Chapter 22

Analysis of Repairable System and Other Recurrent Events Data

W. Q. Meeker, L. A. Escobar, and F. G. Pascual
Iowa State University, Louisiana State University, and Washington State University.

Based on Meeker, Escobar, and Pascual (2021): *Statistical Methods for Reliability Data, Second Edition*, John Wiley & Sons Inc.

May 24, 2021
11h 9min

Chapter 22
Analysis of Repairable System and Other Recurrent Events Data

Topics discussed in this chapter are:

- An introduction to recurrent events data and recurrent events data analysis.
- Estimation of the mean cumulative function (MCF).
- The variance of the MCF estimator, confidence intervals for the MCF, and a simple example.
- Comparison of two MCFs.
- Analysis of recurrent-event data with multiple event types.

Chapter 22

Analysis of Repairable System and Other Recurrent Events Data

Segment 1

An Introduction to Recurrent Events Data

Valve Seat Replacement Times

Introduction

Recurrent events data are a sequence of recurrences $T_1, T_2, \ldots$ in time (a point-process). Data may be from one or more than one observational units.

In general, the interest is on:

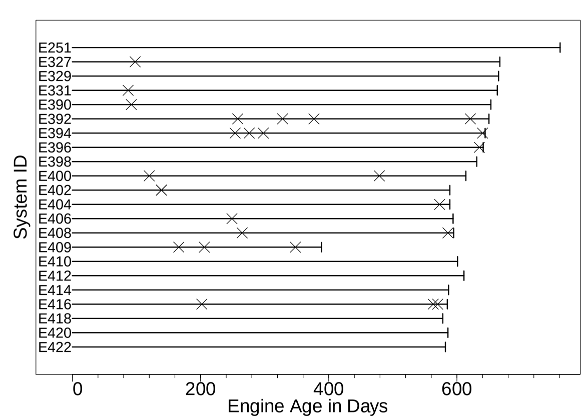
- The number of recurrences in the interval $(0, t]$ as a function of t.
- The expected number of recurrences in the interval $(0, t]$ as a function of t.
- The recurrence rate $\lambda(t)$ as a function of time t.
- The distribution of the times between recurrences, $\tau_j = T_j - T_{j-1}$ $(j = 1, 2, \ldots)$ where $T_0 = 0$.

Valve Seat Replacement Times

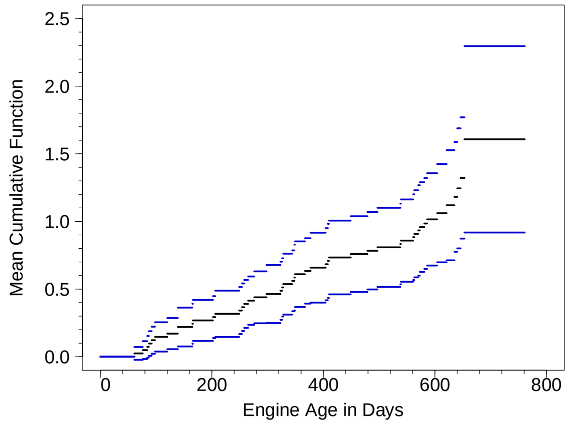
Data collected from valve seats from a fleet of 41 diesel engines operated in and around Beijing, China (days of operation).

- Each engine has 16 valves.
- Most failures caused by operating in a dusty environment.
- Does the replacement rate increase with age?
- How many replacement valves will be needed in the future?
- Can valve life in these systems be modeled as a renewal process?
- Data from Nelson (1995).

Valve Seat Replacement Times Event Plot



Estimate of the Mean Cumulative Replacement Function for the Valve Seat Data



Recurrent Events Data

- Recurrences (e.g., failures, returns, or replacements) are observed in a fixed observation interval $(0, t_a]$.
- The data may be reported on several different ways.
  - ▶ Single system or multiple systems.
  - ▶ Exact recurrence times $t_1 < \ldots < t_r$ $(t_r \leq t_a)$ resulting from continuous inspection in $(0, t_a]$.
  - ▶ Number of interval censored recurrences $d_1, \ldots, d_m$ in the intervals $(0, t_1], (t_1, t_2], \ldots (t_{m-1}, t_m]$, $(t_m = t_a)$ resulting from inspections on $(0, t_a]$.
  - ▶ Window-observation data when events are recorded only in certain windows of time.

