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

Bo Lindqvist

Helge Langseth

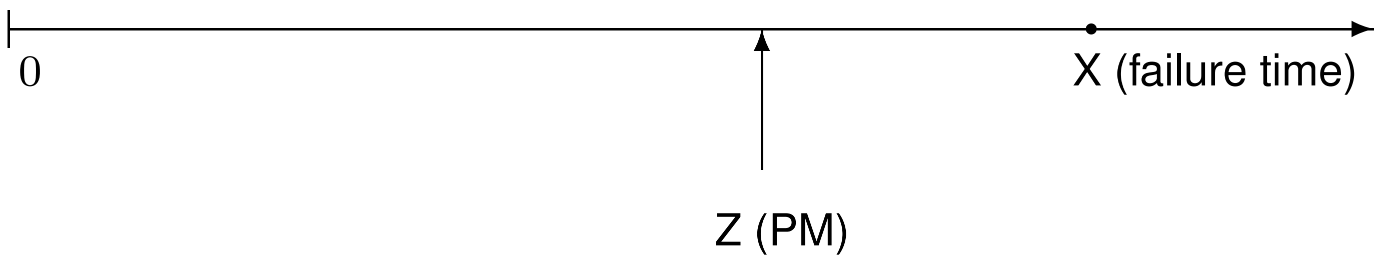
Bård Støve

bo@math.ntnu.no

Department of Mathematical Sciences

Norwegian University of Science and Technology

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

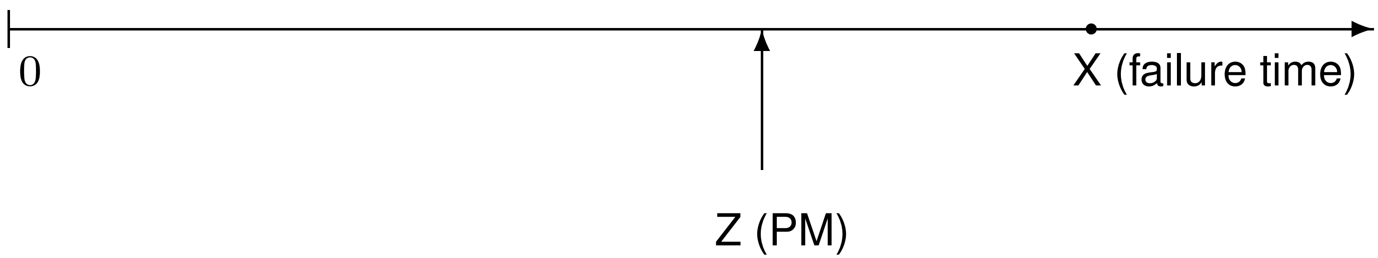
Possible outcomes:

- 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.

Some Notation

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

Some Notation

$$\begin{aligned} S_X(x) &= P(X > x) && \text{(Survival function)} \\ S_Z(z) &= P(Z > z) \\ S^*_X(x) &= P(X > x, X < Z) && \text{(Subsurvival function)} \\ S^*_Z(z) &= P(Z > z, Z < X) \\ \tilde{S}_X(x) &= P(X > x | X < Z) && \text{(Conditional subsurvival function)} \\ \tilde{S}_Z(z) &= P(Z > z | Z < X) \end{aligned}$$

Note that

$$S^*_X(0) + S^*_Z(0) = 1, \quad \tilde{S}_X(x) = \frac{S^*_X(x)}{S^*_X(0)}, \quad \tilde{S}_Z(z) = \frac{S^*_Z(z)}{S^*_Z(0)}$$

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

$$\tilde{S}_X(x) = P(X > x | X < Z) = P(X > x) = S_X(x)$$

so that the marginal distribution of X is identifiable.

