Estimation and model selection for left-truncated and right-censored data: Application to power transformer lifetime modeling

Presenter：徐 紹 凱
Advisor：Takeshi Emura (江 村 剛 志) 博士

Outline

- Introduction
- Left-truncated and right-censored data
- Methods
- Model selection
- Simulation results
- Data analysis
- Conclusion and Discussion

Introduction

- Truncation and censoring are seen commonly in lifetime data as discussed in the books by Cohen (1991), Balakrishnan and Cohen (1991), and Meeker and Escobar (1998).

- Recently, Hong et al. (2009) carried out an analysis of electric power-transformer data from an energy company in the US.

- The EM algorithm was developed in lognormal distribution to left truncated and right censored data by Balakrishnan and Mitra (2011).

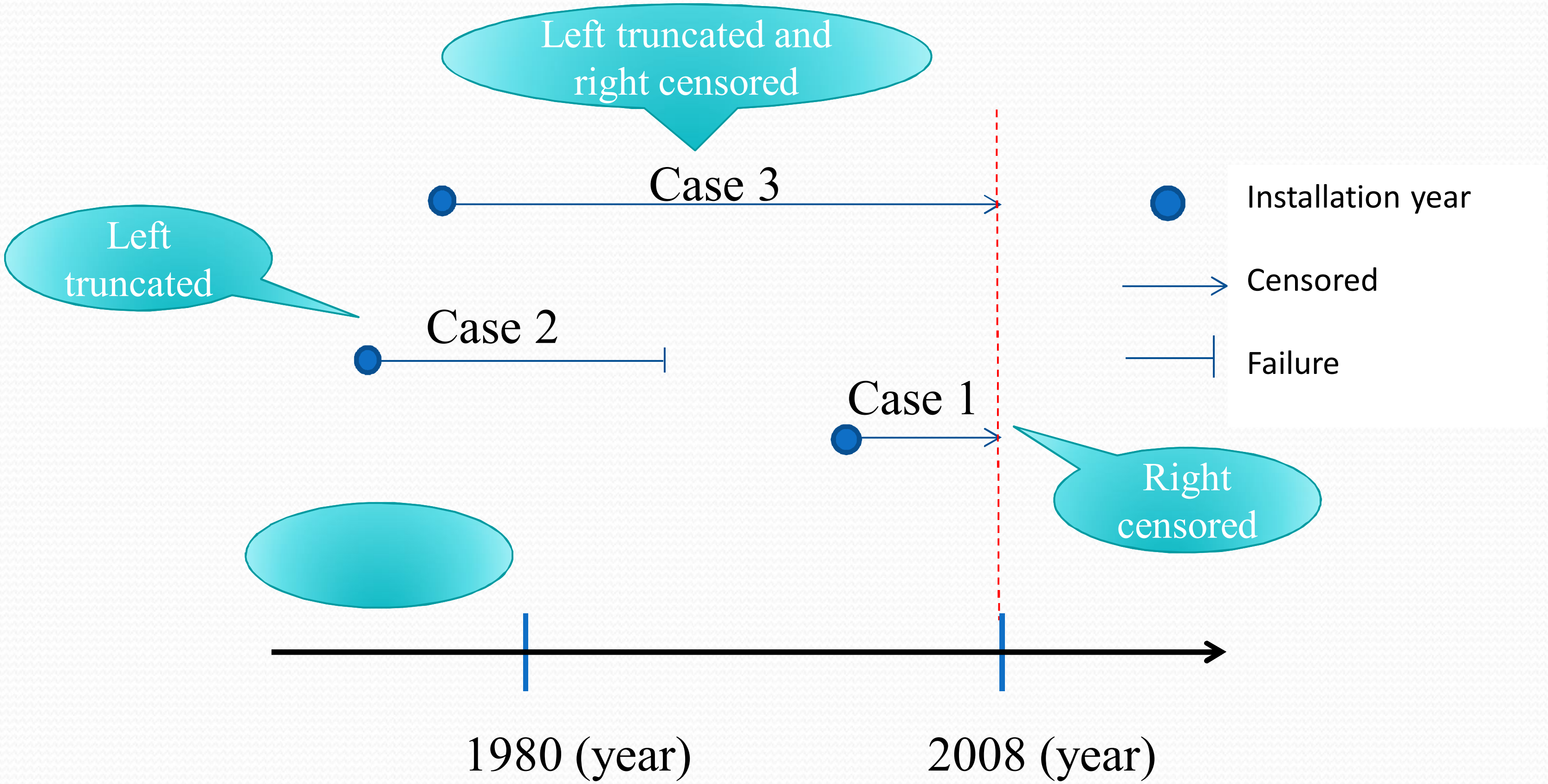
- The EM algorithm was also developed in Weibull distribution to left truncated and right censored by Balakrishnan and Mitra (2012).

Objective:

- Newton-Raphson (NR) method vs. EM algorithm under the lognormal and Weibull distributions.

- Model selection by using Akaike's information criterion or AIC (Akaike, 1974).

