Modeling of Dependence Between Critical Failure and Preventive Maintenance: The Repair Alert Model

Bo Lindqvist

Helge Langseth

Bård Støve

bo@math.ntnu.no

Department of Mathematical Sciences

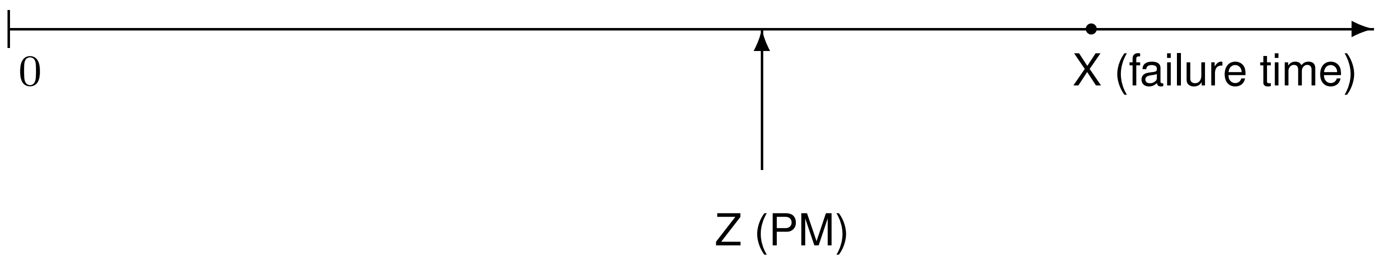
Norwegian University of Science and Technology

Purpose of talk

Explicit modeling of maintenance in repairable systems:

- Several failure mechanisms
- Imperfect repair
- Degraded failures
- Preventive Maintenance (PM) (scheduled/unscheduled)
- Several failure mechanisms
- Periodically tested components

Simultaneous Modeling of Time to Failure and PM



Competing risk problem:

- X is the (potential) failure time of an item
- Z is the time of a (potential) PM action before time X

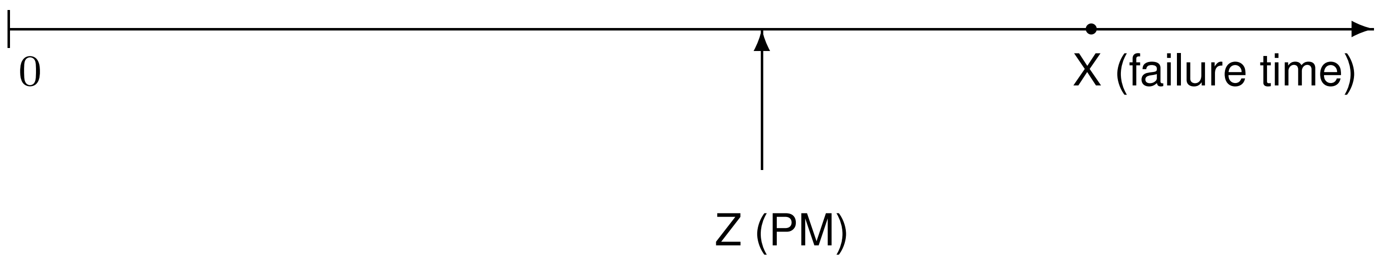
Possible outcomes:

- Failure at X, no PM, Z is not observed
- PM at Z ($< X$), while X is not observed

Typical observations:

- N independent pairs $\{\min(X, Z), I(Z < X)\}$ are observed, represented as $x_1, ..., x_m;\ z_1, ..., z_n$ where $N = m + n$

Cooke's Random Signs Censoring



Definition:

- The event $\{Z < X\}$ is independent of X

Motivation:

- Suppose the item emits some warning of deterioration, *prior to failure.*
  If warning signal is observed, then the item will be preventively maintained and the PM variable Z is observed.

