Chapter 16

Analysis of Repairable System and Other Recurrence Data

William Q. Meeker and Luis A. Escobar
Iowa State University and Louisiana State University

Based on the authors' text *Statistical Methods for Reliability Data*, John Wiley & Sons Inc. 1998.

January 5, 2006
19h 15min

Analysis of Recurrence Data
Chapter 16 Objectives

- Describe typical data from repairable systems and other applications that have recurrence data.
- Explain simple nonparametric graphical methods for presenting recurrence data.
- Show when system test data can be used to estimate the reliability of individual components.
- Describe simple parametric models for recurrence data.
- Illustrate the combined use of simple parametric and nonparametric graphical methods for making inferences from recurrence data.

Introduction

Recurrence data can be viewed as sequence of recurrences $T_1, T_2, \ldots$ in time (a point-process). Data may be from one or more than one observational unit.

In general the interest is on:

- The distribution of the times between recurrences, $\tau_j = T_j - T_{j-1}$ $(j = 1, 2, \ldots)$ where $T_0 = 0$.
- 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 $\nu(t)$ as a function of time t.

Recurrence Data

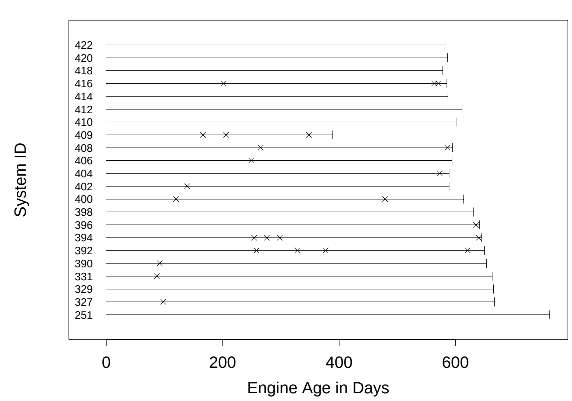
- Recurrences (e.g., failures or replacements) are observed in a fixed observation interval $(0, t_a]$.
- The data may be reported on several different ways.
  - Single system or multiple systems.
  - Exact recurrence times $t_1 < \ldots < t_r$ $(t_r \leq t_a)$ resulting from continuous inspection in $(0, t_a]$.
  - 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 $(0, t_a]$.

Valve Seat Replacement Times
(Nelson and Doganaksoy 1989)

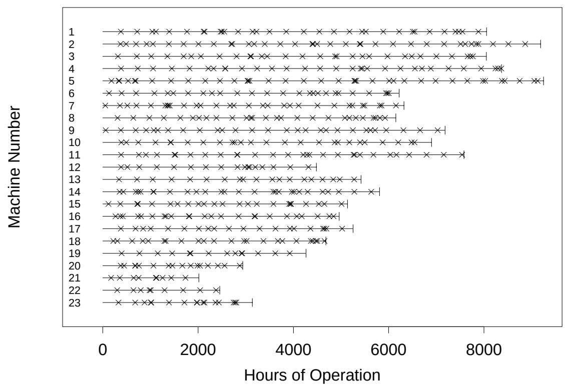
Data collected from valve seats from a fleet of 41 diesel engines operated in and around Beijing, China (days of operation).

- 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?

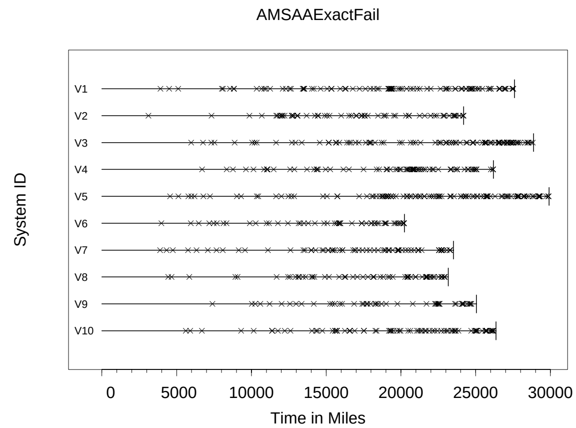
Valve Seat Replacement Times Event Plot
(Nelson and Doganaksoy 1989)



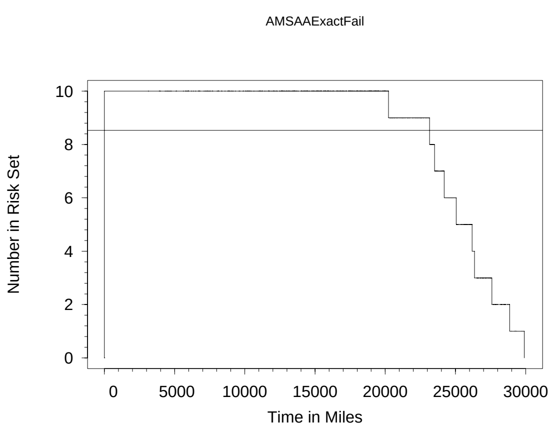
Earth-Moving Machine Maintenance Actions
(Meeker and Escobar 1998)



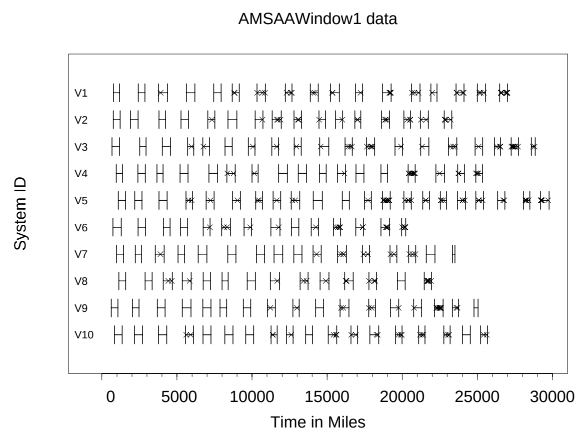
Event Plot of the Simulated AMSAA Vehicle Repairs
Power Rule $\beta = 2.76$ $\eta = 5447$
Continuous Inspection
(Cushing 2003)



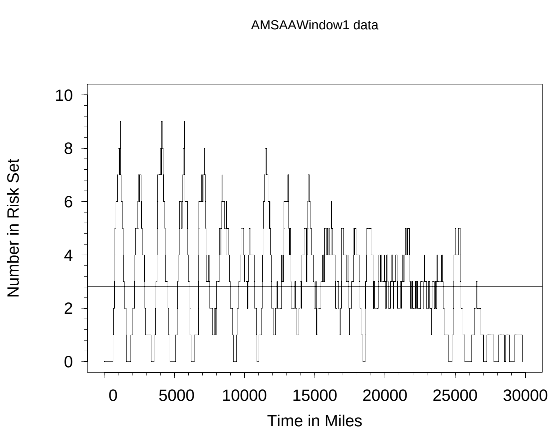
Risk Set Plot of Simulated AMSAA Vehicle Repairs
Power Rule $\beta = 2.76$ $\eta = 5447$
Continuous Inspection
(Cushing 2003)