A Condition for Random Signs Censoring

Functions H_1, H_2 on $[0, \infty)$ form a subsurvival pair if

- H_1, H_2 are nonnegative, nonincreasing
- $\lim_{t\to\infty} H_1(t) = \lim_{t\to\infty} H_1(t) = 0$
- $H_1(0) + H_2(0) = 1$

THEOREM (Cooke, 1993): Let H_1, H_2 be a pair of continuous strictly monotone functions forming a subsurvival pair. Then the following are equivalent:

- There exists a pair (X, Z) of positive random variables such that
  - $\{Z < X\}$ is independent of X (random signs censoring)
  - $S_X^*(x) = H_1(x)$ for all $x \geq 0$, $S_Z^*(z) = H_2(z)$ for all $z \geq 0$
- $$\frac{H_1(x)}{H_1(0)} > \frac{H_2(x)}{H_2(0)} \text{ for all } x > 0$$

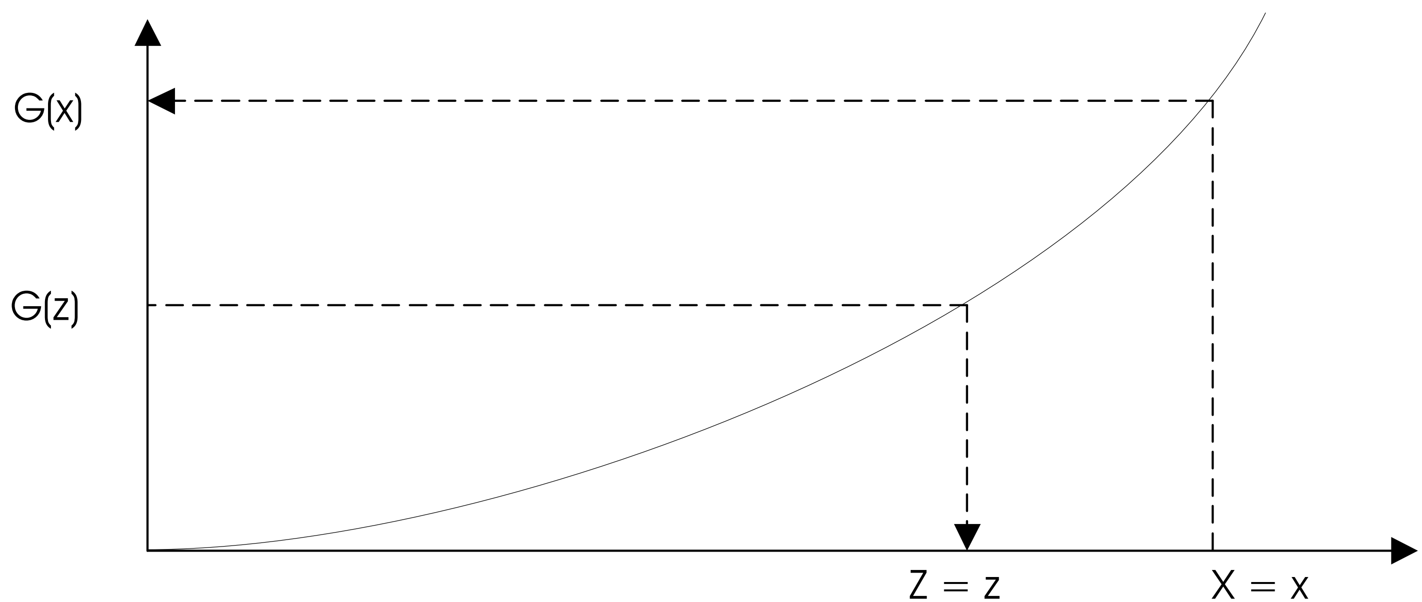
Repair Alert Model

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 $G(x)$ with $G(0) = 0$ such that for all $x > 0$, $P(Z \leq z | Z < X, X = x) = \frac{G(z)}{G(x)}, \;\; 0 \leq z \leq x$

     The function $G(z)$ is called *the cumulative repair alert function*. Its derivative $g(z)$ is called *the repair alert function*.