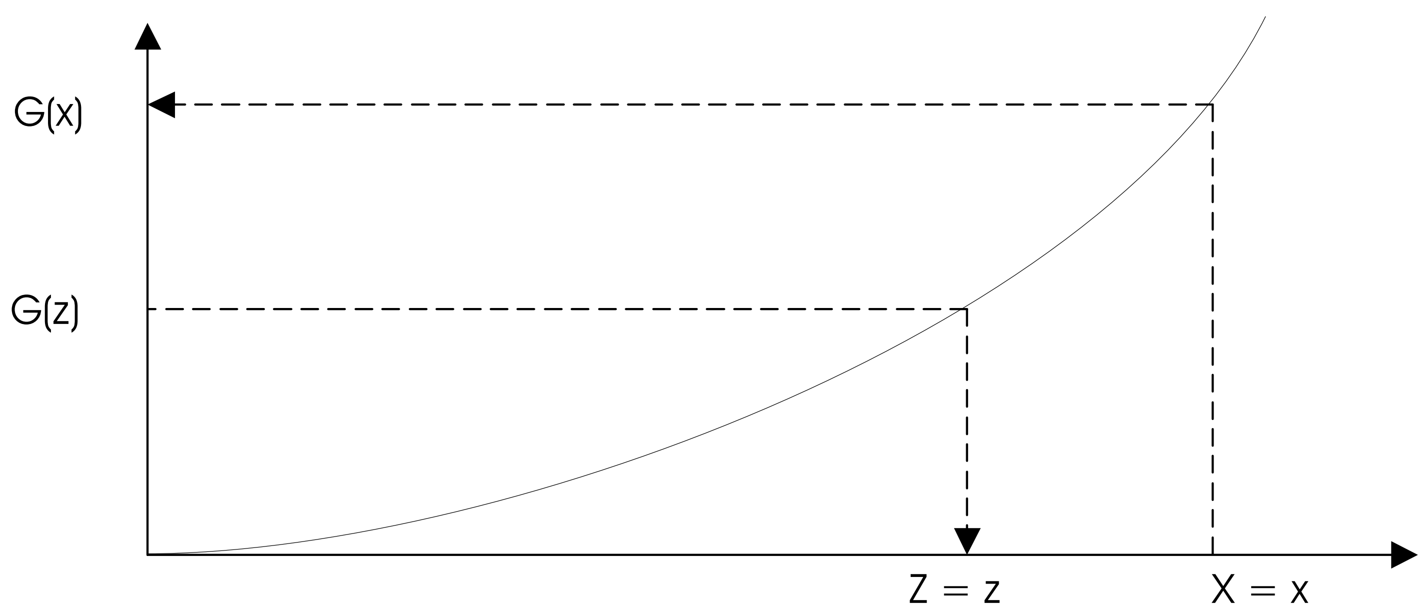
  If the event of observing the signal is independent of the item's age, then *random signs censoring* is appropriate.

Repair Alert Model

Definition:

- The pair (X, Z) satisfies the requirements of the Repair Alert Model provided the following two conditions both hold.
  1. Random signs censoring, that is $\{Z < X\}$ is independent of X
  2. There exists an increasing function $G(x)$ such that for all $x > 0$,
     $P(Z \leq z | Z < X, X = x) = \frac{G(z)}{G(x)}, \quad 0 \leq z \leq x$
     The function $G(z)$ is called *the cumulative repair alert function*. Its derivative $g(z)$ is called *the repair alert function*.

Motivation:



Some Notation

$$\begin{aligned}
F_X(x) &= P(X \leq x) && \text{(Distribution function)} \\
F_Z(z) &= P(Z \leq z) \\
F_X^*(x) &= P(X \leq x, X < Z) && \text{(Subdistribution function)} \\
F_Z^*(z) &= P(Z \leq z, Z < X) \\
\tilde{F}_X(x) &= P(X \leq x | X < Z) && \text{(Conditional subdistribution function)} \\
\tilde{F}_Z(z) &= P(Z \leq z | Z < X)
\end{aligned}$$

- Only the subdistribution functions are in general identifiable from competing risk data
- Under Random Signs Censoring we have by definition

$$\tilde{F}_X(x) = P(X \leq x | X < Z) = P(X \leq x) = F_X(x)$$

making $F_X(x)$ *identifiable.*

Identification of $G(\cdot)$ in Repair Alert Model

Recall assumptions:

- $\{Z < X\}$ is independent of X
- $P(Z \leq z | Z < X, X = x) = \frac{G(z)}{G(x)}, \quad 0 \leq z \leq x$

From this:

$$\begin{aligned}
\tilde{F}_Z(z) &= P(Z \leq z | Z < X) \\
&= \int_0^\infty P(Z \leq z | Z < X, X = x) P(x \leq X \leq x + dx | Z < X) \\
&= \int_0^\infty \min\left(\frac{G(z)}{G(x)}, 1\right) \cdot f_X(x) dx = F_X(z) + G(z) \int_z^\infty \frac{f_X(x)}{G(x)} dx
\end{aligned}$$