Left-truncated and right-censored data



Methods

- Lognormal:

1. Two-dimensional Newton-Raphson.
2. EM algorithms. (Balakrishnan and Mitra, 2011).

- Weibull:

1. One-dimensional Newton-Raphson.
2. EM algorithm. (Balakrishnan and Mitra, 2012).

- Exponential:

No need to use numerical method.

Notations

- c : right censored time.
- c^L: log-transformed right censored time.
- δ: censoring indicator

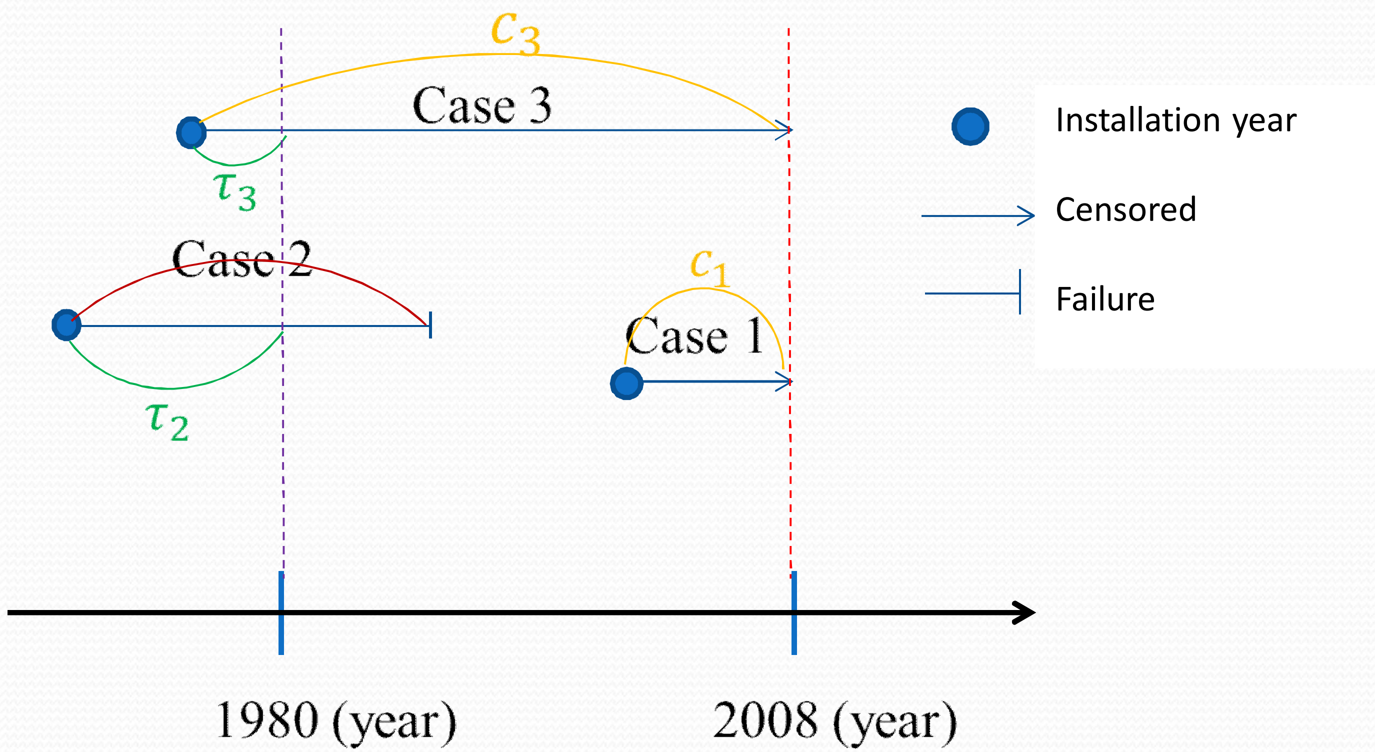
$$\delta = \begin{cases} 0, & if\ censored \\ 1, & if\ not\ censored \end{cases}.$$

- τ: left truncated time.
- τ^L: log-transformed left truncated time.
- v: truncation indicator.

$$v = \begin{cases} 0, & if\ truncated \\ 1, & if\ not\ truncated \end{cases}.$$

- S_1: the set of machines that installed on or after 1980.
- S_2: the set of machines that installed before 1980.

Example

- 
c_3
Case 3
τ_3
Case 2
τ_2
c_1
Case 1
1980 (year)
2008 (year)
Installation year
Censored
Failure

Lognormal

- Let X be the original lifetime variable that follows a lognormal distribution with parameters μ and σ

$$f_X(x) = \frac{1}{\sigma x\sqrt{2\pi}} \exp\left\{ \frac{-(\log x - \mu)^2}{2\sigma^2} \right\}, x > 0.$$

- Let $T = \log X$ be a normal distribution with density

$$f_T(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ \frac{-(t - \mu)^2}{2\sigma^2} \right\}, -\infty < t < \infty.$$