Motivation:



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

Recall assumptions:

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

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 $a > 0$:

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

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

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

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

Extension of Cooke's theorem

THEOREM: Let H_1, H_2 be a pair of continuous strictly monotone functions forming a subsurvival pair. Then the following are equivalent:

- There exists a pair (X, Z) of positive random variables and a nondecreasing function $G(x)$ with $G(0) = 0$ such that
  - $\{Z < X\}$ is independent of X
  - For all $x > 0$, $P(Z \leq z | Z < X, X = x) = \frac{G(z)}{G(x)}, \ \ 0 \leq z \leq x$
  - $S^*_X(x) = H_1(x)$ for all $x \geq 0$, $S^*_Z(z) = H_2(z)$ for all $z \geq 0$
-

$$\frac{H_1(x)}{H_1(0)} > \frac{H_2(x)}{H_2(0)} \text{ for all } x > 0$$

Nonparametric Estimation of $G(\cdot)$ in Repair Alert Model

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

Goal is to estimate this based on data $x_1, ..., x_m,\ z_1, ..., z_n$.

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

Hence, with $t = \tilde{F}_Z^{-1}(y)$:

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

Thus:

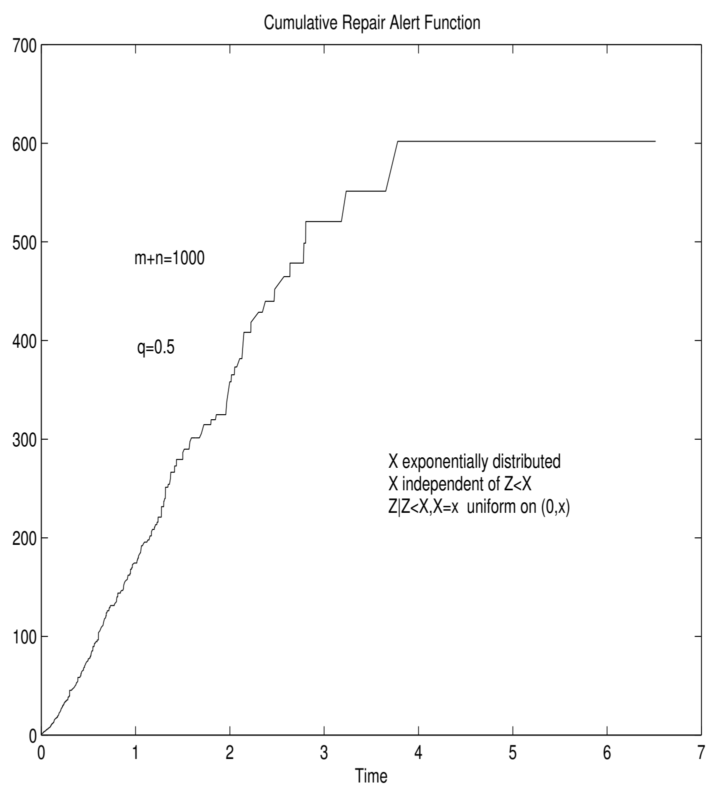
$$\int_{\tilde{F}_Z(x_1)}^{\tilde{F}_Z(x_j)} \frac{dy}{y - \hat{F}_X(\tilde{F}_Z^{-1}(y))} = \sum_{i=1}^{j-1} \int_{\tilde{F}_Z(x_i)}^{\tilde{F}_Z(x_{i+1})} \frac{dy}{y - i/m} = \sum_{i=1}^{j-1} \ln \frac{\tilde{F}_Z(x_{i+1}) - i/m}{\tilde{F}_Z(x_i) - i/m}.$$