Differentiate:

$$\tilde{f}_Z(z) = f_X(z) + g(z) \int_z^\infty \frac{f_X(x)}{G(x)} dx - G(z) \frac{f_X(z)}{G(z)} = g(z) \int_z^\infty \frac{f_X(x)}{G(x)} dx.$$

Combine:

$$\tilde{F}_Z(z) = F_X(z) + G(z) \frac{\tilde{f}_Z(z)}{g(z)}$$

Identification of $G(\cdot)$ in Repair Alert Model

Basic formula:

$$\tilde{F}_Z(z) = F_X(z) + G(z)\frac{\tilde{f}_Z(z)}{g(z)}$$

Rearranging:

$$\frac{g(z)}{G(z)} = \frac{\tilde{f}_Z(z)}{\tilde{F}_Z(z) - F_X(z)}.$$

Integrating from fixed point $a > 0$, assuming $G(a) > 0$:

$$\int_a^w \frac{g(z)}{G(z)} dz = \int_a^w \frac{\tilde{f}_Z(z)}{\tilde{F}_Z(z) - F_X(z)} dz$$

$$\int_{G(a)}^{G(w)} \frac{dy}{y} = \int_{\tilde{F}_Z(a)}^{\tilde{F}_Z(w)} \frac{dy}{y - F_X(\tilde{F}_Z^{-1}(y))}.$$

$$\frac{G(w)}{G(a)} = \exp\{\int_{\tilde{F}_Z(a)}^{\tilde{F}_Z(w)} \frac{dy}{y - F_X(\tilde{F}_Z^{-1}(y))}\}.$$

Hence, $G(z)$ is identifiable from data, modulo a constant.

Nonparametric Estimation of Cumulative Repair Function

Let $x_1, ..., x_m$ and $z_1, ..., z_n$ be the observed X and Z.
$F_X(t)$ is estimated by

(1) $$\hat{F}_X(t) = \frac{i}{m} \;\; for \;\; x_i \leq t < x_{i+1}, \;\; i = 0, 1, ..., m.$$

With $t = \tilde{F}_Z^{-1}(y)$ we get

(2) $$\hat{F}_X(\tilde{F}_Z^{-1}(y)) = \frac{i}{m} \;\; for \;\; \tilde{F}_Z(x_i) \leq y < \tilde{F}_Z(x_{i+1}), \;\; i = 0, 1, ..., m.$$

Thus

(3) $$\int_{\tilde{F}_Z(x_1)}^{\tilde{F}_Z(x_j)} \frac{dy}{y - \hat{F}_X(\tilde{F}_Z^{-1}(y))} = \sum_{i=1}^{j-1} \int_{\tilde{F}_Z(x_i)}^{\tilde{F}_Z(x_{i+1})} \frac{dy}{y - i/m} = \sum_{i=1}^{j-1} \ln \frac{\tilde{F}_Z(x_{i+1}) - i/m}{\tilde{F}_Z(x_i) - i/m}.$$