Multiple Systems - Data and Model

- **Data:** For a single system, $N(s, t)$ denotes the cumulative number of recurrences in the interval $(s, t]$. And $N(t) = N(0, t)$.
- **Model:** The mean cumulative function (MCF) at time t is defined as $\Lambda(t) = \mathrm{E}[N(t)]$, where the expectation is over the variability of each system and the unit to unit variability in the population.
- When $\Lambda(t)$ is differentiable,
$$\lambda(t) = \frac{d\mathrm{E}[N(t)]}{dt} = \frac{d\Lambda(t)}{dt}$$
defines the recurrence rate per system (or **mean** recurrence rate for a collection of systems).
- Some times the interest is on cost over time and $\Lambda(t) = \mathrm{E}[C(t)]$ is the mean cumulative cost per unit in $(0, t]$.

Chapter 22

Analysis of Repairable System and Other Recurrent Events Data

Segment 2

Estimation of the MCF

Earth-Moving Machine Maintenance Actions Cylinder Replacement Data

Nonparametric Methods for Recurrent Events Data

Under the general cumulative recurrent events model the nonparametric analysis provides:

- Nonparametric estimate of the MCF $\Lambda(t)$.
- Variance of the nonparametric estimator of the MCF $\Lambda(t)$.
- Nonparametric confidence interval for $\Lambda(t)$.
- Nonparametric confidence interval for the difference between two cumulative occurrence models.

Nonparametric Estimate of a Population MCF Definition and Assumptions

Here we present a nonparametric estimate of an MCF $\Lambda(t)$. The estimator is nonparametric in the sense that the method does not require specification of a model for the point process recurrence rate.

- Suppose that there is a fleet of n units generating recurrent events.
- Suppose also that the time at which observation on a unit is terminated is not systematically related to any factor related to the recurrence time distribution.

Nonparametric Estimate of MCF Input Data Notation Conventions

- Let t_{ij} be recurrence time j for system i, $j = 1, \ldots, m_i$ and $i = 1, \ldots, n$.
- Order the unique recurrence times from smallest to largest and collect the distinct recurrences times say $t_1 < \ldots < t_m$. Thus m is the number of unique event times.
- Some applications focus on the number of recurrent events and others focus on values (e.g., cost) associated with the recurrrent events.
- Let $d_i(t_j)$ the total number of recurrences or other quantitative value (such as cost) for unit i **at** time t_j.
- Let

$$\delta_i(t_j) = \begin{cases} 1 & \text{if system } i \text{ is being observed } \textbf{at} \text{ time } t_j \\ 0 & \text{otherwise.} \end{cases}$$

Estimation of the MCF $\Lambda(t)$ with Multiple Systems Notation for Computational Elements

- The total number of system recurrences **at** time t_k is

$$d.(t_k) = \sum_{i=1}^{n} \delta_i(t_k) d_i(t_k),$$

- The size of the risk set **at** time t_k is

$$\delta.(t_k) = \sum_{i=1}^{n} \delta_i(t_k),$$

- The mean number of system recurrences **at** time t_k (or proportion of recurrences if a system can have only one recurrence at a time) is

$$\bar{d}(t_k) = \frac{d.(t_k)}{\delta.(t_k)}$$

Estimation of the MCF $\Lambda(t)$ with Multiple Systems

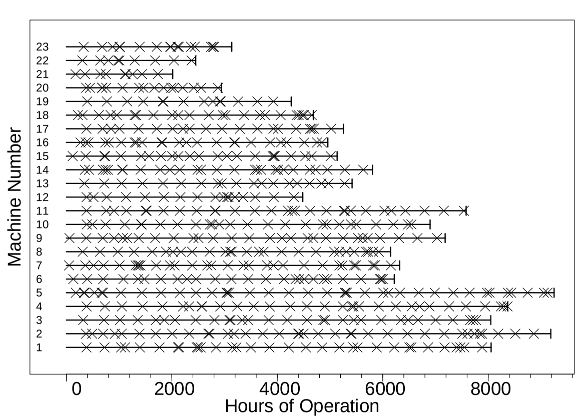
The nonparametric estimate of the MCF $\widehat{\Lambda}(t)$ is constant between events (which occur at the t_j's) and at time t_j, the estimate jumps to