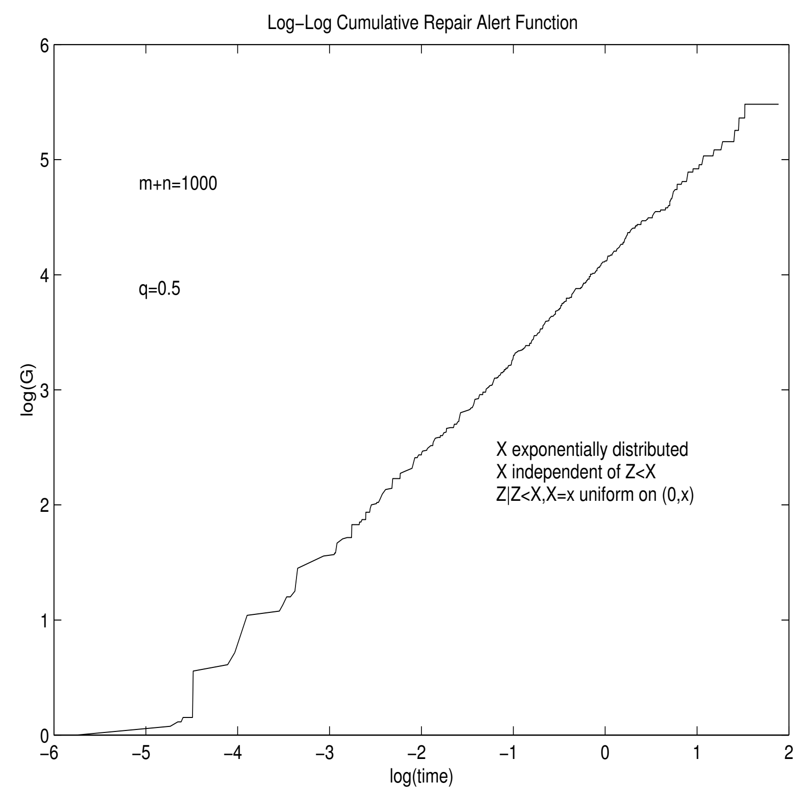
By estimating $\tilde{F}_Z(\cdot)$ this finally gives the estimator:

$$\frac{\hat{G}(x_j)}{\hat{G}(x_1)} = \prod_{i=1}^{j-1} \frac{\hat{\tilde{F}}_Z(x_{i+1}) - i/m}{\hat{\tilde{F}}_Z(x_i) - i/m} = \prod_{i=1}^{j-1} \frac{\#\ \{z_k : z_k \leq x_{i+1}\}/n - i/m}{\#\ \{z_k : z_k \leq x_i\}/n - i/m}$$

Nonparametric Estimation with Simulated Data

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





Parametric Inference for Repair Alert Model

Data:

$$x_1, ..., x_m;\ z_1, ..., z_n$$

We are interested in estimating

- Density of X, $f_X(x)$ (for example exponential, Weibull, etc.)
- Repair probability $q = P(Z < X)$
- Repair alert functions $g(x)$ or $G(x)$

Construction of likelihood function:

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

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

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:

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

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

The EM-algorithm in Exponential-Power Model

Augment data by assuming the X to be observed also when $Z < X$.

Likelihood contributions to complete likelihood are now simplified:

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

The resulting iterative EM-algorithm boils down to:

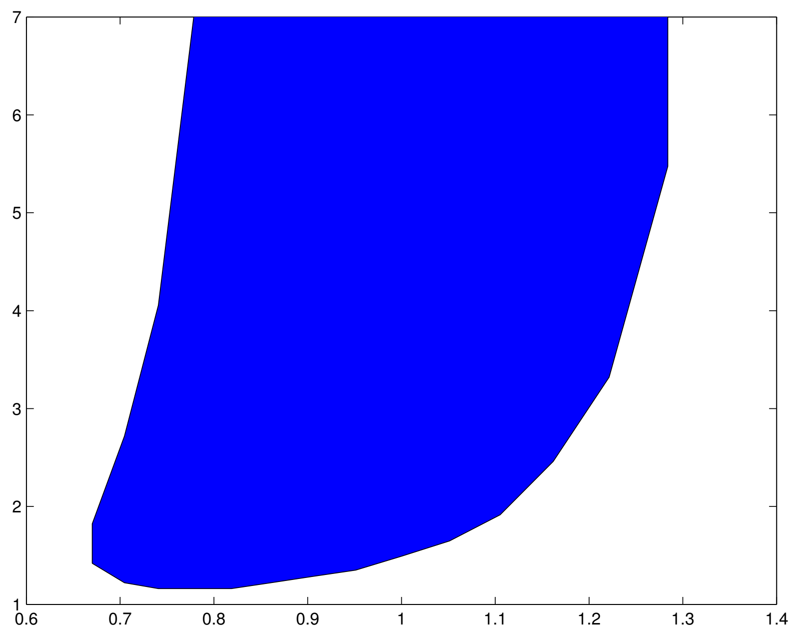
$$\begin{aligned} \hat{q} &= \frac{n}{N} \\ \hat{\lambda}_{j+1} &= \frac{N}{\sum_{i=1}^{m} x_i + (1/\hat{\lambda}_j) \sum_{i=1}^{n} \frac{\int_{\hat{\lambda}_j z_i}^{\infty} w^{-(\hat{\beta}_j - 1)} e^{-w} dw}{\int_{\hat{\lambda}_j z_i}^{\infty} w^{-\hat{\beta}_j} e^{-w} dw}} \\ \hat{\beta}_{j+1} &=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