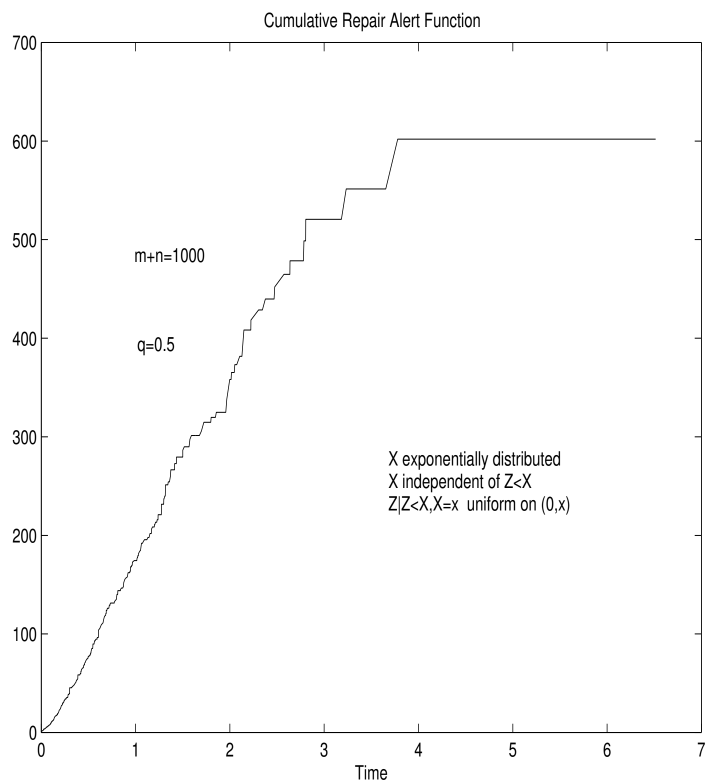
yielding the estimator

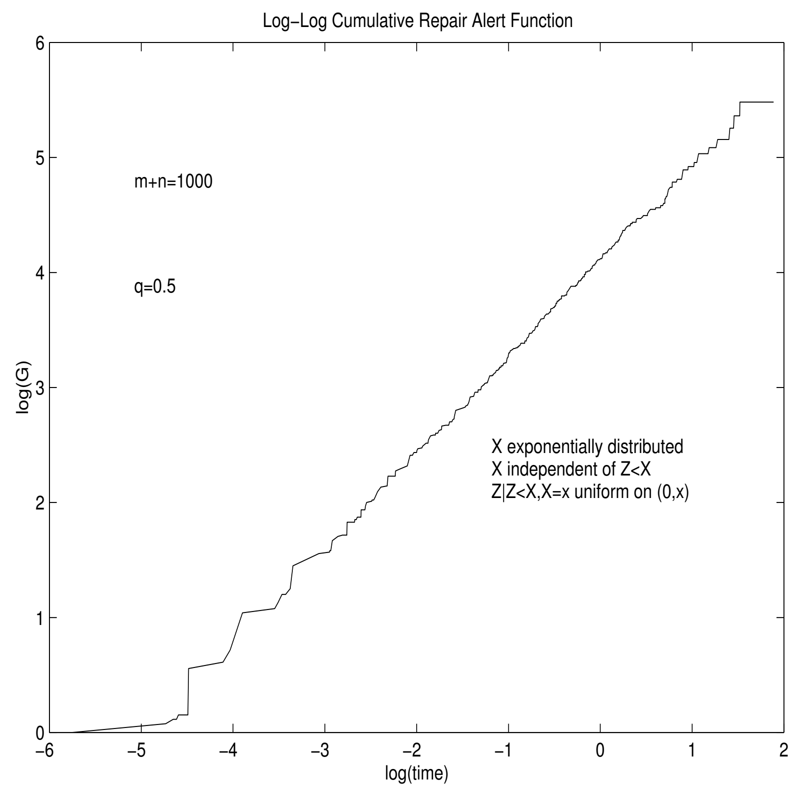
(4) $$\frac{\hat{G}(x_j)}{\hat{G}(x_1)} = \prod_{i=1}^{j-1} \frac{\hat{\tilde{F}}_Z(x_{i+1}) - i/m}{\hat{\tilde{F}}_Z(x_i) - i/m}.$$

The estimator for $\tilde{F}_Z(t)$ is

Simulated example: Repair alert model

- Failure time X is exponentially distributed, failure rate $\lambda = 1$
- $q = P(Z < X) = 0.5$
- Cumulative repair function is $G(x) = x$, i.e. Z given $Z < X, X = x$ is uniform on $(0, x)$
- $m + n = 1000$





Parametric statistical inference for repair alert model

Suppose we observe N independent copies of the pair

$$\{\min(X, Z), I(Z < X)\}$$

This gives data $x_1, ..., x_m;\ z_1, ..., z_n,$
with obvious meaning, where $n + m = N$.

We are interested in estimating

- Density of X, $f_X(x)$ (for example exponential, Weibull, etc.)
- Repair probability q
- Repair function $g(x)$

LIKELIHOOD FUNCTION
Likelihood contribution from an observation is under the repair alert model:

$$\begin{aligned}
f(x_i, X < Z) &= (1-q) f_X(x_i) \quad \text{for } x_i \\
f(z_i, Z < X) &= q \tilde{f}_Z(z_i) \quad \text{for } z_i \\
&= q g(z_i) \int_{z_i}^{\infty} (f_X(x)/G(x)) dx
\end{aligned}$$

An Exponential-Power Repair Alert Model

$$\boxed{f_X(x) = \lambda e^{-\lambda x},\ q = P(Z < X),\ G(x) = x^{\beta}}$$

Likelihood contributions:

$$\begin{aligned}
f(x_i, X < Z) &= (1-q) f_X(x_i) = (1-q)\lambda e^{-\lambda x_i} \\
f(z_i, Z < X) &= q g(z_i) \int_{z_i}^{\infty} (f_X(x)/G(x)) dx \\
&= q\beta z_i^{\beta-1} \int_{z_i}^{\infty} \lambda e^{-\lambda x} x^{-\beta} dx \\
&= q\lambda\beta(\lambda z_i)^{\beta-1} \int_{\lambda z_i}^{\infty} w^{-\beta} e^{-w} dw
\end{aligned}$$