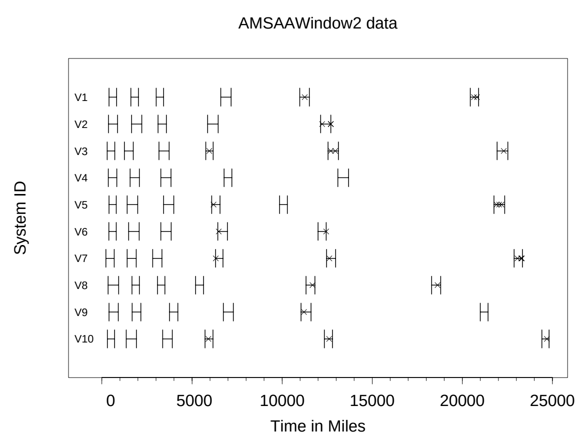
Event Plot of the Simulated AMSAA Vehicle Repairs
Power Rule $\beta = 2.76$ $\eta = 5447$
Random Inspection Windows
(Cushing 2003)



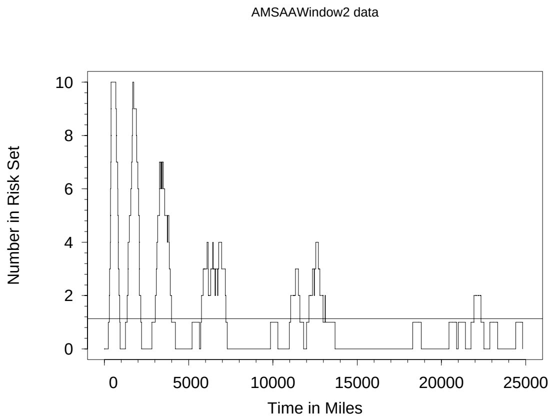
Risk Set Plot of Simulated AMSAA Vehicle Repairs
Power Rule $\beta = 2.76$ $\eta = 5447$
Random Inspection Windows
(Cushing 2003)



Simulated AMSAA Vehicle Repairs
Power Rule $\beta = 2.76$ $\eta = 5447$
Biased Inspection Windows
(Cushing 2003)



Risk Set Plot of Simulated AMSAA Vehicle Repairs
Power Rule $\beta = 2.76$ $\eta = 5447$
Biased Inspection Windows
(Cushing 2003)



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 $\mu(t) = \mathrm{E}[N(t)]$, where the expectation is over the variability of each system and the unit to unit variability in the population.
- Assuming that $\mu(t)$ is differentiable,

$$\nu(t) \;=\; \frac{d\mathrm{E}[N(t)]}{dt} = \frac{d\mu(t)}{dt}$$

defines the recurrence rate per system (or **average** recurrence rate for a collection of systems).
- Some times the interest is on cost over time and $\mu(t) = \mathrm{E}[C(t)]$ is the average cumulative cost per unit in $(0,t]$.

Nonparametric Methods for Recurrence Data

Under the general cumulative recurrence model the nonparametric analysis provides:

- Nonparametric estimate of the MCF $\mu(t)$.
- Nonparametric confidence interval for $\mu(t)$.
- 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 $\mu(t)$. 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 available a random sample (or entire population) of n units generating recurrences.
- 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

- $N_i(t)$ denotes the cumulative number of recurrences for the unit i at time t.
- Let $t_{ij}, j = 1, \ldots, m_i$ be the recurrence times for system i.
- Order all the recurrence times from smallest to largest and collect the distinct recurrences times say $t_1 < \ldots < t_m$.
- Let $d_i(t_j)$ the total number of recurrences for unit i at t_j.
- Let $\delta_i(t_j) = 1$ if system i is still being observed at time t_j and $\delta_i(t_j) = 0$ otherwise.

Estimation of the $\mu(t)$ 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 t_k is

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

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

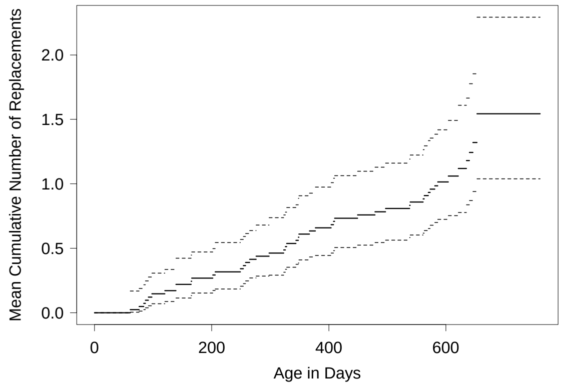
$$\bar{d}(t_k) = \frac{d.(t_k)}{\delta.(t_k)}$$

Estimation of the $\mu(t)$ with Multiple Systems

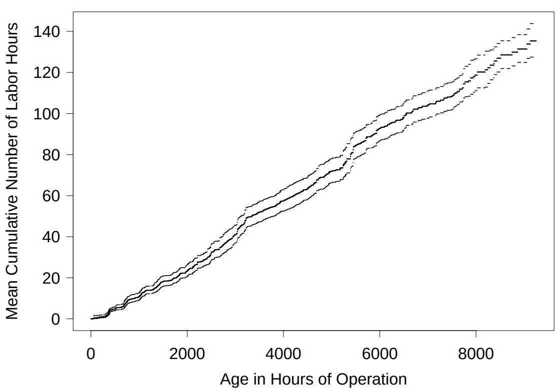
The nonparametric estimate of the MCF $\widehat{\mu}(t)$ is constant between events (which occur at the t_j's) and at t_j, the estimate jumps to

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

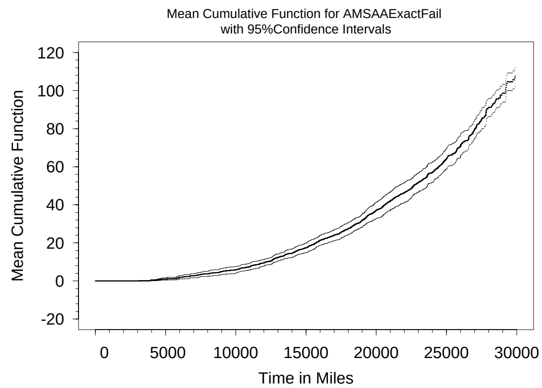
Estimate of Mean Cumulative Replacement Function for the Valve Seat



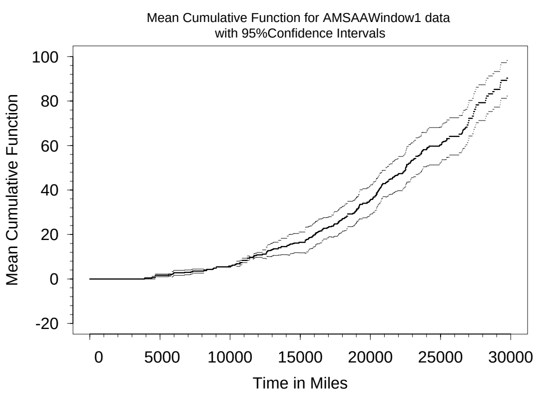
MCF plot of Earth-Moving Machine Maintenance Actions (Meeker and Escobar 1998)