- Let $\phi(.)$ and $\Phi(.)$ denote the probability density function and cumulative distribution function of the standard normal distribution.
- Define $y_i = \min(t_i, c_i{}^L), i = 1, \ldots, n,$ (observed lifetime, log-transform).
- The likelihood function for the left truncation and right censoring data is

$$L(\mu,\sigma)=\prod_{i\in S_1}\left\{\frac{1}{\sigma}\phi\left(\frac{y_i-\mu}{\sigma}\right)\right\}^{\delta_i}\left\{1-\Phi\left(\frac{y_i-\mu}{\sigma}\right)\right\}^{1-\delta_i}\times\prod_{i\in S_2}\left\{\frac{\frac{1}{\sigma}\phi\left(\frac{y_i-\mu}{\sigma}\right)}{1-\Phi\left(\frac{\tau_i{}^L-\mu}{\sigma}\right)}\right\}^{\delta_i}\left\{\frac{1-\Phi\left(\frac{y_i-\mu}{\sigma}\right)}{1-\Phi\left(\frac{\tau_i{}^L-\mu}{\sigma}\right)}\right\}^{1-\delta_i}.$$

- The log-likelihood function (without the constant term) is given by

$$\log L(\mu,\sigma)=\sum_{i=1}^{n}\left\{-\delta_i[\log\sigma+\delta_i\frac{(y_i-\mu)^2}{2\sigma^2}]+(1-\delta_i)\log\left[1-\Phi\left(\frac{y_i-\mu}{\sigma}\right)\right]\right\}-\sum_{i=1}^{n}\left\{(1-v_i)\log\left[1-\Phi\left(\frac{\tau_i{}^L-\mu}{\sigma}\right)\right]\right\}.$$

Newton-Raphson

- At the (k+1)-th step of iteration process, the updated parameter is obtained as

$$\begin{bmatrix} \mu_{k+1} \\ \sigma_{k+1} \end{bmatrix} = \begin{bmatrix} \mu_k \\ \sigma_k \end{bmatrix} - J_f^{-1}(\mu_k, \sigma_k) \begin{bmatrix} f_1(\mu_k, \sigma_k) \\ f_2(\mu_k, \sigma_k) \end{bmatrix}$$

where

$$f_1(\mu_k, \sigma_k) = \frac{\partial}{\partial \mu} \log L(\mu_k, \sigma_k) \qquad f_2(\mu_k, \sigma_k) = \frac{\partial}{\partial \sigma} \log L(\mu_k, \sigma_k)$$

$$J_f(\mu_k, \sigma_k) = \begin{bmatrix} \frac{\partial}{\partial \mu} f_1(\mu_k, \sigma_k) & \frac{\partial}{\partial \sigma} f_1(\mu_k, \sigma_k) \\ \frac{\partial}{\partial \mu} f_2(\mu_k, \sigma_k) & \frac{\partial}{\partial \sigma} f_2(\mu_k, \sigma_k) \end{bmatrix}.$$

The iteration process then continues until convergence, i.e., until $\left\|\mu^{(k+1)} - \mu^{(k)}\right\| < \varepsilon$ and $\left\|\sigma^{(k+1)} - \sigma^{(k)}\right\| < \varepsilon$ for some pre-fixed $\varepsilon > 0$.

- For the two-parameter lognormal distribution, the first-order derivatives of the log-likelihood with respect to the parameters are given by

$$\frac{\partial}{\partial\mu}\log L(\mu,\sigma)=\sum_{i=1}^{n}\left[\frac{\delta_i}{\sigma^2}(y_i-\mu)+\frac{(1-\delta_i)}{\sigma}h_{y_i}(\mu,\sigma)-\frac{(1-v_i)}{\sigma}h_i(\mu,\sigma)\right],$$

$$\frac{\partial}{\partial\sigma}\log L(\mu,\sigma)=\sum_{i=1}^{n}\left[-\frac{\delta_i}{\sigma}+\frac{\delta_i}{\sigma^3}(y_i-\mu)^2+(1-\delta_i)h_{y_i}(\mu,\sigma)\left(\frac{y_i-\mu}{\sigma^2}\right)-(1-v_i)h_i(\mu,\sigma)\left(\frac{\tau_i^L-\mu}{\sigma^2}\right)\right],$$

where

$$h_{y_i}(\mu,\sigma)=\frac{\phi\left(\frac{y_i-\mu}{\sigma}\right)}{1-\Phi\left(\frac{y_i-\mu}{\sigma}\right)}\quad\text{and}\quad h_i(\mu,\sigma)=\frac{\phi\left(\frac{\tau_i^L-\mu}{\sigma}\right)}{1-\Phi\left(\frac{\tau_i^L-\mu}{\sigma}\right)}.$$

