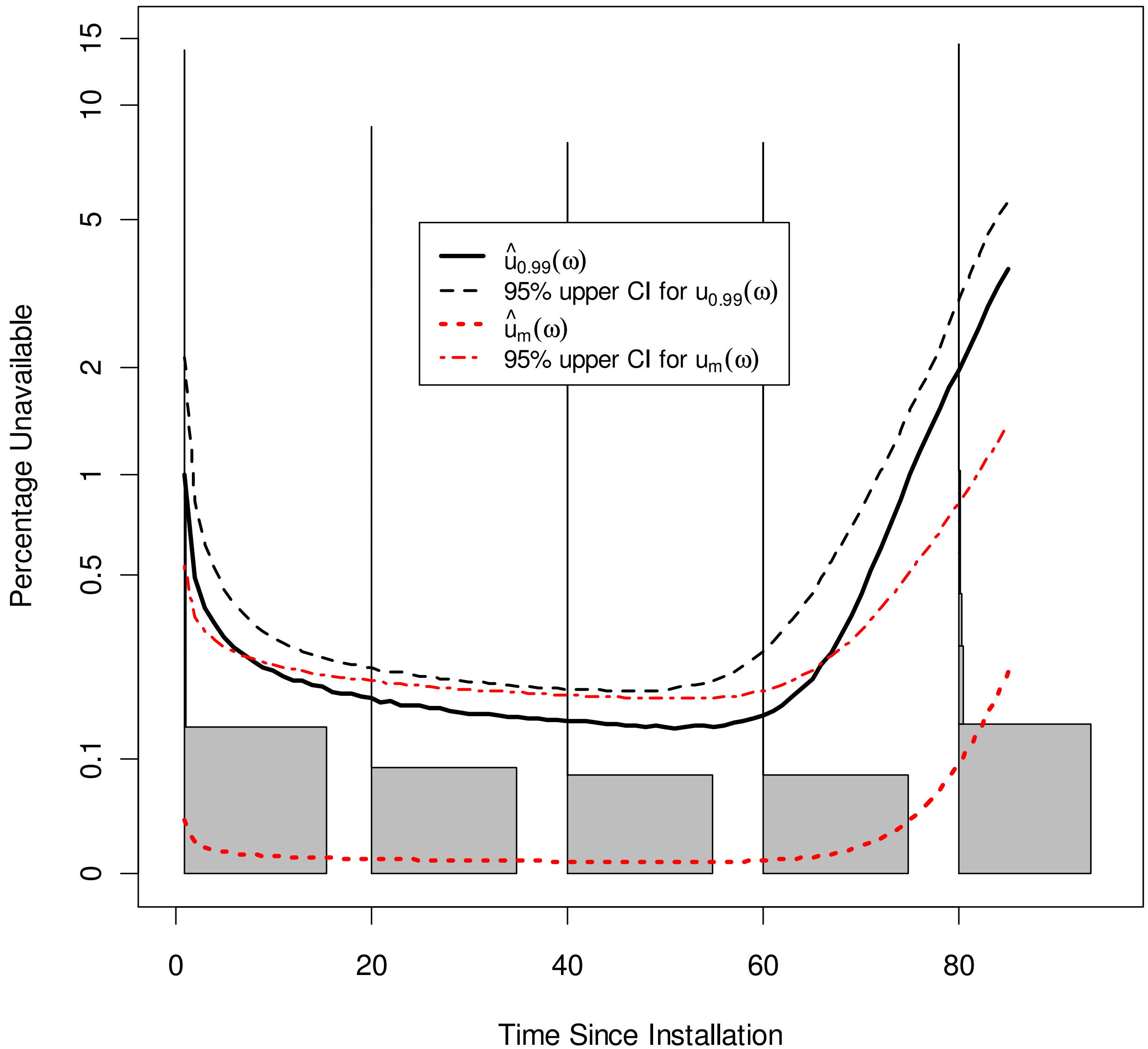
Conditional Analysis

- In some applications, it is also useful to estimate the system unavailability metric for a particular unit that has been in service for a period of time
- This information is useful for maintenance scheduling and system health monitoring for individual units.
- The only difference is to compute $\widehat{\xi}_{k\omega}$ by

$$\widehat{\xi}_{kt_i} = \Lambda_k(t_i + \Delta) - \Lambda_k(t_i).$$

- Here t_i is the current age for unit *i* and $\triangle$ is the future time interval that is of interest.