$$\begin{aligned}\widehat{\Lambda}(t_j) &= \sum_{k=1}^{j} \frac{\sum_{i=1}^{n} \delta_i(t_k) d_i(t_k)}{\sum_{i=1}^{n} \delta_i(t_k)} \\ &= \sum_{k=1}^{j} \frac{d.(t_k)}{\delta.(t_k)} \\ &= \sum_{k=1}^{j} \bar{d}(t_k), \quad j = 1, \ldots, m.\end{aligned}$$

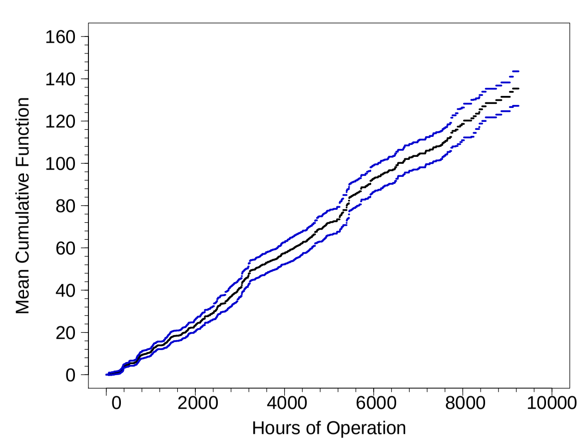
Earth-Moving Machine Maintenance Actions

- Fleet of 23 earth-moving machines put into service over time
- Preventive maintenance every 300-400 hours of operation
- Major overhaul every 2000-3000 hours of operation
- Many unscheduled maintenance actions
- The response is the number of labor hours for each maintenance action

Event Plot of Earth-Moving Machine Maintenance Actions



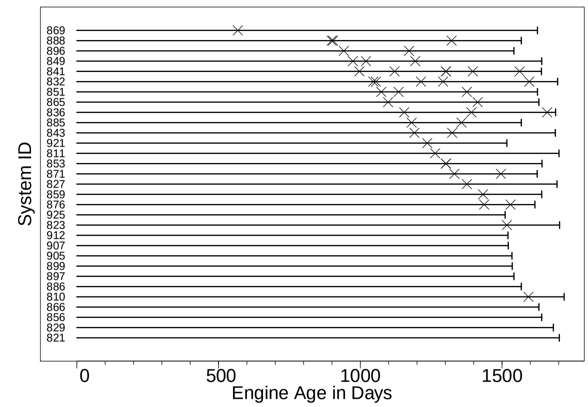
MCF plot of Earth-Moving Machine Maintenance Actions



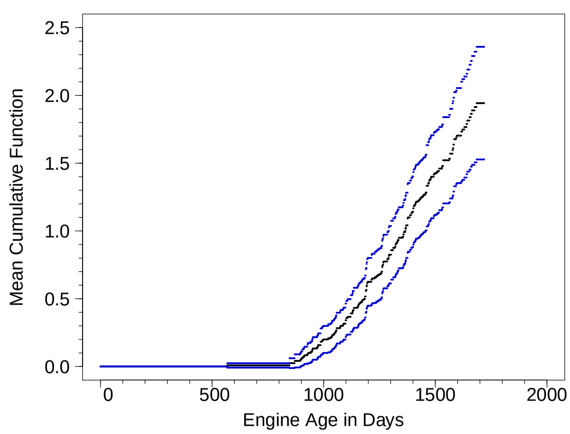
Cylinder Replacement Data

- Cylinder replacement times on 120 locomotive diesel engines.
- Cylinders can develop leaks or have low compression for some other reason.
- Such cylinders are replaced by a rebuilt cylinder.
- Each engine has 16 cylinders.
- More than one cylinder may be replaced at an inspection.
- Is preventive replacement of cylinders appropriate?
- Data from Nelson and Doganaksoy (1989).