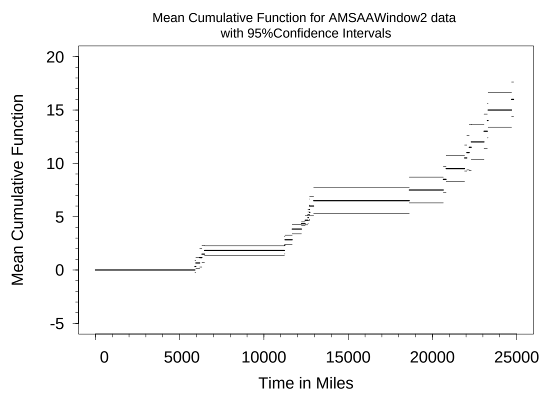
MCF Plot of the Simulated AMSAA Vehicle Repairs Power Rule $\beta = 2.76$ $\eta = 5447$ Continuous Inspection (Cushing 2003)



MCF Plot of Simulated AMSAA Vehicle Repairs Power Rule $\beta = 2.76$ $\eta = 5447$ Random Inspection Windows (Cushing 2003)



MCF Plot of Simulated AMSAA Vehicle Repairs Power Rule $\beta = 2.76$ $\eta = 5447$ Biased Inspection Windows (Cushing 2003)



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

- Suppose that the observation times are fixed. Then the number of recurrences is random.
- Suppose that the systems are independent.
- Define $d(t_k)$ as the random variable that describes the number of system recurrences at t_k for a system sampled at random from the population of systems.
- Direct computations give

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

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

- To estimate $\text{Var}[d(t_k)]$, 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

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

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

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

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

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

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

- Using the unbiased estimates can result in a *negative estimate* for $\text{Var}[\widehat{\mu}(t)]$. The probability of this event is small unless the number of units under observation is small (e.g., less 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	$\delta_{\cdot}$	$d_{\cdot}$	$\bar{d}$	$\widehat{\mu}(t_j)$
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:

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

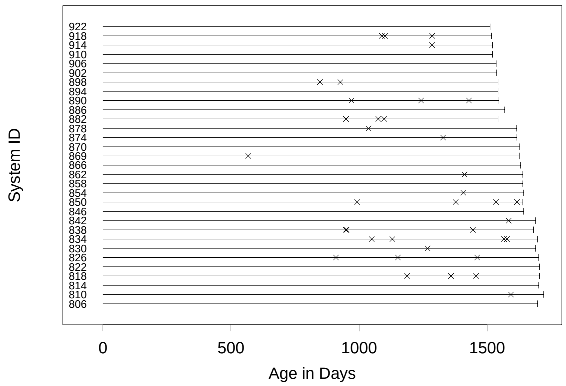
  Similar computations yield:

$$\begin{aligned}\widehat{\text{Var}}[\widehat{\mu}(t_2)] &= 6/81 = .0741 \\ \widehat{\text{Var}}[\widehat{\mu}(t_3)] &= 24/81 = .296 \\ \widehat{\text{Var}}[\widehat{\mu}(t_4)] &= 163/216 = .755\end{aligned}$$

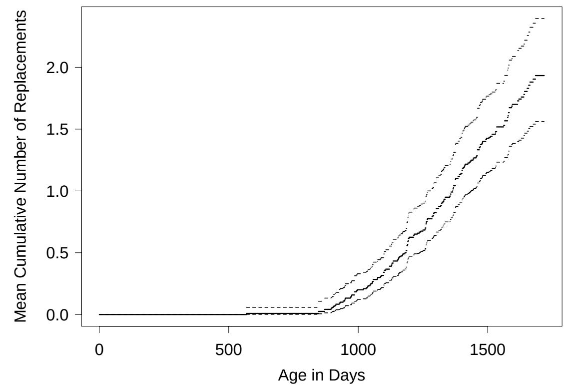
Cylinder Replacement Data

- Nelson and Doganaksoy provide 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?

Cylinder Replacement Time Event Plot (Subset of Systems) (Nelson and Doganaksoy 1989)



Estimate of Mean Cumulative Replacement Function for the Diesel Cylinders



Estimation of $\mu(t)$ 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 $\mu(t)$ is the same. But to take in consideration sampling from a finite population the following estimates are used in computing $\widehat{\text{Var}}[\widehat{\mu}(t)]$:

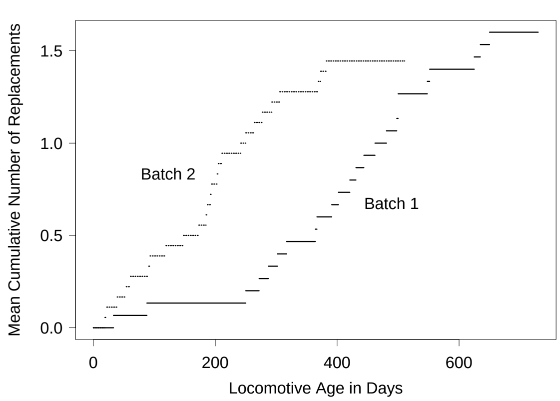
$$\widehat{\text{Var}}[d(t_k)] = \left[1 - \frac{\delta.(t_k)}{N}\right] \sum_{i=1}^{n} \frac{\delta_i(t_k)}{\delta.(t_k)} [d_i(t_k) - \bar{d}(t_k)]^2$$
$$\widehat{\text{Cov}}[d(t_k), d(t_v)] = \left[1 - \frac{\delta.(t_v)}{N}\right] \sum_{i=1}^{n} \frac{\delta_i(t_v)}{\delta.(t_v)} [d_i(t_k) - \bar{d}(t_k)]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

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

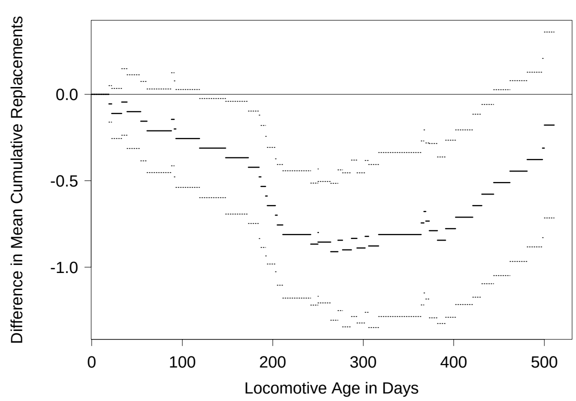
Braking Grid Replacement Frequency Comparison (Doganaksoy and Nelson 1991)

- 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 between two different production batches of braking grids is desired.

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



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



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

- When there is a single system the point estimate $\widehat{\mu}(t)$ is the number of system recurrences up to t.
- Due to the limited information (a sample of size one at each recurrence time), the nonparametric estimate for $\widehat{\text{Var}}[\widehat{\mu}(t)]$ used in the multiple systems case can't be used for single systems.