- The second-order derivatives of the log-likelihood with respect to the parameters are then obtained as follows:

$$\frac{\partial^2}{\partial\mu^2}\log L(\mu,\sigma)=\sum_{i=1}^{n}\left\{-\frac{\delta_i}{\sigma^2}+\frac{(1-\delta_i)}{\sigma^2}\left[\left(\frac{y_i-\mu}{\sigma}\right)h_{y_i}(\mu,\sigma)-[h_{y_i}(\mu,\sigma)]^2\right]\right.$$

$$\left.-\frac{(1-v_i)}{\sigma^2}\left[\left(\frac{\tau_i^L-\mu}{\sigma}\right)h_i(\mu,\sigma)-[h_i(\mu,\sigma)]^2\right]\right\},$$

$$\frac{\partial^2}{\partial\sigma^2}\log L(\mu,\sigma)$$

$$=\sum_{i=1}^{n}\left\{\frac{\delta_i}{\sigma^2}-\frac{3\delta_i}{\sigma^4}(y_i-\mu)^2-2(1-\delta_i)\left(\frac{y_i-\mu}{\sigma^3}\right)h_{y_i}(\mu,\sigma)+(1-v_i)\frac{2(\tau_i^L-\mu)}{\sigma^3}h_i(\mu,\sigma)\right.$$

$$+(1-\delta_i)\left(\frac{y_i-\mu}{\sigma^2}\right)\left[\frac{(y_i-\mu)^2}{\sigma^3}h_{y_i}(\mu,\sigma)-\frac{(y_i-\mu)}{\sigma^2}[h_{y_i}(\mu,\sigma)]^2\right]$$

$$\left.-(1-v_i)\left(\frac{\tau_i^L-\mu}{\sigma^2}\right)\left[\frac{(\tau_i^L-\mu)^2}{\sigma^3}h_i(\mu,\sigma)-\frac{(\tau_i^L-\mu)}{\sigma^2}[h_i(\mu,\sigma)]^2\right]\right\},$$

$$\frac{\partial^2}{\partial\mu\partial\sigma}\log L(\mu,\sigma)=\sum_{i=1}^{n}\left\{-\frac{2\delta_i}{\sigma^3}(y_i-\mu)-\frac{(1-\delta_i)}{\sigma^2}h_{y_i}(\mu,\sigma)+\frac{(1-v_i)}{\sigma^2}h_i(\mu,\sigma)\right.$$

$$+\frac{(1-\delta_i)}{\sigma}\left[\frac{(y_i-\mu)^2}{\sigma^3}h_{y_i}(\mu,\sigma)-\frac{(y_i-\mu)}{\sigma^2}[h_{y_i}(\mu,\sigma)]^2\right]$$

$$\left.-\frac{(1-v_i)}{\sigma}\left[\frac{(\tau_i^L-\mu)^2}{\sigma^3}h_i(\mu,\sigma)-\frac{(\tau_i^L-\mu)}{\sigma^2}[h_i(\mu,\sigma)]^2\right]\right\}.$$

EM algorithm

E-step: the conditional expectation of the complete data likelihood, given the observed incomplete data and the current value of parameter.

M-step: the obtained expression of the conditional expectation is maximized with respect to the parameter.

- We introduce the EM algorithm for maximizing the MLE proposed by Balakrishnan and Mitra (2011).

- Let $\theta = (\mu, \sigma)'$ denote the parameter vector.

- The complete data likelihood (no censoring) would simply be

$$L_c(t,\theta) = \prod_{i \in S_1} \left\{ \frac{1}{\sigma} \phi\left(\frac{t_i - \mu}{\sigma} \right) \right\} \times \prod_{i \in S_2} \left\{ \frac{\frac{1}{\sigma} \phi\left(\frac{t_i - \mu}{\sigma} \right)}{1 - \Phi\left(\frac{\tau_i^L - \mu}{\sigma} \right)} \right\}.$$

- The log-likelihood function

$$\log L_c(t,\theta) = -n \log \sigma - \sum_{i=1}^{n} \left\{ \frac{(t_i - \mu)^2}{2\sigma^2} \right\} - \sum_{i=1}^{n} (1 - v_i) \log\left[1 - \Phi\left(\frac{\tau_i^L - \mu}{\sigma} \right) \right].$$

- $\delta = (\delta_1, \delta_2, \dots, \delta_n)'$ denote the vector of censoring indicator.
- $y = (y_1, y_2, \dots, y_n)'$ denote the vector of observed lifetimes,

where $y_i = \min(t_i, c_i)$.

- E-step:

$$Q(\theta, \theta^{(k)}) = E_{\theta^{(k)}}\left[\log L_c(t, \theta) \mid y, \delta\right].$$

- Using these results of Cohen (1991), then obtaining

$$Q(\theta, \theta^{(k)}) = -n\log\sigma - \sum_{i=1}^{n}(1-v_i)\log\left[1-\Phi\left(\frac{\tau_i^{\,L}-\mu}{\sigma}\right)\right]$$

$$-\frac{1}{2\sigma^2}\left\{\sum_{i;\delta_i=1}^{n} y_i^2 + \sum_{i;\delta_i=0}\left[\sigma^{(k)2}(1+\xi_i^{(k)}q_i^{(k)}) + 2\sigma^{(k)}\mu^{(k)}q_i^{(k)} + \mu^{(k)2}\right]\right\}$$

$$+\frac{\mu}{\sigma^2}\left\{\sum_{i;\delta_i=1}^{n} y_i + \sum_{i;\delta_i=0}(\mu^{(k)} + \sigma^{(k)}q_i^{(k)})\right\} - \frac{n\mu^2}{2\sigma^2}.$$

- M-step:

$$\theta^{(k+1)} = \arg\max_{\theta} Q(\theta, \theta^{(k)}).$$

- As suggested by Balakrishnan and Mitra (2011), the maximization is done by

1. Approximating the hazard function by the Taylor expansion. (EM1)
2. EM gradient algorithm (Lange, 1995). (EM2)

- The E-step and the M-step are then repeated until convergence to the MLE of the parameter θ.