Cylinder Replacement Time Event Plot (Subset of Systems)



Estimate of Mean Cumulative Replacement Function for the Diesel Cylinders



Chapter 22

Analysis of Repairable System and Other Recurrent Events Data

Segment 3

Variance of the MCF Estimator Confidence Intervals for the MCF and a Simple Example

Variance of $\widehat{\Lambda}(t)$

- Suppose that the observation times are fixed. Then the number of recurrent events is random.
- Suppose that the systems are independent.
- Define $d(t_k)$ as the random variable that describes the number of system recurrences or the amount of the variable (such as cost) of interest **at** t_k for a system sampled at random from the population of systems.
- Recall that

$$\widehat{\Lambda}(t_j) = \sum_{k=1}^{j} \bar{d}(t_k), \quad j = 1, \ldots, m.$$

- Then, direct computations give

$$\begin{aligned}\text{Var}[\widehat{\Lambda}(t_j)] &= \sum_{k=1}^{j} \text{Var}[\bar{d}(t_k)] + 2\sum_{k=1}^{j-1}\sum_{v=k+1}^{j} \text{Cov}[\bar{d}(t_k), \bar{d}(t_v)] \\ &= \sum_{k=1}^{j} \frac{\text{Var}[d(t_k)]}{\delta_{.}(t_k)} + 2\sum_{k=1}^{j-1}\sum_{v=k+1}^{j} \frac{\text{Cov}[d(t_k), d(t_v)]}{\delta_{.}(t_k)}.\end{aligned}$$

Estimate of Var$[\widehat{\Lambda}(t)]$

- To estimate Var$[d(t_k)]$, we use the assumption that $d_i(t_k)$, $i = 1, \ldots, n$ is a random sample from $d(t_k)$.

  The moment estimators are

$$\widehat{\text{Var}}[d(t_k)] = \sum_{i=1}^{n} \frac{\delta_i(t_k)}{\delta_{.}(t_k)}[d_i(t_k) - \bar{d}(t_k)]^2$$

$$\widehat{\text{Cov}}[d(t_k), d(t_v)] = \sum_{i=1}^{n} \frac{\delta_i(t_v)}{\delta_{.}(t_v)}[d_i(t_k) - \bar{d}(t_k)]d_i(t_v).$$

- Substituting these into the variance formula, and after simplifications, one gets

$$\begin{aligned}\widehat{\text{Var}}[\widehat{\mu}(t_j)] &= \sum_{k=1}^{j} \frac{\widehat{\text{Var}}[d(t_k)]}{\delta_{.}(t_k)} + 2\sum_{k=1}^{j-1}\sum_{v=k+1}^{j} \frac{\widehat{\text{Cov}}[d(t_k), d(t_v)]}{\delta_{.}(t_k)} \\ &= \sum_{i=1}^{n}\left\{\sum_{k=1}^{j} \frac{\delta_i(t_k)}{\delta_{.}(t_k)}\left[d_i(t_k) - \bar{d}_{.}(t_k)\right]\right\}^2.\end{aligned}$$

Comment on Other Estimates of Var$[\widehat{\Lambda}(t)]$

- An alternative to the moment estimators of variances and covariances, one can use Nelson's (slightly different) unbiased estimators given by

$$\widehat{\text{Var}}[d(t_k)] = \sum_{i=1}^{n} \frac{\delta_i(t_k)}{\delta.(t_k) - 1}[d_i(t_k) - \bar{d}(t_k)]^2$$
$$\widehat{\text{Cov}}[d(t_k), d(t_v)] = \sum_{i=1}^{n} \frac{\delta_i(t_v)}{\delta.(t_v) - 1}[d_i(t_k) - \bar{d}(t_k)]d_i(t_v).$$

- Using the unbiased estimates can result in a *negative estimate* for Var$[\widehat{\Lambda}(t)]$. The probability of this event is small unless the number of units under observation is small (e.g., fewer than 20).