Nonparametric Comparison of Two Samples of Recurrence Data

- Suppose that there are two independent samples of recurrence data with mean cumulative functions given by $\mu_1(t)$ and $\mu_2(t)$, respectively.
- Let $\Delta_\mu(t)$ represent the mean cumulative difference at t.
- A nonparametric estimate of $\Delta_\mu(t)$ is
$$\widehat{\Delta}_\mu(t) = \widehat{\mu}_1(t) - \widehat{\mu}_2(t)$$
with estimated variance given by
$$\widehat{\text{Var}}[\widehat{\Delta}_\mu(t)] = \widehat{\text{Var}}[\widehat{\mu}_1(t)] + \widehat{\text{Var}}[\widehat{\mu}_2(t)].$$
- An approximate $100(1-\alpha)\%$ confidence interval for $\Delta_\mu(t)$ is
$$\left[\underset{\sim}{\Delta}_\mu, \quad \tilde{\Delta}_\mu\right] = \left[\widehat{\Delta}_\mu - z_{(1-\alpha/2)}\widehat{\text{se}}_{\widehat{\Delta}_\mu}, \quad \widehat{\Delta}_\mu + z_{(1-\alpha/2)}\widehat{\text{se}}_{\widehat{\Delta}_\mu}\right].$$

Parametric Methods for Analyzing Recurrence Data

Some important parametric models:

- Poisson processes:
  - Homogeneous (HPP).
  - Nonhomogeneous (NHPP).
- Renewal processes (RP).
- Superimposed renewal processes (SRP).

Poisson Processes

Poisson processes provide a simple parametric model for the analysis of point-process recurrence data.

- A point process on $[0, \infty)$ is said to be a Poisson process if it satisfies the following three conditions:
  - $N(0) = 0$.
  - The number of recurrences occurring on disjoint time intervals are independent (independent increments).
  - The recurrence rate, $\nu(t)$, is positive and such that $\mu(a,b) = \text{E}[N(a,b)] = \int_a^b \nu(u)du < \infty$, when $0 \le a < b < \infty$.
- For a Poisson process, it follows that the number of recurrences in $(a, b]$, say $N(a,b)$, is Poisson distributed with pdf
$$\Pr\left[N(a,b) = d\right] = \frac{[\mu(a,b)]^d}{d!} \exp\left[-\mu(a,b)\right], \; d = 0, 1, \ldots$$

Homogeneous Poisson Processes

A **homogeneous** Poisson process (HPP) is a Poisson process with a **constant** recurrence rate, say $\nu(t) = 1/\theta$. In this case:

- $N(a,b)$ has a Poisson distribution with parameter $\mu(a,b) = (b-a)/\theta$.
- The expected number of recurrences in $(a, b]$ is $\mu(a,b) = (b-a)/\theta$. Equivalently the expected number of recurrences per unit time over $(a, b]$ is constant and equal to $1/\theta$ (stationary increments).
- The times between recurrences, $\tau_j = T_j - T_{j-1}$, are independent and identically distributed each with an $\text{EXP}(\theta)$ distribution. This follows directly from the relationship
$$\Pr(\tau_j > t) \;=\; \Pr\left[N(T_{j-1}, T_j) = 0\right] = \exp(-t/\theta).$$
- Then the time to the kth recurrence has a $\text{GAM}(\theta, k)$ distribution.

Nonhomogeneous Poisson Processes

A **nonhomogeneous** Poisson process (NHPP) is a Poisson process with a nonconstant recurrence rate.

- In this case the times between recurrence are neither independent nor identically distributed.
- The expected number of recurrences per unit time over $(a, b]$ is
$$\frac{\mu(a,b)}{b-a} \;=\; \frac{1}{b-a}\int_a^b \nu(u)du$$
- Model is often specified in terms of the recurrence rate $\nu(t)$.
- Here we suppose that $\nu(u) = \nu(u; \boldsymbol{\theta})$ is a known function of an unknown vector of parameters $\boldsymbol{\theta}$.

NHPP Power Recurrence Rate Model

- The power recurrence rate model is

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

- The corresponding mean cumulative number of recurrences over $(0,t]$ is

$$\mu(t;\beta,\eta) \;=\; \left(\frac{t}{\eta}\right)^{\beta}$$

- $\beta=1$ implies an HPP.

NHPP Loglinear Recurrence Rate Model

- The loglinear recurrence rate is

$$\nu(t;\gamma_0,\gamma_1) \;=\; \exp(\gamma_0+\gamma_1 t).$$

- The corresponding mean cumulative number of recurrences over $(0,t]$ is

$$\mu(t;\gamma_0,\gamma_1) \;=\; \frac{\exp(\gamma_0)}{\gamma_1}\left[\exp(\gamma_1 t)-1\right]$$

- When $\gamma_1=0$, $\nu(t;\gamma_0,0)=\exp(\gamma_0)$ which implies an HPP.

Renewal Processes

Definition: A sequence of recurrences $T_1, T_2, \ldots$ is a renewal process if the time between recurrences $\tau_j = T_j - T_{j-1}$, $j=1,2,\ldots$ ($T_0=0$) are independent and identically distributed.

To avoid trivialities we suppose that $\Pr(T_1=0)\neq 1$.

- The HPP is a renewal process but the NHPP is not.
- Some questions of interest include:
  - the distribution of the τ_j's.
  - the distribution of the time until the kth recurrence $k=1,2,\ldots$.
  - the number of occurrences or renewals $N(t)$ in the interval $(0,t]$ and the associated recurrence rate.
  - prediction of future recurrences in a given time interval.

Inferences with Data From a Renewal Process

- If a renewal process provides an adequate model for recurrences, the techniques for single distribution analysis can be applied to model the times between recurrences.
- For example, Lognormal, Weibull, or other distribution used in Chapters 4 - 5, 7 - 11 can be used in this case to model the times between recurrences.

Superimposed Renewal Processes (SRP)

- **Definition:** Consider a collection of n independent renewal processes. The union of all the events from these processes is a **superimposed** (SRP) renewal process.
- In general a SRP is not a renewal process (unless it is an HPP).
- **Drenick's Theorem:** Under mild regularity conditions, when n is large and the system has run long enough to eliminate transients, a SRP behaves as an HPP.
  - This is a kind of central limit theorem for renewal processes. And it is sometimes used to justify the use of the exponential distribution to model times between system failures in large repairable systems.
  - Large samples and long times needed for good approximations.
  - A generalization (Khinchin's Theorem) shows, again under mild conditions, that for a large number of systems, the SRP will converge to an NHPP.

Tools for Checking Point Process Assumptions

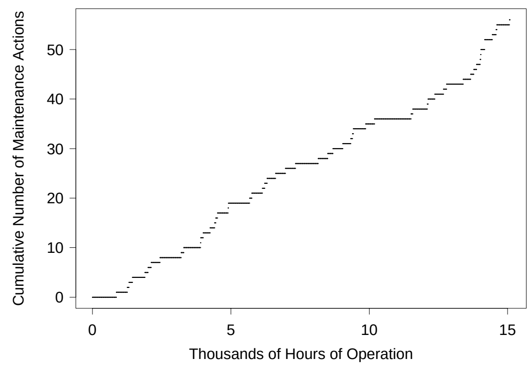
- Cumulative number of recurrences versus time (special case of MCF plot with only one unit). Nonlinearity in this plot indicates non-identically distributed interrecurrence times, which for Poisson processes indicates a nonconstant recurrence rate.
- Plot of times between recurrences versus unit age or **time series plot** of times between recurrences versus recurrence number. Look for trends or cycles to indicate non-identically distributed interrecurrences times.
- Plot of time between recurrences versus lagged time between recurrences to see if times between recurrences have autocorrelation (a form of non-independence).