Simulated Example in Exponential-Power Model

- $N = m + n = 100$

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

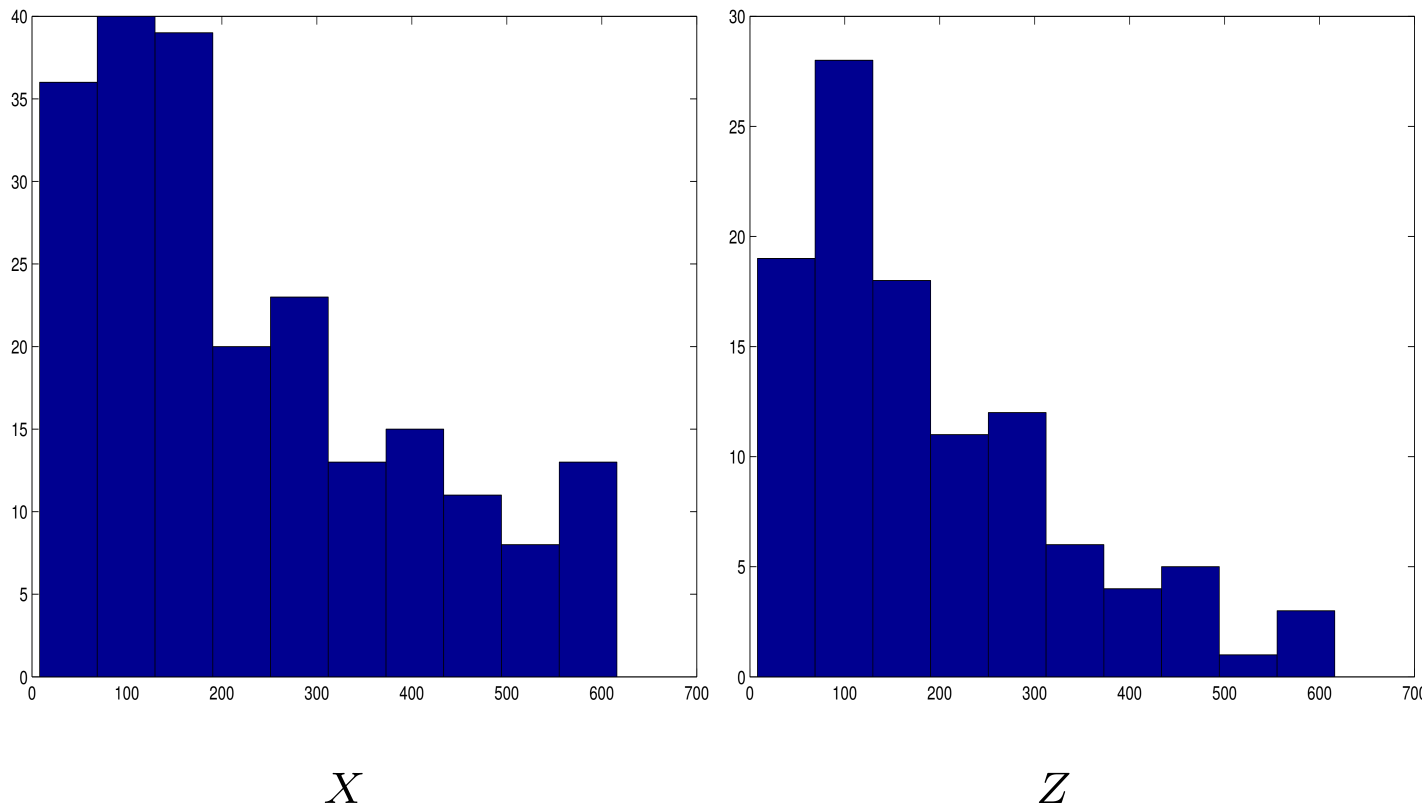


The 95% confidence region for λ (horizontal axis) and β from the simulated data

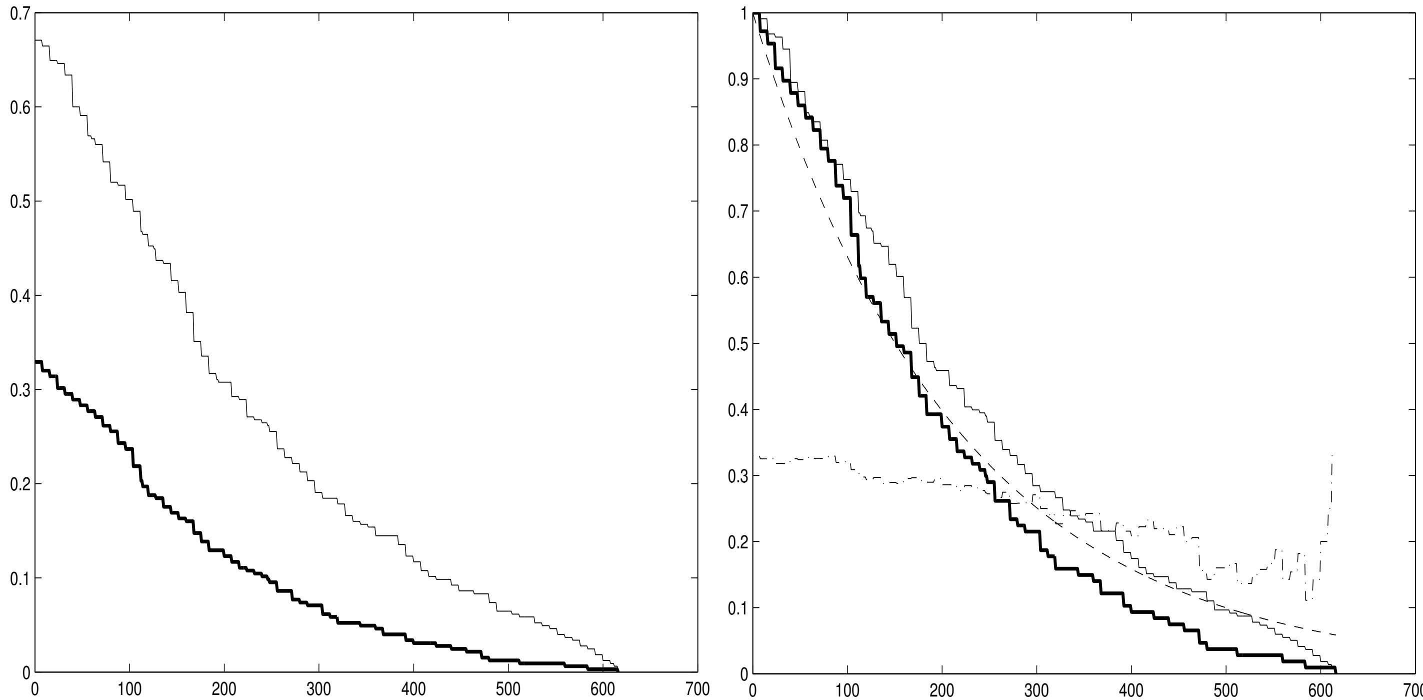
Example: VHF data (Mendenhall and Hader, 1958)