Simple Example for 3 Systems Data

Consider 3 systems with the following system failures and censoring times

System	System Failures	Censoring Time
1	5, 8	12
2		16
3	1, 8, 16	20

Then the collection of all system failures is

$$t_1 = 1, t_2 = 5, t_3 = 8, t_4 = 16$$

Simple Example Estimation of $\mu(t)$

- Point estimation:

j	t_j	δ_1	δ_2	δ_3	d_1	d_2	d_3	$\delta.$	$d.$	$\bar{d}$	$\widehat{\mu}(t_j)$
1	1	1	1	1	0	0	1	3	1	1/3	1/3
2	5	1	1	1	1	0	0	3	1	1/3	2/3
3	8	1	1	1	1	0	1	3	2	2/3	4/3
4	16	0	1	1	0	0	1	2	1	1/2	11/6

- Estimation of variances:

$$\widehat{\text{Var}}[\widehat{\mu}(t_1)] = [(1/3) \times (0 - 1/3)]^2 + [(1/3) \times (0 - 1/3)]^2 + [(1/3) \times (1 - 1/3)]^2 = \frac{6}{81}.$$

Similar computations yield:

$$\widehat{\text{Var}}[\widehat{\mu}(t_2)] = 6/81 = 0.0741$$
$$\widehat{\text{Var}}[\widehat{\mu}(t_3)] = 24/81 = 0.296$$
$$\widehat{\text{Var}}[\widehat{\mu}(t_4)] = 163/216 = 0.755$$

Estimation of the MCF $\Lambda(t)$ with Finite Populations

Sometimes with field data, the number of systems is small and the inference of interest is on the number of recurrences and cost of *those* units.

- In this case, finite population methods are appropriate.

- The point estimator for the MCF $\Lambda(t)$ is the same. But to take in consideration sampling from a finite population the following estimates are used in computing $\widehat{\text{Var}}[\widehat{\Lambda}(t)]$:

$$\widehat{\text{Var}}[d(t_k)] = \left[1 - \frac{\delta.(t_k)}{N}\right] \sum_{i=1}^{n} \frac{\delta_i(t_k)}{\delta.(t_k)}[d_i(t_k) - \bar{d}(t_k)]^2$$
$$\widehat{\text{Cov}}[d(t_k), d(t_v)] = \left[1 - \frac{\delta.(t_v)}{N}\right] \sum_{i=1}^{n} \frac{\delta_i(t_v)}{\delta.(t_v)}[d_i(t_k) - \bar{d}(t_k)]d_i(t_v)$$

where N is the total number of systems in the population of interest.

Chapter 22

Analysis of Repairable System and Other Recurrent Events Data

Segment 4

Comparison of Two MCFs

Braking Grid Replacement Frequency Comparison

Braking Grid Replacement Frequency Comparison

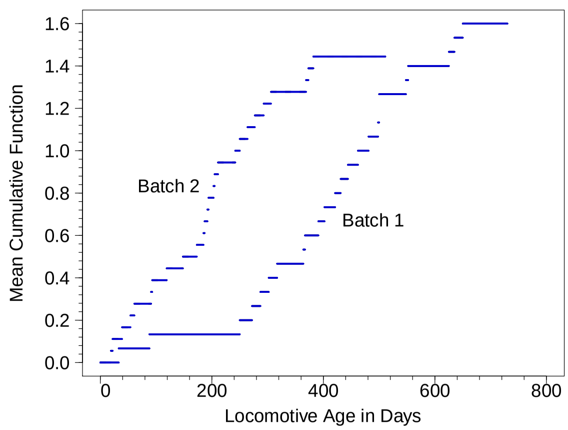
- A particular type of locomotive has six braking grids.

- Data available on locomotive age when a braking grid is replaced and the age at the end of the observation period.

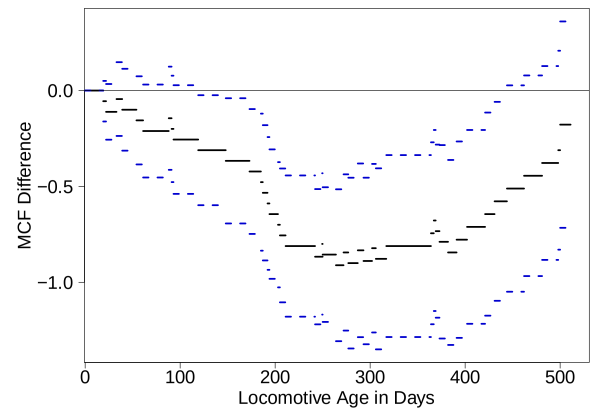
- A comparison of two different braking grids production batches is desired.

- The data are from Doganaksoy and Nelson (1998).