Weibull

- We denote the lifetime variable by X

- The c.d.f. and p.d.f. of the Weibull random variable X can be expressed as

$$F(x;\eta,\beta) = 1 - \exp\left[-\left(\frac{x}{\eta} \right)^{\beta} \right], x \geq 0,$$

$$f(x;\eta,\beta) = \left(\frac{\beta}{\eta} \right) \left(\frac{x}{\eta} \right)^{\beta-1} \exp\left[-\left(\frac{x}{\eta} \right)^{\beta} \right], x \geq 0, \eta > 0, \beta > 0,$$

where η is the scale parameter and β is the shape parameter.

- Let $x = (x_1, x_2, \dots, x_n)'$ denote the vector of observed lifetimes.

- The likelihood function for the left truncation and right censoring data is

$$L(\eta,\beta)=\prod_{i\in S_1}\left\{\left(\frac{\beta}{\eta}\right)\left(\frac{x_i}{\eta}\right)^{\beta-1}\exp\left[-\left(\frac{x_i}{\eta}\right)^{\beta}\right]\right\}^{\delta_i}\left\{\exp\left[-\left(\frac{x_i}{\eta}\right)^{\beta}\right]\right\}^{1-\delta_i}$$

$$\times\prod_{i\in S_2}\left\{\frac{\left(\frac{\beta}{\eta}\right)\left(\frac{x_i}{\eta}\right)^{\beta-1}\exp\left[-\left(\frac{x_i}{\eta}\right)^{\beta}\right]}{\exp\left[-\left(\frac{\tau_i}{\eta}\right)^{\beta}\right]}\right\}^{\delta_i}\left\{\frac{\exp\left[-\left(\frac{x_i}{\eta}\right)^{\beta}\right]}{\exp\left[-\left(\frac{\tau_i}{\eta}\right)^{\beta}\right]}\right\}^{1-\delta_i}.$$

- The log-likelihood function is given by

$$\log L(\eta,\beta)=\sum_{i=1}^{n}\left\{-\delta_i[\log\beta-\log\eta+(\beta-1)(\log x_i-\log\eta)]-\left(\frac{x_i}{\eta}\right)^{\beta}\right\}+\sum_{i=1}^{n}(1-v_i)\left(\frac{\tau_i}{\eta}\right)^{\beta}.$$

- For the two-parameter Weibull distribution, the first-order derivative of the log-likelihood with respect to the parameter η is given by

$$\frac{\partial}{\partial \eta}\log L(\eta, \beta) = \sum_{i=1}^{n}\left\{ \delta_i \left(-\frac{\beta}{\eta} \right) + \beta \frac{x_i^{\beta}}{\eta^{\beta+1}} \right\} - \sum_{i=1}^{n}\left\{ (1-v_i)\beta \frac{(\tau_i)^{\beta}}{\eta^{\beta+1}} \right\}.$$

- Let $\frac{\partial}{\partial \eta}\log L(\eta, \beta) = 0$ and then obtain an explicit solution,

$$\hat{\eta} = \left\{ \frac{\sum_{i=1}^{n} x_i^{\beta} - \sum_{i=1}^{n}[(1-v_i)(\tau_i)^{\beta}]}{\sum_{i=1}^{n}\delta_i} \right\}^{\frac{1}{\beta}}.$$

- The first-order and second-order derivatives of the log-likelihood with respect to the parameter β are given by

$$\frac{\partial}{\partial\beta}\log L(\eta,\beta)=\sum_{i=1}^{n}\left\{\delta_i\left[\frac{1}{\beta}+\log(x_i)-\log(\eta)\right]+\frac{x_i^{\beta}}{\eta^{\beta}}\log\left(\frac{t_i}{\eta}\right)\right\}+\sum_{i=1}^{n}\left\{(1-v_i)\frac{(\tau_i)^{\beta}}{\eta^{\beta}}\log\left(\frac{\tau_i}{\eta}\right)\right\},$$

$$\frac{\partial^2}{\partial\beta^2}\log L(\eta,\beta)=\sum_{i=1}^{n}\left\{\delta_i\left[-\frac{1}{\beta^2}\right]-\frac{x_i^{\beta}}{\eta^{\beta}}\left[\log\left(\frac{x_i}{\eta}\right)\right]^2\right\}+\sum_{i=1}^{n}\left\{(1-v_i)\frac{(\tau_i)^{\beta}}{\eta^{\beta}}\left[\log\left(\frac{\tau_i}{\eta}\right)\right]^2\right\}.$$

Newton-Raphson

- At the (k+1)-th step of iteration process, the parameter β is obtained as

$$\beta_{k+1} = \beta_k - \frac{f(\beta_k)}{f'(\beta_k)},$$

where

$$f(\beta_k) = \frac{\partial}{\partial \beta} \log L(\hat{\eta}, \beta_k) \quad \text{and} \quad f'(\beta_k) = \frac{\partial^2}{\partial \beta^2} \log L(\hat{\eta}, \beta_k).$$

The iteration process then continues until convergence, i.e, until $\left\|\beta^{(k+1)} - \beta^{(k)}\right\| < \varepsilon$ for some pre-fixed $\varepsilon > 0$.