Complete log-likelihood:

$$l(\lambda, \beta, q) = m \ln(1-q) + n \ln q + (n+m) \ln \lambda + n \ln \beta - \lambda \sum_{i=1}^{m} x_i +$$

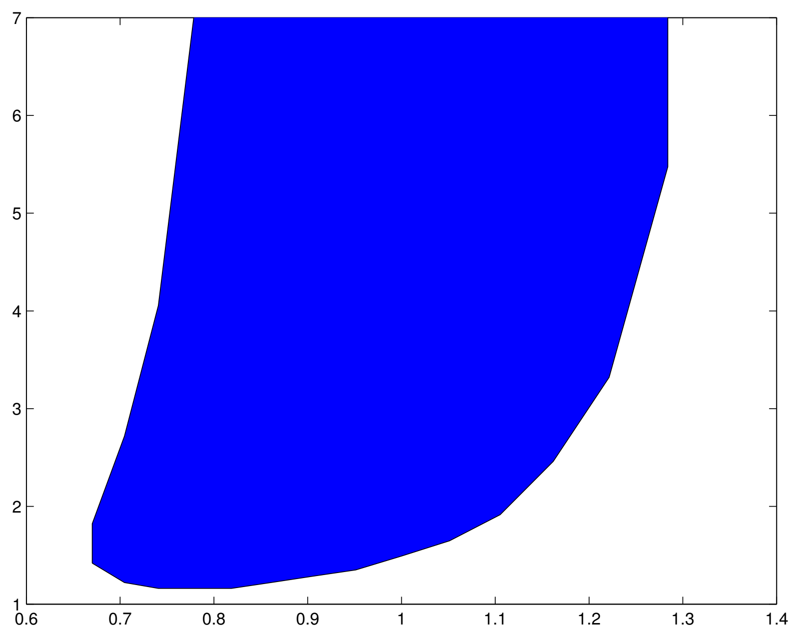
$$\sum_{i=1}^{n} (\beta - 1) \ln(\lambda z_i) + \sum_{i=1}^{n} \ln\left(\int_{\lambda z_i}^{\infty} w^{-\beta} e^{-w} dw\right).$$

Simulated parametric example

- $N = m + n = 100$

Parameter	True value	Estimate	Lower bound	Upper bound
λ	1	0.9838	0.7621	1.2566
β	3	4.5301	1.5010	∞
q	0.5	0.5700	0.4730	0.6670

Maximum likelihood estimates and approximate 95% *confidence intervals for simulated data*



The 95% confidence region for λ (horizontal axis) and β from the simulated data

The EM-algorithm – general description

$\mathcal{Y}$ is the observed data; $\mathcal{X}$ is a piece of unknown data; θ is the parameter of interest; and $l_C(\theta; \mathcal{Y}, \mathcal{X})$ is the hypothetical complete-data log-likelihood, defined for all $\theta \in \Omega$. Starting with an initial parameter value $\theta^{(0)} \in \Omega$, the EM algorithm repeats the following two steps until convergence.

E-step: Compute $l^{(j)}(\theta) = E_{\mathcal{X}|\mathcal{Y},\theta^{(j-1)}}[l_C(\theta; \mathcal{Y}, \mathcal{X})]$, where the expectation is taken with respect to the conditional distribution of the missing data $\mathcal{X}$ given the observed data $\mathcal{Y}$, and the current numerical value $\theta^{(j-1)}$ is used in evaluating the expected value.

M-step: Find $\theta^{(j)} \in \Omega$ that maximizes $l^{(j)}(\theta)$.