Comparison of MCFs for the Braking Grids from Production Batches 1 and 2



Difference $\widehat{\Lambda}_1 - \widehat{\Lambda}_2$ Between Sample MCFs for Batches 1 and 2 and Pointwise Approximate 95% Confidence Intervals for the Population Difference



Nonparametric Comparison of Two Samples of Recurrent Events Data

- Suppose that there are two independent samples of recurrent events data with MCF estimates given by $\widehat{\Lambda}_1(t)$ and $\widehat{\Lambda}_2(t)$, respectively.
- Let $\Delta(t) = \Lambda_1(t) - \Lambda_2(t)$ represent the MCF difference at t.
- A nonparametric estimate of $\Delta(t)$ is
$$\widehat{\Delta}(t) = \widehat{\Lambda}_1(t) - \widehat{\Lambda}_2(t)$$
with estimated variance given by
$$\widehat{\text{Var}}[\widehat{\Delta}(t)] = \widehat{\text{Var}}[\widehat{\Lambda}_1(t)] + \widehat{\text{Var}}[\widehat{\Lambda}_2(t)].$$
- The standard error for $\widehat{\Delta}(t)$ is
$$\text{se}_{\widehat{\Delta}(t)} = \sqrt{\widehat{\text{Var}}[\widehat{\Delta}(t)]}.$$
- An approximate $100(1-\alpha)\%$ confidence interval for $\Delta(t)$ is
$$\left[\underset{\sim}{\Delta}_{\mu}, \quad \widetilde{\Delta}_{\mu}\right] = \left[\widehat{\Delta} - z_{(1-\alpha/2)}\text{se}_{\widehat{\Delta}}, \quad \widehat{\Delta} + z_{(1-\alpha/2)}\text{se}_{\widehat{\Delta}}\right].$$

Chapter 22

Analysis of Repairable System and Other Recurrent Events Data

Segment 5

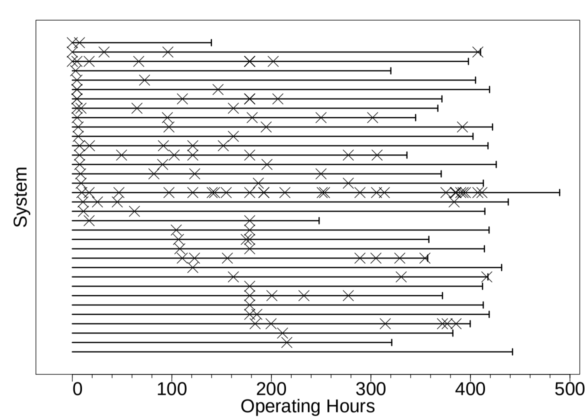
Analysis of Recurrent Events Data with Multiple Event Types

Analyzing the System E Recurrent Events Data

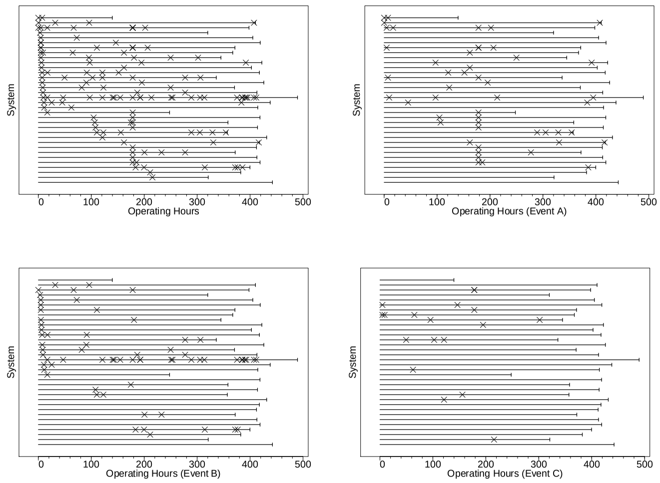
System E Recurrent Events Data

- System operators and others by technicians who periodically perform tests on the system to detect, identify, and mitigate potential failure-causing problems.
- 34 systems were monitored in the field for the occurrence of event types A, B, and C.
- There was interest in how much each event type contributes to the total MCF.
- Information would be fed back to improve system design for better reliability and availability.