EM algorithm

- We introduce the EM algorithm for maximizing the MLE proposed by Balakrishnan and Mitra (2012).

- Let $T = \log X$ be the log-transformed variable which follows an extreme value distribution.

- The density is given by

$$f_T(t) = \frac{1}{\sigma}\exp\left[\left(\frac{t-\mu}{\sigma}\right) - \exp\left(\frac{t-\mu}{\sigma}\right)\right], -\infty < t < \infty, -\infty < \mu < \infty, \sigma > 0,$$

where $\mu = \log(\eta)$ and $\sigma = \dfrac{1}{\beta}$.

- Under the extreme value model, had there been no censoring, the complete data likelihood would be

$$L_c(t,\theta)=\prod_{i\in S_1}\left\{\frac{1}{\sigma}-\exp\left[\left(\frac{t_i-\mu}{\sigma}\right)-\exp\left(\frac{t_i-\mu}{\sigma}\right)\right]\right\}$$

$$\times\prod_{i\in S_2}\left\{\frac{\exp\left[\exp\left(\frac{\tau_i^L-\mu}{\sigma}\right)\right]}{\sigma}\exp\left[\left(\frac{t_i-\mu}{\sigma}\right)-\exp\left(\frac{t_i-\mu}{\sigma}\right)\right]\right\}.$$

- And the corresponding log-likelihood function is

$$\log L_c(t,\theta)=-n\log\sigma-\sum_{i=1}^{n}\left\{\left(\frac{t_i-\mu}{\sigma}\right)-\exp\left(\frac{t_i-\mu}{\sigma}\right)\right\}-\sum_{i=1}^{n}(1-v_i)\exp\left(\frac{\tau_i^L-\mu}{\sigma}\right).$$

- $\delta = (\delta_1, \delta_2, \ldots, \delta_n)'$ denote the vector of censoring indicator.
- $y = (y_1, y_2, \ldots, y_n)'$ denote the vector of observed lifetimes, where $y_i = \min(t_i, c_i)$.

- E-step:

$$Q(\theta, \theta^{(k)}) = E_{\theta^{(k)}}\left[\log L_c(t, \theta) \mid y, \delta \right].$$

- Then, it can be derived

$$Q(\theta, \theta^{(k)}) = -n \log \sigma - \sum_{i=1}^{n} (1 - v_i) \exp\left(\frac{\tau_i^{\,L} - \mu}{\sigma} \right) + \sum_{\delta_i = 1} \left\{ \left(\frac{t_i - \mu}{\sigma} \right) - \exp\left(\frac{t_i - \mu}{\sigma} \right) \right\}$$

$$+ \sum_{\delta_i = 0} \left\{ \frac{\mu^{(k)} + \sigma^{(k)} E_{1i}^{(k)} - \mu}{\sigma} \right\} - \sum_{\delta_i = 0} \left\{ e^{\left(\frac{\mu^{(k)} - \mu}{\sigma} \right)} M_{\left(\frac{T_i - \mu}{\sigma} \right)}\left(\frac{\sigma^{(k)}}{\sigma} \right) \right\}.$$

- M-step:

$$\theta^{(k+1)} = \arg\max_{\theta} Q(\theta, \theta^{(k)}) .$$

- As suggested by Balakrishnan and Mitra (2012), the maximization is done by EM gradient algorithm (Lange, 1995).

- The E-step and the M-step are then repeated until convergence to the MLE of the parameter θ.

Exponential

- Let X be the lifetime variable that follows an exponential distribution with parameter η .

- The c.d.f. and p.d.f. of the exponential random variable X can be expressed as

$$F(x;\eta) = 1 - \exp\left(-\frac{x}{\eta}\right), x \geq 0,$$

$$f(x;\eta) = \frac{1}{\eta}\exp\left(-\frac{x}{\eta}\right), t \geq 0.$$

- Let $x = (x_1, x_2, \ldots, x_n)'$ denote the vector of observed lifetimes.