The EM-algorithm in our model

Likelihood contributions to complete likelihood are now simplified:

$$
\begin{aligned}
f(x_i, X < Z) &= (1-q) f_X(x_i) = (1-q)\lambda e^{-\lambda x_i} \\
f(x_i, z_i, Z < X) &= q\lambda e^{-\lambda x_i} \frac{\beta z_i^{\beta-1}}{x_i^{\beta}}
\end{aligned}
$$

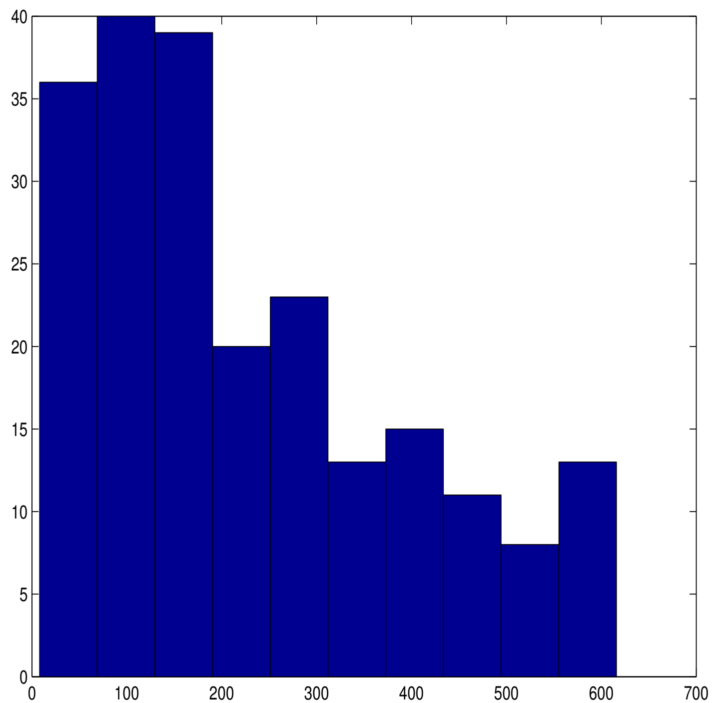
Resulting iterative algorithm:

$$
\begin{aligned}
\hat{q} &= \frac{n}{N} \\
\hat{\lambda}_{j+1} &= \frac{N}{\sum_{i=1}^{m} x_i + (1/\hat{\lambda}_j) \sum_{i=1}^{n} \frac{\int_{\hat{\lambda}_j z_i}^{\infty} w^{-(\hat{\beta}_j - 1)} e^{-w} dw}{\int_{\hat{\lambda}_j z_i}^{\infty} w^{-\hat{\beta}_j} e^{-w} dw}} \\
\hat{\beta}_{j+1} &= \frac{n}{\sum_{i=1}^{n} \left[\frac{\int_{\hat{\lambda}_j z_i}^{\infty} \ln(w) w^{-\hat{\beta}_j} e^{-w} dw}{\int_{\hat{\lambda}_j z_i}^{\infty} w^{-\hat{\beta}_j} e^{-w} dw} - \ln(\hat{\lambda}_j z_i) \right]}
\end{aligned}
$$

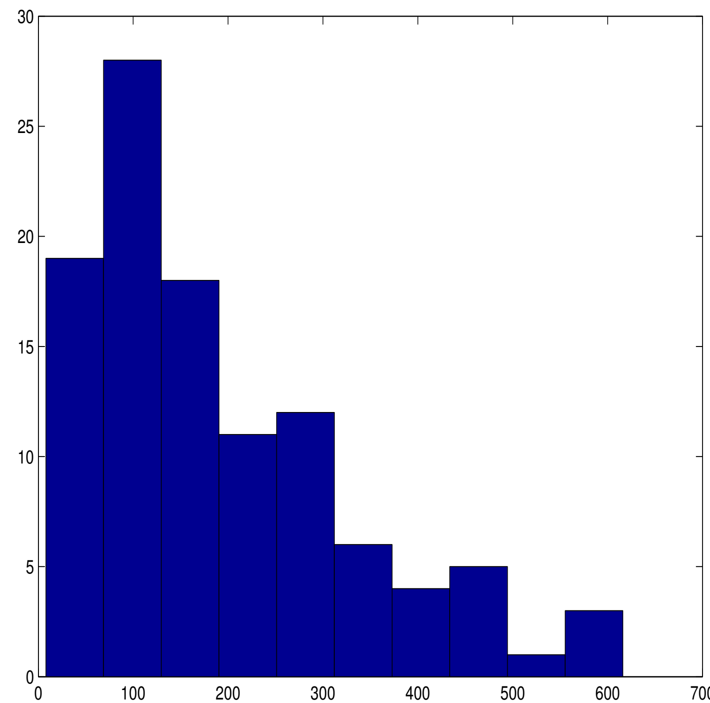
Example: VHF data (Mendenhall and Hader, 1958)

Times to failure for ARC-1 VHF communication transmitter-receivers of a single commercial airline.