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), $n = 107$



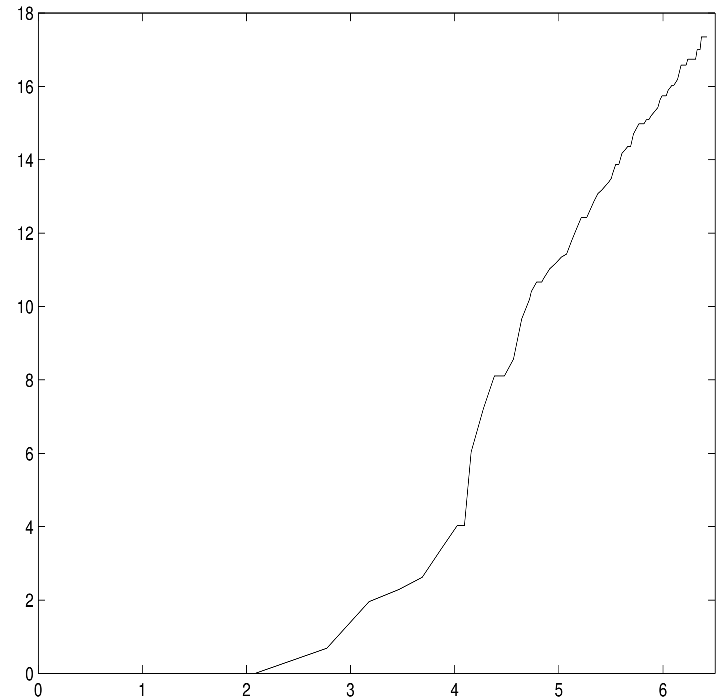
Example: VHF data (Mendenhall and Hader, 1958)



- *Left: Estimated subsurvival functions for X, thin line, and Z, thin 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 $\approx .15$

VHF data: Nonparametric Repair Alert Model



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

VHF data: Fitted Exponential-Power 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*

Properties of Repair Alert Model

- Time to next event, $Y = \min(X, Z)$:

$$F_Y(y) = F_X(y) + qG(y) \int_y^\infty \frac{f_X(x)}{G(x)} dx$$

- Expected time to next event for general $G(\cdot)$:

$$E(Y) = E(X) - qE\left[\frac{\mathcal{G}(X)}{G(X)}\right]$$

where $\mathcal{G}(x) = \int_0^x G(t)dt$.

- Expected time to next event for $G(x) = x^\beta$:

$$E(Y) = E(X)\left(1 - \frac{q}{\beta + 1}\right)$$

- Expected cost per time unit in the long run if C_{PM}, C_{CR} are costs of PM and failure, respectively:

$$\frac{qC_{PM} + (1-q)C_{CR}}{E(X)\left(1 - \frac{q}{\beta+1}\right)}$$

Main Conclusions

- The Repair Alert Model is a special case of Random Signs Censoring, obtainable under the same conditions on the subsurvival distributions
- As for Random Signs Censoring, the marginal distribution of the failure time X is identifiable under the Repair Alert Model
- The Repair Alert Function $G(x)$ is easy to interprete and is identifiable from the conditional subdistribution functions
- The Repair Alert Function $G(x)$ can be estimated from data both nonparametrically and parametrically