Data plots will also tend to reveal features of the data or the process that might otherwise escape detection.

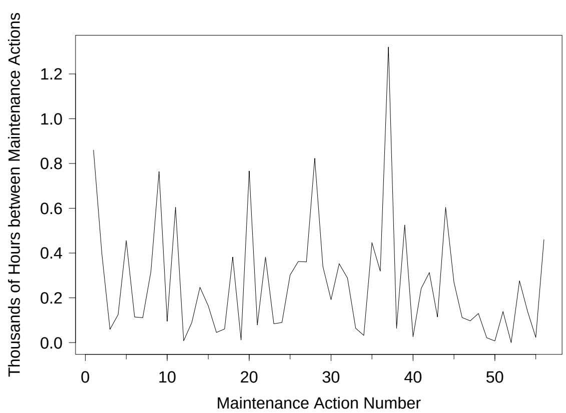
Times of Unscheduled Maintenance Actions for a USS Grampus Diesel Engine

- Unscheduled maintenance actions caused by failure of imminent failure.
- Unscheduled maintenance actions are inconvenient and expensive.
- Data available for 16,000 operating hours.
- Data from Lee (1980).
- Is the system deteriorating (i.e., are failures occurring more rapidly as the system ages)?
- Can the occurrence of unscheduled maintenance actions be modeled by an HPP?

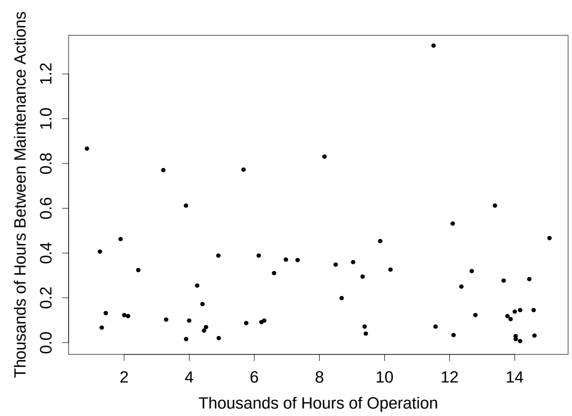
Cumulative Number of Unscheduled Maintenance Actions Versus Operating Hours for a USS Grampus Diesel Engine Lee (1980)



Times Between Unscheduled Maintenance Actions Versus Maintenance Action Number for a USS Grampus Diesel Engine



Times Between Unscheduled Maintenance Actions Versus Engine Operating Hours for a USS Grampus Diesel Engine



Assessing Independence of Times Between Recurrences

Before modeling data as a Poisson process it is necessary to check that the assumption of independent inter-recurrence times is consistent with the data.

- Plot the times between recurrences τ_i versus τ_{i+k} for several values of k. If times between recurrences are independent, then these plots should not show any trend.
- The serial correlation coefficient of lag-k which is defined as

$$\rho_k \;=\; \text{Cov}(\tau_j, \tau_{j+k}) / \sqrt{\text{Var}(\tau_j)\text{Var}(\tau_{j+k})}.$$

Serial Correlation Estimate

- If $\tau_1, \ldots, \tau_r$ are observed time between recurrences then

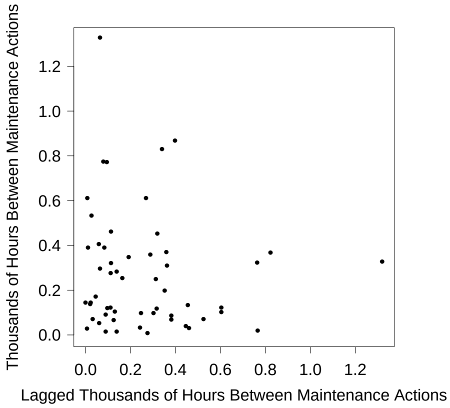
$$\widehat{\rho}_k \;=\; \frac{\sum_{j=1}^{r-k}(\tau_j - \bar{\tau})(\tau_{j+k} - \bar{\tau})}{\sqrt{\sum_{j=1}^{r-k}(\tau_j - \bar{\tau})^2 \; \sum_{j=1}^{r-k}(\tau_{j+k} - \bar{\tau})^2}}$$

where

$$\bar{\tau} = \frac{\sum_{j=1}^{r} \tau_j}{r}.$$

When $\rho_k = 0$ and r large $\sqrt{r-k} \times \widehat{\rho}_k \,\dot{\sim}\, \text{NOR}(0, 1)$ which is used to assess deviations from 0.

USS Grampus Diesel Engine Plot of Times Between Unscheduled Maintenance Actions Versus Lagged Times Between Unscheduled Maintenance Actions



Military Handbook Test (MIL-HDBk-189, 1981)

A simple method of testing $\beta = 1$ against $\beta \neq 1$ in the power recurrence rate model is based on the fact that under the null hypothesis of an HPP and conditional on the number of recurrences r

$$\frac{2r}{\widehat{\beta}} \sim \chi^2_{(2r)}$$

This follows directly from the following:

- Under the assumption of a HPP and conditional on r
$$\frac{t_1}{t_a} < \ldots < \frac{t_r}{t_a}$$
are distributed as the order statistics from a uniform in $(0, 1)$.

- Then under the HPP model,
$$X^2_{\text{MHB}} = -2 \sum_{j=1}^{r} \log(t_j/t_a) = 2r/\widehat{\beta} \sim \chi^2_{(2r)}.$$

Laplace Test for Trend

- Laplace's test has a similar basis for testing for trend in the log-linear recurrence rate NHPP model.

- In this case if the underlying process is HPP ($\gamma_1 = 0$)
$$Z_{\text{LP}} \;=\; \frac{\sum_{j=1}^{r} t_j/t_a - r/2}{\sqrt{r/12}}$$
follows a NOR$(0, 1)$ distribution.

- Values of Z_{LP} in excess of $z_{(1-\alpha/2)}$ provide evidence of a nonconstant recurrence rate.

- This is a powerful test for testing HPP versus NHPP with a log-linear recurrence rate.

Lewis-Robinson Test for Trend

- Both X^2_{MHB} and the Z_{LP} test can give misleading results when the underlying process is a renewal process but is not an HPP.

- The Lewis-Robinson test for trend uses
$$Z_{\text{LR}} = Z_{\text{LP}} \times \frac{\bar{\tau}}{S_\tau}$$
where $\bar{\tau}$ and S_τ are, respectively, the sample mean and standard deviation of the times between recurrence.

- In large samples, Z_{LR} follows approximately a NOR$(0, 1)$ distribution if the underlying process is a renewal process.

- Z_{LR} was derived from heuristic arguments to allow for non-exponential times between recurrences by adjusting for a different coefficient of variation

- Lawless and Thiagarajah (1996) indicate that Z_{LR} is preferable to Z_{LP} as a general test of trend in point process data.

The NHPP Likelihood - Single Unit

- With **interval** recurrence data.