- X = time to confirmed failure, $m = 218$
- Z = time to unconfirmed failure (censoring), $n = 107$

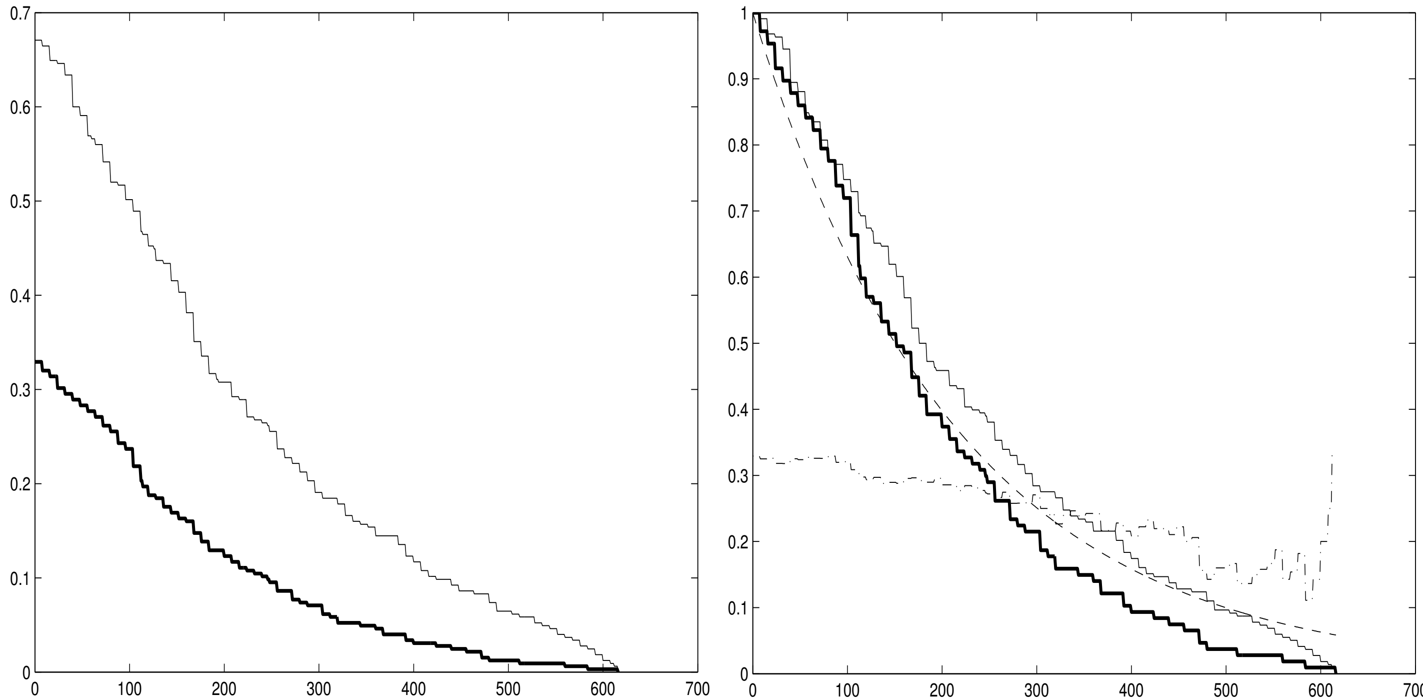


X



Z

Example: VHF data (Mendenhall and Hader, 1958)