Percentage of Unavailability for System A



Cost Analysis

❖ The unavailable time of a system will result in expenditures

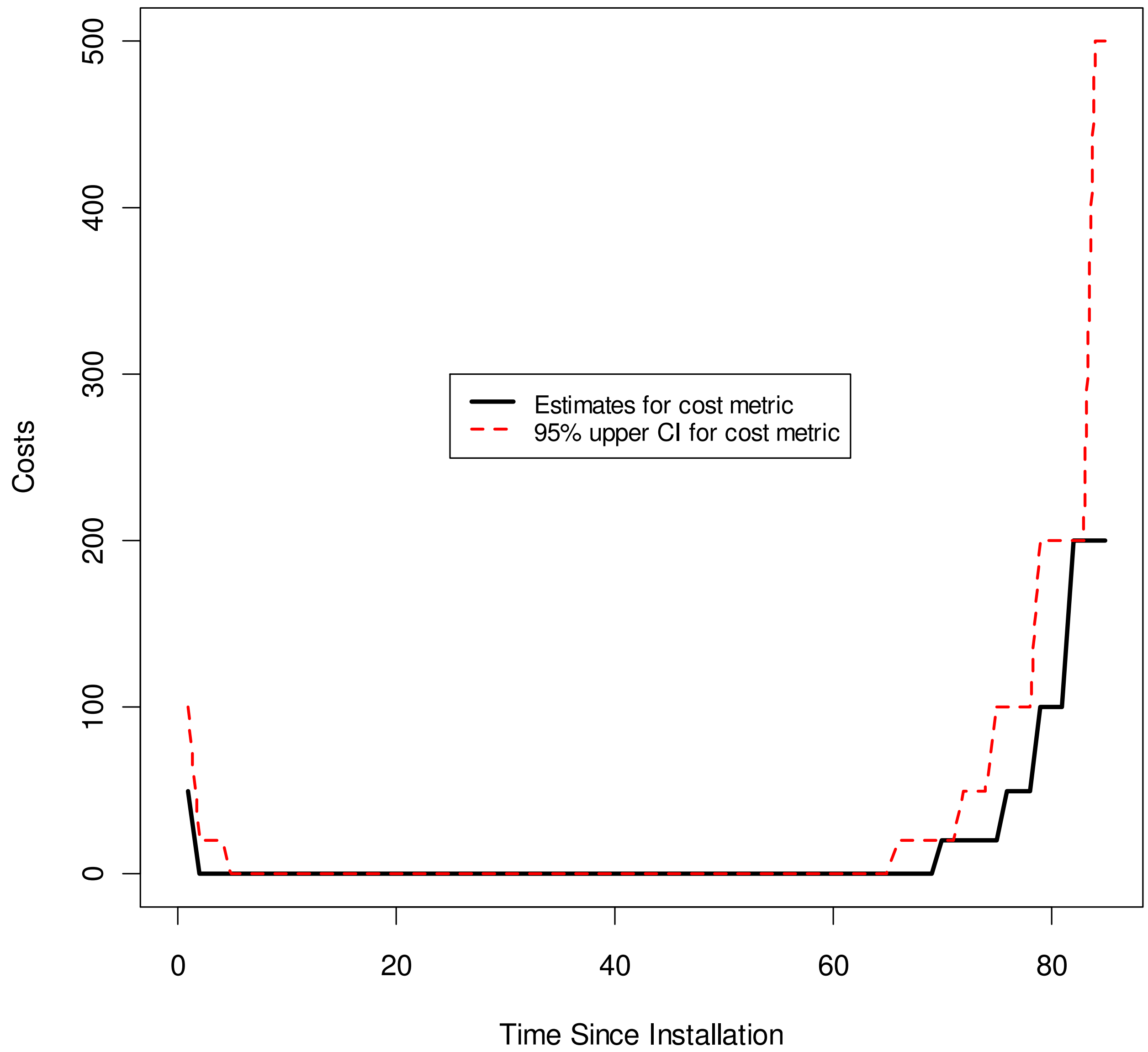
❖ For a given cost function, the cost C_ω can be expressed as $C_\omega = f(Y_\omega)$

❖ A conservative metric for the cost is $f[u_p(\omega)]$, if the cost function is non-decreasing.

❖ Example

$$C_\omega = f(Y_\omega) = \begin{cases} 0, & Y_\omega < 0.5\%; \\ 20, & 0.5\% \le Y_\omega < 1\% \\ 50, & 1\% \le Y_\omega < 1.5\% \\ 100, & 1.5\% \le Y_\omega < 2.5\% \\ 200, & 2.5\% \le Y_\omega < 5\% \\ 500, & Y_\omega \ge 5\% \end{cases} .$$

Cost Analysis for System A



Extension of the Q-Unavailability Metric

- We have discussed the definition of the Q-unavailability metric based on NHPP.
- The definition is generic and the extension to a general repairable system is straightforward.
- Considering a general repairable system, the ith unavailable period is denoted by $[T_i^L, T_i^R], i = 1, 2, \cdots$
- For an interested operation period [L, R], the percentage of the total unavailability time is

$$Y = \frac{\sum_i (\min\{T_i^R, R\} - \max\{T_i^L, L\})}{R - L}$$

- The Q-unavailability metric is defined as $u_p = F_Y^{-1}(p)$

where $F_Y(y) = \Pr(Y \leq y)$ is the cdf.

Conclusions and Areas for future research

- We proposed a new system unavailability metric and developed its estimation based on complicated field data.
- The extended NHPP is used to model the event process.
- The truncated lognormal distribution is used to model the event duration.
- The combination of modeling of event intensity and event duration enables the estimation of unavailability metrics.

- In the future, an alternative approach would be developing nonparametric methods to model the event intensity.
- In the modeling of the time-to-event process, we assumed the System A population is homogeneous.
- It will be interesting, however, to consider population nonhomogeneity in future applications, e.g., frailty model.