  Suppose that the unit has been observed for a period $(0, t_a]$ and the data are the number of recurrences $d_1, \ldots, d_m$ in the nonoverlapping intervals $(t_0, t_1]$, $(t_1, t_2]$, $\ldots$, $(t_{m-1}, t_m]$ (with $t_0 = 0$, $t_m = t_a$).

$$\begin{aligned}
L(\boldsymbol{\theta}) &= \Pr\left[N(t_0, t_1) = d_1, \ldots, N(t_{m-1}, t_m) = d_m\right] \\
&= \prod_{j=1}^{m} \Pr\left[N(t_{j-1}, t_j) = d_j\right] \\
&= \prod_{j=1}^{m} \frac{\left[\mu(t_{j-1}, t_j; \boldsymbol{\theta})\right]^{d_j}}{d_j!} \exp\left[-\mu(t_{j-1}, t_j; \boldsymbol{\theta})\right] \\
&= \prod_{j=1}^{m} \frac{\left[\mu(t_{j-1}, t_j; \boldsymbol{\theta})\right]^{d_j}}{d_j!} \times \exp\left[-\mu(t_0, t_a; \boldsymbol{\theta})\right]
\end{aligned}$$

The NHPP Likelihood (Continued)

- If the number of intervals m increases and there are **exact** recurrences at $t_1 \leq \ldots \leq t_r$ (here $r = \sum_{j=1}^{m} d_j$, $t_0 \leq t_1$, $t_r \leq t_a$), then using a limiting argument it follows that the likelihood in terms of the density approximation is
$$L(\boldsymbol{\theta}) \;=\; \prod_{j=1}^{r} \nu(t_j; \boldsymbol{\theta}) \times \exp\left[-\mu(0, t_a; \boldsymbol{\theta})\right]$$

- For simplicity, above we assumed that observation is continuous and thus the observation intervals are contiguous.

- In both cases (the interval data or exact recurrences data) the same methods used in Chapters 7, 8 can be used to obtain the ML estimate $\widehat{\boldsymbol{\theta}}$ and confidence regions for $\boldsymbol{\theta}$ or functions of $\boldsymbol{\theta}$.

The NHPP Likelihood with (Possibly) Noncontiguous Observation Windows - Single Unit

- Suppose that there are recurrence data within windows of observation. Suppose that the unit has been observed intermittently over the period $(0, t_a]$ and the data are the number of recurrences $d_1, \dots, d_p$ in the nonoverlapping windows of observation $(t_{1L}, t_{1U}]$, $(t_{2L}, t_{2U}]$, ..., $(t_{pL}, t_{pU}]$ (with $t_{1L} \geq 0, t_{(k-1)U} \leq t_{kL}, t_{pU} \leq t_a$). Any recurrences outside of these windows are not recorded. The likelihood is

$$\begin{aligned} L(\boldsymbol{\theta}) &= \Pr\left[N(t_{1L}, t_{1U}) = d_1, \dots, N(t_{pL}, t_{pU}) = d_p\right] \\ &= \prod_{k=1}^{p} \Pr\left[N(t_{kL}, t_{kU}) = d_k\right] \\ &= \prod_{k=1}^{p} \frac{[\mu(t_{kL}, t_{kU}; \boldsymbol{\theta})]^{d_k}}{d_k!} \exp\left[-\mu(t_{kL}, t_{kU}; \boldsymbol{\theta})\right] \end{aligned}$$

The NHPP Likelihood with (Possibly) Noncontiguous Observation Windows with Intervals Within Windows - Single Unit

- The above can be generalized to smaller intervals within the observation windows (e.g., when there are multiple inspections within each window). Suppose that within window i there are m_k contiguous intervals $(t_{k,0}, t_{k,1}]$, $(t_{k,1}, t_{k,2}]$, ..., $(t_{k,m-1}, t_{k,m}]$ (with $t_{k,0} = t_{mL}$, $t_{k,m_k} = t_{mU}$) and the data are the number of recurrences $d_{k,j}$ in interval $(t_{k,j-1}, t_{k,j}], k = 1, \dots, n$, and $j = 1, \dots, m_k$. The likelihood is

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{k=1}^{p} \prod_{j=1}^{m_k} \Pr\left[N(t_{k,j-1}, t_{k,j}) = d_{k,j}\right] \\ &= \prod_{k=1}^{p} \prod_{j=1}^{m_k} \frac{\left[\mu(t_{k,j-1}, t_{k,j}; \boldsymbol{\theta})\right]^{d_{k,j}}}{d_{k,j}!} \exp\left[-\mu(t_{kL}, t_{kU}; \boldsymbol{\theta})\right] \end{aligned}$$

The NHPP Likelihood with (Possibly) Noncontiguous Observation Windows with Exact Failures Within Windows - Single Unit

- Using a limiting argument similar to the one used for continuous observation with the observation window covering period $(0, t_a]$, if there are **exact** recurrences at $t_1 \leq \dots \leq t_r$, within the n windows, the likelihood in terms of the density approximation is

$$L(\boldsymbol{\theta}) = \left\{\prod_{j=1}^{r} \nu(t_j; \boldsymbol{\theta})\right\} \left\{\prod_{k=1}^{p} \exp\left[-\mu(t_{kL}, t_{kU}; \boldsymbol{\theta})\right]\right\}$$

The NHPP Likelihood for Multiple Systems

- We suppose that there are n independent NHPP system with the same intensity function.

- The overall likelihood is simply the product of the likelihoods for the individual units

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} L_i(\boldsymbol{\theta})$$

The NHPP with Power Recurrence Rate and Exact Recurrence Times

- The likelihood for a single system is

$$L(\beta, \eta) = \left(\frac{\beta}{\eta^{\beta}}\right)^{r} \prod_{j=1}^{r} t_j^{\beta-1} \times \exp\left[-\mu(t_a; \beta, \eta)\right]$$

- The ML estimates of the parameters are:

$$\begin{aligned} \widehat{\beta} &= \frac{r}{\sum_{j=1}^{r} \log\left(t_a/t_j\right)} \\ \widehat{\eta} &= \frac{t_a}{r^{1/\widehat{\beta}}} \end{aligned}$$

- The relative likelihood is

$$R(\beta, \eta) = \left(\frac{\beta}{\widehat{\beta}} \times \frac{\widehat{\eta}^{\widehat{\beta}}}{\eta^{\beta}}\right)^{r} \left(\prod_{j=1}^{r} t_j\right)^{\beta-\widehat{\beta}} \exp\left[r - \mu(t_a; \beta, \eta)\right]$$

NHPP with a Loglinear Recurrence Rate and Exact Recurrence Times

- The likelihood for a single system is

$$L(\gamma_0, \gamma_1) = \exp\left(r\gamma_0 + \gamma_1 \sum_{j=1}^{r} t_j\right) \times \exp\left[-\mu(t_a; \gamma_0, \gamma_1)\right]$$