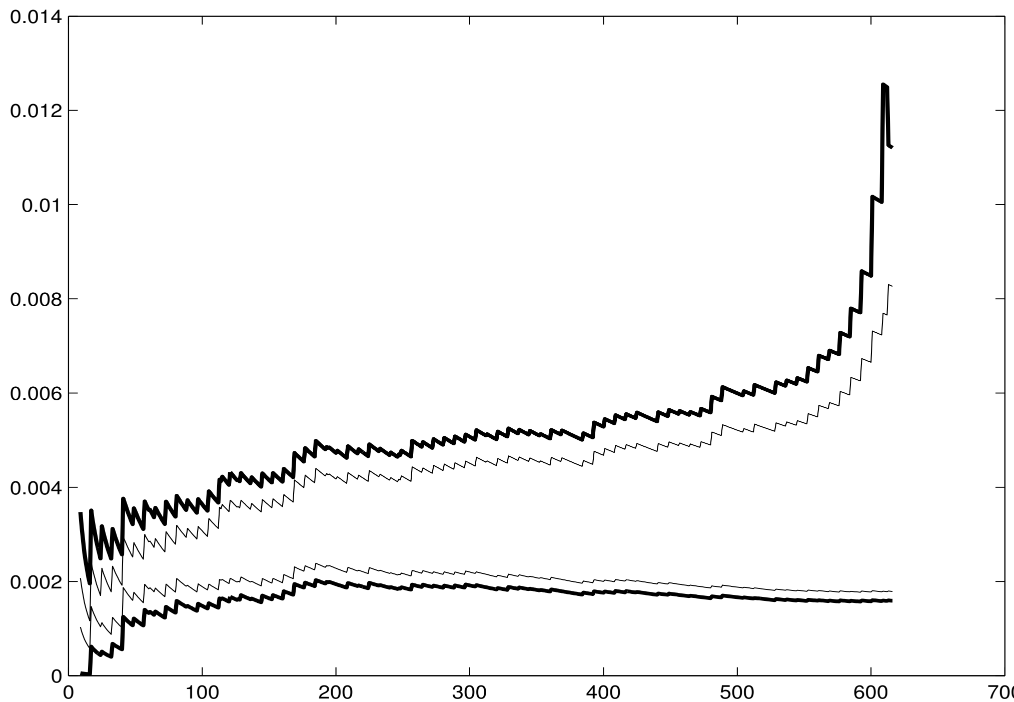


- *Left: Estimated subsurvival functions for X, thin line, and Z, thick line*
- Right: *Estimated conditional subsurvival functions for X, thin line, Z, thick line, estimated $\Phi(t) = P(Z < X|X > t, Z > t)$, and the estimated CSF corresponding to an independent exponential model*

Null hypothesis of independent exponentials is accepted at p-value $\approx .15$

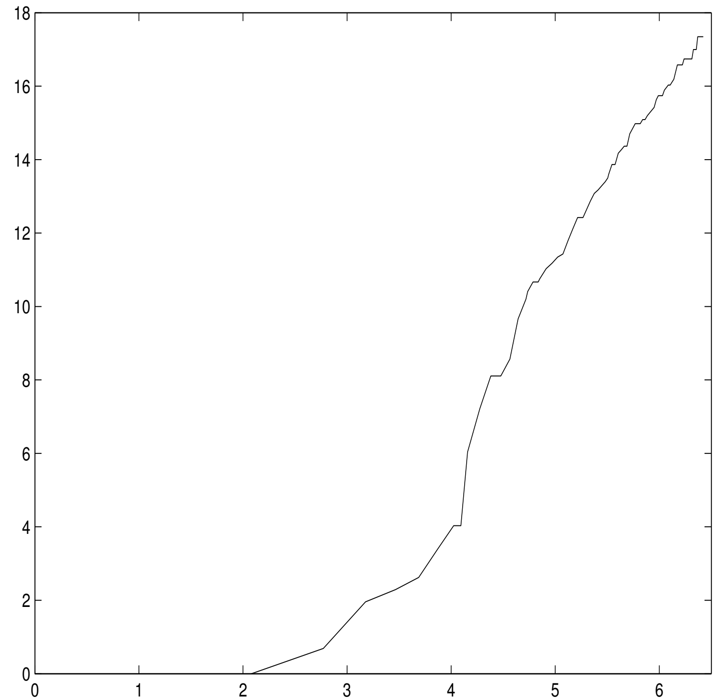
Estimated failure rate for X, VHF data

Assuming	*Failure rate for X*
Repair Alert, X exponentially distributed	$4.353 \cdot 10^{-3}$
X, Z independent exponentials	$3.092 \cdot 10^{-3}$



Bounds for the failure rate of X. Thick line: accounting for sampling fluctuations. Thin line: without sampling fluctuations

VHF data: Nonparametric Repair Alert Model



The estimated $\ln \frac{G(x_j)}{G(x_1)}$ *plotted against* $\ln(x_j)$

VHF data: Parametric Repair Alert Model

MODEL

- Failure time X is exponentially distributed, failure rate λ
- Cumulative repair function is $G(x) = x^{\beta}$

Parameter	Estimate	Lower bound	Upper bound
λ	$4.458 \cdot 10^{-3}$	$3.916 \cdot 10^{-3}$	$5.069 \cdot 10^{-3}$
β	8.9809	3.0345	∞
q	0.3292	0.2781	0.3803

Maximum likelihood estimates and approximate 95% *confidence intervals for parametric repair alert model*