- The likelihood function for the left truncation and right censoring data is

$$L(\eta) = \prod_{i \in S_1} \left\{ \frac{1}{\eta} \exp\left(-\frac{x_i}{\eta} \right) \right\}^{\delta_i} \left\{ \exp\left(-\frac{x_i}{\eta} \right) \right\}^{1-\delta_i} \times \prod_{i \in S_2} \left\{ \frac{\frac{1}{\eta} \exp\left(-\frac{x_i}{\eta} \right)}{\exp\left(-\frac{\tau_i}{\eta} \right)} \right\}^{\delta_i} \left\{ \frac{\exp\left(-\frac{x_i}{\eta} \right)}{\exp\left(-\frac{\tau_i}{\eta} \right)} \right\}^{1-\delta_i} .$$

- The log-likelihood function is given by

$$\log L(\eta) = \sum_{i=1}^{n} \left\{ -\delta_i \log \eta - \frac{x_i}{\eta} \right\} + \sum_{i=1}^{n} (1 - v_i) \left(\frac{\tau_i}{\eta} \right).$$

- The first-order and second-order derivatives of the log-likelihood with respect to the parameter η are given by

$$\frac{\partial}{\partial\eta}\log L(\eta)=\sum_{i=1}^{n}\left\{-\frac{\delta_i}{\eta}+\frac{x_i}{\eta^2}\right\}-\sum_{i=1}^{n}\left\{(1-v_i)\left(\frac{\tau_i}{\eta^2}\right)\right\},$$

$$\frac{\partial}{\partial\eta^2}\log L(\eta)=\sum_{i=1}^{n}\left\{\frac{\delta_i}{\eta^2}-2\frac{x_i}{\eta^3}\right\}+2\sum_{i=1}^{n}\left\{(1-v_i)\left(\frac{\tau_i}{\eta^3}\right)\right\}.$$

- Assume $\frac{\partial}{\partial \eta} \log L(\eta) = 0$, then we can obtain

$$\hat{\eta} = \frac{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \{(1 - v_i)\tau_i\}}{\sum_{i=1}^{n} \delta_i}.$$

- Checking $\frac{\partial}{\partial \eta^2} \log L(\hat{\eta}) < 0.$

- Then we can say $\hat{\eta}$ is the MLE of η.

Model selection

- AIC: Akaike's information criterion (Akaike, 1974).

$$AIC = -2\log L + 2k$$

- BIC: Bayesian information criterion (Schwarz, 1978).

$$BIC = -2\log L + k\log n$$

k: the number of unknown parameters in the model.

L: the maximized value of the likelihood function.

n: sample size.

Simulation

- All the simulation results are based on 1000 Monte Carlo runs.
- Truncation percentage = 30% or 60 %.
- Sample size n =50, 100 and 200.
- Stopping criterion $\varepsilon = 0.001$.

- How to generate data?

1. The installation years were simulated by unequal probability.

The installation years

(1960-1979); Truncated				
Year	1960	1961	1962	1963
Probability	0.1	0.1	0.1	0.1
Year	1964	1965	1966~1979	
Probability	0.1	0.1	0.4/14 (each year)	

(1980-1995); Untruncated				
Year	1980	1981	1982	1983
Probability	0.15	0.15	0.15	0.15
Year	1984	1985~1995		
Probability	0.15	0.25/11 (each year)		

2. The lifetimes of the machines, in years, are simulated from distributions (Lognormal, Weibull, Exponential).

3. To know the censoring percentage. (?)

$$Installation + Lifetime \geq 2008$$

Lognormal

- $(\mu, \sigma) = (3.5, 0.5)$ or $(3.0, 0.2)$.

- Initial values:

μ: sample mean of the lifetime.

σ: sample standard deviation of the lifetime.

Table 1 (n=100)

(μ, σ)	Trunc. (%) Cen. (%)	Method	$B(\hat{\mu})$	$B(\hat{\sigma})$	$MSE(\hat{\mu})$	$MSE(\hat{\sigma})$	AI
(3.5, 0.5)	30; 62.3	EM1	0.00001	0.00028	0.00508	0.00445	14.0
		EM2	-0.00033	-0.00001	0.00507	0.00445	17.5
		NR	0.00216	0.00260	0.00517	0.00449	7.0
	60; 51.1	EM1	0.00703	-0.00102	0.00431	0.00363	10.3
		EM2	0.00662	-0.00134	0.00430	0.00362	13.5
		NR	0.00772	0.00026	0.00437	0.00364	6.4
(3.0, 0.2)	30; 17.4	EM1	-0.00103	-0.00134	0.00047	0.00025	3.9
		EM2	-0.00110	-0.00150	0.00047	0.00025	4.3
		NR	-0.00010	-0.00136	0.00047	0.00025	3.9
	60; 10.1	EM1	0.00107	-0.00142	0.00054	0.00025	3.7
		EM2	0.00096	-0.00140	0.00054	0.00025	3.7
		NR	0.00099	-0.00135	0.00054	0.00025	3.5

EM1 is the EM algorithm method approximating the hazard function by the Taylor expansion. EM2 is the EM gradient algorithm. NR is the Newton-Raphson method.