- The ML estimates are obtained by solving

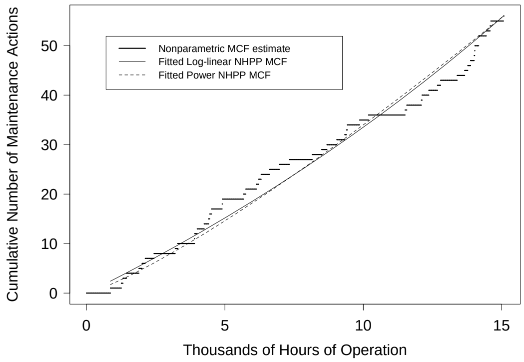
$$\sum_{j=1}^{r} t_j + \frac{r}{\widehat{\gamma_1}} - \frac{r t_a \exp(\widehat{\gamma_1} t_a)}{\exp(\widehat{\gamma_1} t_a) - 1} = 0$$

$$\exp(\widehat{\gamma_0}) = \frac{r\widehat{\gamma_1}}{\exp(t_a \widehat{\gamma_1}) - 1}$$

- The relative likelihood is

$$\begin{aligned} R(\gamma_0, \gamma_1) &= \exp\left[r(\gamma_0 - \widehat{\gamma_0}) + (\gamma_1 - \widehat{\gamma_1}) \sum_{j=1}^{r} t_j\right] \times \\ &\quad \exp\left\{r - \mu(t_a; \gamma_0, \gamma_1)\right\} \end{aligned}$$

Cumulative Number of Unscheduled Maintenance Actions Versus Operating Hours with Power and Loglinear NHPP Models for a USS Grampus Diesel Engine



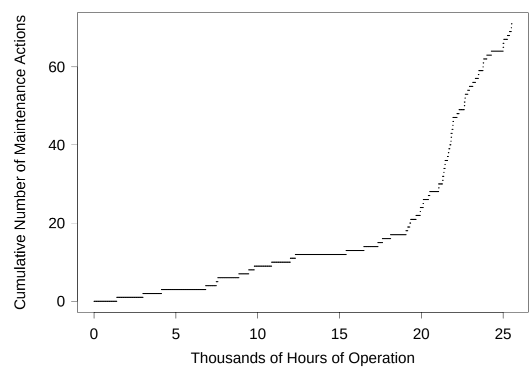
Results of Fitting NHPP Models to the USS Grampus Diesel Engine Data

- Both models seem to fit the data very well.
- For the power recurrence rate model, $\widehat{\beta}$=1.22 and $\widehat{\eta}$ =0.553.
- For the loglinear recurrence rate model, $\widehat{\gamma}_0$=1.01 and $\widehat{\gamma}_1$ =.0377.
- Times between recurrences are consistent with a HPP:
  - the Lewis-Robinson test gave $Z_{\text{LR}} = 1.02$ with p-value $p = .21$.
  - the MIL-HDBk-189 test gave $X^2_{\text{MHB}} = 92$ with p-value $p = .08$.

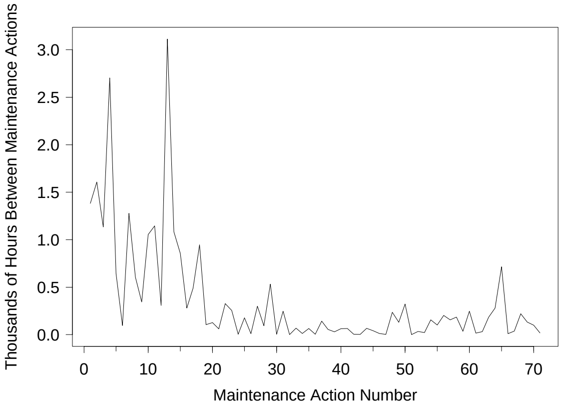
Times Between Unscheduled Maintenance Actions for a USS Halfbeak Diesel Engine

- Unscheduled maintenance actions caused by failure or imminent failure
- Unscheduled maintenance actions are in convenient and expensive
- Data available for 25,518 operating hours.
- Data from Ascher and Feingold (1984, page 75)
- Is the system deteriorating (i.e., are failures occurring more rapidly as the system ages)?
- Can the occurrence of unscheduled maintenance actions be modeled by an HPP?

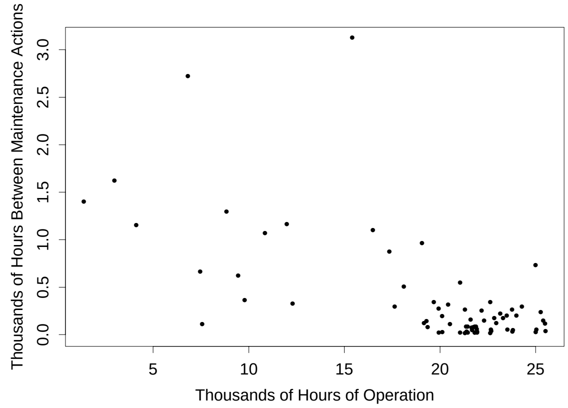
Cumulative Number of Unscheduled Maintenance Actions Versus Operating Hours for a USS Halfbeak Diesel Engine Ascher and Feingold (1984)



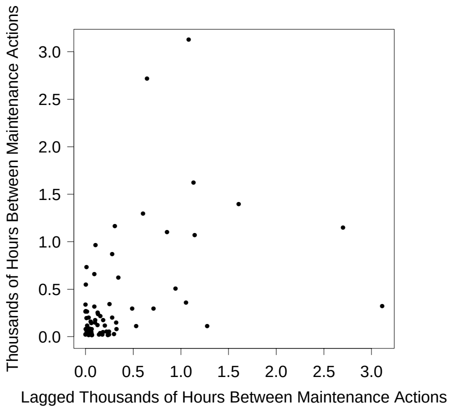
Times Between Unscheduled Maintenance Actions Versus Maintenance Action Number for a USS Halfbeak Diesel Engine Versus



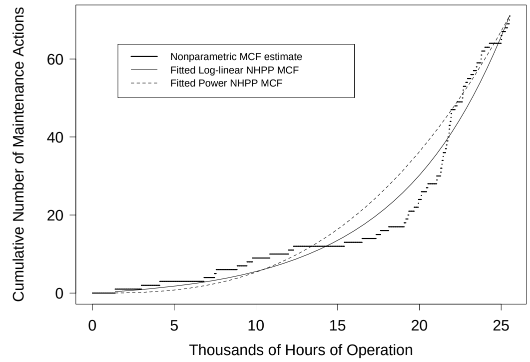
Times Between Unscheduled Maintenance Actions Versus Engine Operating Hours for a USS Halfbeak Diesel Engine



USS Halfbeak Diesel Engine Plot of Times Between Unscheduled Maintenance Actions Versus Lagged Times Between Unscheduled Maintenance Actions



Cumulative Number of Unscheduled Maintenance Actions Versus Operating Hours with Power and Loglinear NHPP Models for a USS Halfbeak Diesel Engine



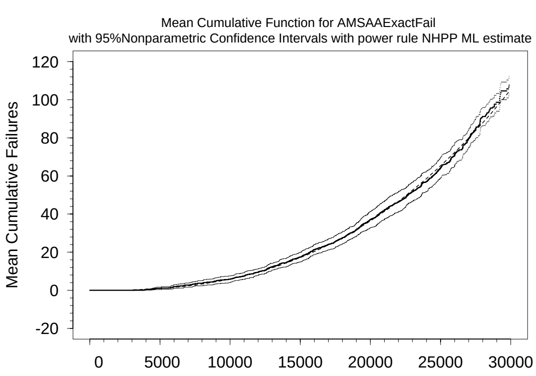
Results of Fitting NHPP Models to the USS Halfbeak Diesel Engine Data

- Both models seem to fit the data reasonably well, but the loglinear recurrence rate model fits better than the power recurrence rate.
- For the power recurrence rate model, $\widehat{\beta}$=2.76 and $\widehat{\eta}$ =5.45.
- For the loglinear recurrence rate model, γ_0=−1.43 and γ_1 =.149.
- The evidence against an HPP is strong:
  - ▶ the Lewis-Robinson test gave $Z_{\text{LR}} = 4.70$ with p-value =0.
  - ▶ the MIL-HDBk-189 test gave $X^2_{\text{MHB}} = 51$ with p-value $= 0$.