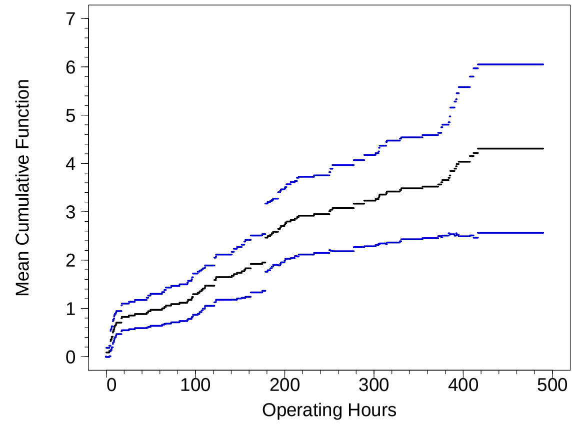
System E Event Plot Showing All Event Types



System E Event Plots



System E MCF Estimate
Ignoring Event-Type Information



MCF Estimation Using Event-Type Information

- Separate MCF estimates can be computed for each event type.
- Suppose that there are L event types.
- Let $d_{\ell i}(t_j)$ be the total number of recurrences of event type ℓ for system i at time t_j for $\ell = 1, \ldots, L$, $i = 1, \ldots, n$, and $j = 1, \ldots, m$.
- Then the MCF estimate for event type ℓ at time t_j is

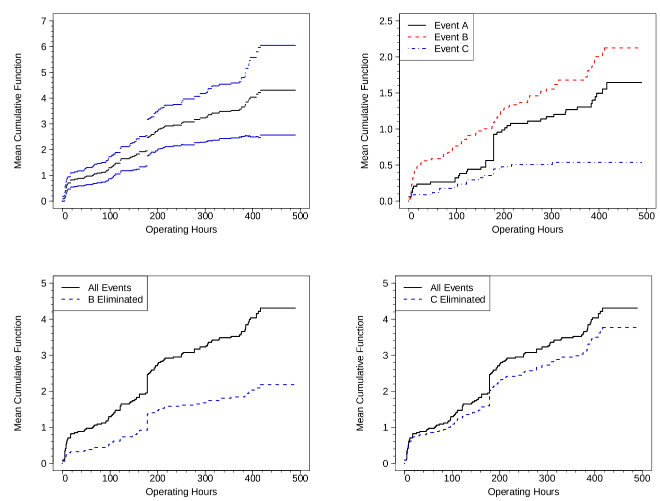
$$\widehat{\mu}_{\ell}(t_j) = \sum_{k=1}^{j} \frac{\sum_{i=1}^{n} \delta_i(t_k) d_{\ell i}(t_k)}{\sum_{i=1}^{n} \delta_i(t_k)}$$

- Then

$$\widehat{\mu}(t_j) = \sum_{\ell=1}^{L} \widehat{\mu}_{\ell}(t_j)$$

- The effect of eliminating one or more event types can be estimated by summing over the remaining event types.

System E Event MCF Estimate



Other Topics In Recurrent Events Data Analysis

- Window-observation data.
- Parametric models for recurrent event data (e.g., for prediction).
  - ▶ Nonhomogeneous Poisson process.
  - ▶ Renewal process.
  - ▶ Trend renewal process.
- Covariate adjustment/regression models for recurrent event data.
  - ▶ Fixed covariates.
  - ▶ Dynamic (time-varying) covariates.
  - ▶ Random effects (frailities) to describe unit-to-unit variability.

References

Doganaksoy, N. and W. B. Nelson (1998). A method to compare two samples of recurrence data. *Lifetime Data Analysis 4*, 51–63. []

Meeker, W. Q., L. A. Escobar, and F. G. Pascual (2021). *Statistical Methods for Reliability Data* (Second Edition). Wiley. [1]

Nelson, W. B. (1995). Confidence limits for recurrence data—applied to cost or number of product repairs. *Technometrics 37*, 147–157. []

Nelson, W. B. and N. Doganaksoy (1989). A computer program for an estimate and confidence limits for the mean cumulative function for cost or number of repairs of repairable products. TIS Report 89CRD239, General Electric Company Research and Development, Schenectady, NY. []