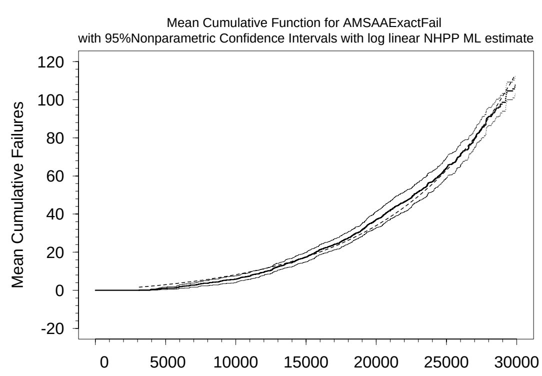
Prediction of Future Recurrences with a Poisson Process

- The expected number of recurrences in an interval $[a, b]$ is $\int_a^b \nu(u, \boldsymbol{\theta}) du$. Then the ML point prediction estimate is $\int_a^b \nu(u, \widehat{\boldsymbol{\theta}}) du$.
- A point prediction for the power recurrence rate is
$$\int_a^b \nu(u, \widehat{\boldsymbol{\theta}}) du = \left(\frac{1}{\widehat{\eta}}\right)^{\widehat{\beta}} \left(b^{\widehat{\beta}} - a^{\widehat{\beta}}\right).$$
- A point prediction for the loglinear recurrence rate is
$$\int_a^b \nu(u, \widehat{\boldsymbol{\theta}}) du = \frac{\exp(\widehat{\gamma_0})}{\widehat{\gamma_1}} \left[\exp(\widehat{\gamma_1} b) - \exp(\widehat{\gamma_1} a)\right].$$
- There is a similar expression for the case of a loglinear power recurrence rate.
- Need a method to obtain prediction intervals. Could use bootstrap.

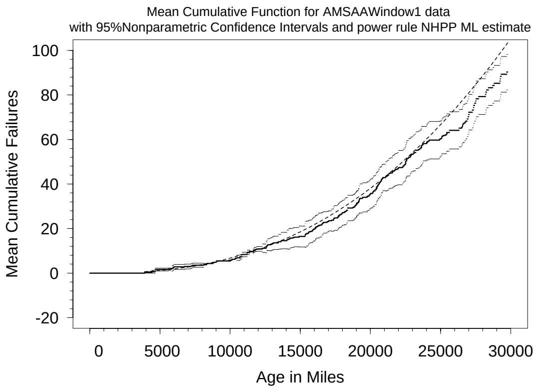
Power Rule NHPP and MCF Plot of the Simulated AMSAA Vehicle Repairs Power Rule $\beta = 2.76$ $\eta = 5447$ Continuous Inspection



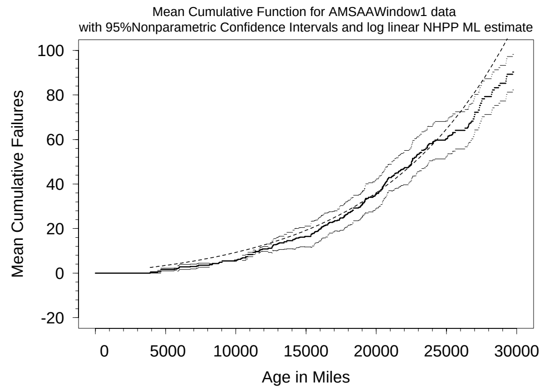
Log-Linear NHPP and MCF Plot of the Simulated AMSAA Vehicle Repairs Power Rule $\beta = 2.76$ $\eta = 5447$ Continuous Inspection



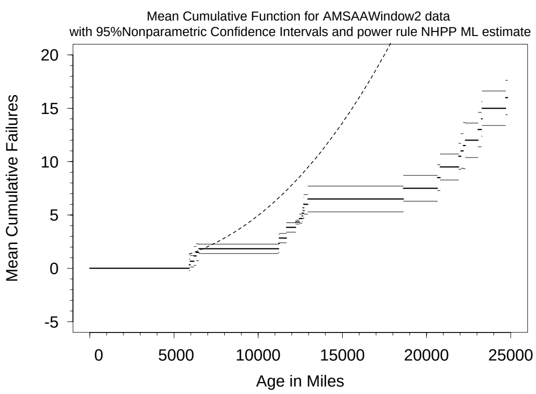
Power Rule NHPP and MCF Plot of the Simulated AMSAA Vehicle Repairs Random Inspection Windows



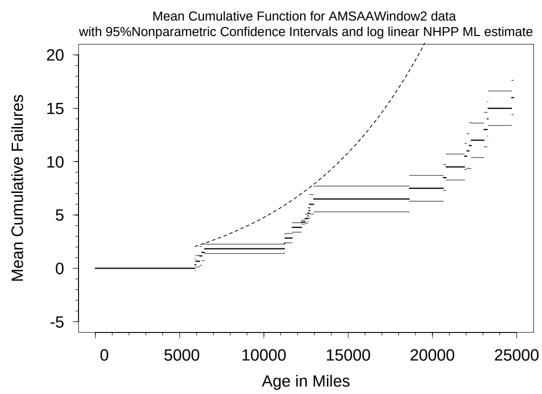
Log-Linear NHPP and MCF Plot of the Simulated AMSAA Vehicle Repairs Random Inspection Windows



Power Rule NHPP and MCF Plot of the Simulated AMSAA Vehicle Repairs Biased Inspection Windows



Log-Linear NHPP and MCF Plot of the Simulated AMSAA Vehicle Repairs Biased Inspection Windows



Comparison of NHPP Estimation Results

True parameters: $\beta = 2.76$ $\eta = 5447$

```
AMSAAExactFail
          MLE  Std.Err. 95% Lower 95% Upper
 eta 5063.071 310.79797   4453.92  5672.223
beta    2.617   0.09519      2.43     2.804
AMSAAWindow1
          MLE Std.Err. 95% Lower 95% Upper
 eta 4686.747 515.5076  3676.370  5697.123
beta    2.509   0.1562     2.202     2.815
AMSAAWindow2
          MLE  Std.Err. 95% Lower 95% Upper
 eta 5263.816 985.9663   3331.36  7196.275
beta    2.494   0.3135      1.88     3.109
```


Other Topics in the Analysis of Recurrence Data

- Adjustment for covariates.
- Reliability growth applications.
- Random effects and mixture models (important unit-to-unit differences)
- Bayesian methods
- Methods of analysis when unit identification is not possible.