Confidence interval

-
- For instance, the confidence interval for μ is obtain as

$$[\ \hat{\mu} - z_{\alpha/2}\sqrt{\hat{Var}(\hat{\mu})},\ \hat{\mu} + z_{\alpha/2}\sqrt{\hat{Var}(\hat{\mu})}\ ].$$

Table 2 (Coverage probability for μ)

μ	Sample size	Trunc.(%) Cen. (%)	Nominal CL (%)	Method		
				EM1	EM2	NR
3.5	100	30; 62.3	90	0.952 (3.357, 3.643)	0.954 (3.357, 3.642)	0.902 (3.383, 3.621)
			95	0.981 (3.330, 3.670)	0.981 (3.330, 3.670)	0.95 (3.360, 3.644)
		60; 51.1	90	0.948 (3.379, 3.635)	0.949 (3.378, 3.635)	0.897 (3.403, 3.613)
			95	0.984 (3.354, 3.660)	0.984 (3.354, 3.659)	0.943 (3.383, 3.633)
	200	30; 62.3	90	0.953 (3.399, 3.600)	0.953 (3.399, 3.599)	0.902 (3.418, 3.585)
			95	0.978 (3.380, 3.619)	0.978 (3.380, 3.618)	0.953 (3.402, 3.601)
		60; 51.1	90	0.961 (3.410, 3.590)	0.961 (3.410, 3.589)	0.91 (3.427, 3.574)
			95	0.983 (3.393, 3.607)	0.983 (3.393, 3.607)	0.956 (3.413, 3.588)

Table 3 (Coverage probability for σ)

σ	Sample size	Trunc.(%) Cen. (%)	Nominal CL (%)	Method		
				EM1	EM2	NR
0.5	100	30; 62.3	90	0.903 (0.389, 0.611)	0.903 (0.389, 0.611)	0.885 (0.398, 0.607)
			95	0.945 (0.368, 0.632)	0.945 (0.368, 0.632)	0.933 (0.378, 0.627)
		60; 51.1	90	0.887 (0.401, 0.597)	0.887 (0.401, 0.597)	0.888 (0.402, 0.598)
			95	0.943 (0.382, 0.616)	0.943 (0.382, 0.615)	0.94 (0.383, 0.617)
	200	30; 62.3	90	0.906 (0.421, 0.577)	0.905 (0.421, 0.576)	0.892 (0.427, 0.575)
			95	0.952 (0.406, 0.592)	0.951 (0.406, 0.591)	0.936 (0.413, 0.589)
		60; 51.1	90	0.897 (0.429, 0.566)	0.896 (0.428, 0.566)	0.896 (0.430, 0.567)
			95	0.943 (0.415, 0.579)	0.943 (0.415, 0.579)	0.941 0.417, 0.580)

- We encounter a big problem in the process of the simulation.

➡ Methods does not converge in some situations.

- What kind of situations that methods diverge frequently?

Table 4

Method	(μ, σ)	Sample size	Truncation (%)	Censoring (%)	Convergence	NF
EM1	(3.5, 0.5)	50	30 60	62.3 51.1	Yes Yes	
		100	30 60	62.3 51.1	Yes Yes	
EM2	(3.5, 0.5)	50	30 60	62.3 51.1	No Yes	0.6
		100	30 60	62.3 51.1	Yes Yes	
NR	(3.5, 0.5)	50	30 60	62.3 51.1	No No	18.7 4.8
		100	30 60	62.3 51.1	No Yes	1.3

NF is the average number of runs that fails to converge in 1000 Monte Carlo runs.

- We encounter a big problem in the process of the simulation.

⇒ Methods does not converge in some situations.

- What kind of situations that methods diverge frequently?

⇒ Small sample sizes and high censoring percentage.

Table 5 (n=50)

(μ, σ)	Trunc. (%) Cen. (%)	Method	$B(\hat{\mu})$	$B(\hat{\sigma})$	$MSE(\hat{\mu})$	$MSE(\hat{\sigma})$	AI
(3.5, 0.5)	30; 62.3	EM1	0.00235	-0.00594	0.01038	0.00802	14.17
	60; 51.1	EM1	0.00372	-0.00712	0.00802	0.00713	10.33
(3.0, 0.2)	30; 17.4	EM1	0.00141	-0.00254	0.00089	0.00050	3.97
	60; 10.1	EM1	0.00102	-0.00248	0.00103	0.00052	3.77

Weibull

- $(\eta, \beta) = (35, 3)$ or $(40, 3)$.
- $(\mu, \sigma) = (3.55, 0.33)$ or $(3.69, 0.33)$.

- Initial values:

We denote T as the log-transformed lifetime variable which follows extreme value distribution. Then

$$E(T) = \mu - \gamma\sigma \text{ and } Var(T) = \frac{\pi^2}{6}\sigma^2$$

where γ=0.5772 is Euler's constant.

Table 6 (n=50)

(μ, σ)	Trunc. (%) Cen. (%)	Method	$B(\hat{\mu})$	$B(\hat{\sigma})$	$MSE(\hat{\mu})$	$MSE(\hat{\sigma})$	AI
(3.55, 0.33)	30; 56.8	EM	-0.00414	-0.00848	0.00569	0.00333	21.8
		NR	0.00028	-0.00610	0.00585	0.00333	5.3
	60;42.3	EM	-0.00321	-0.00125	0.00415	0.00310	12.8
		NR	-0.00149	-0.00037	0.00421	0.00310	4.7
(3.69, 0.33)	30; 66.3	EM	-0.00722	0.00957	0.00463	0.00433	33.1
		NR	0.00114	-0.00002	0.00471	0.00442	6.0
	60; 53.3	EM	-0.00328	-0.00244	0.00528	0.00451	19.5
		NR	0.00055	0.00013	0.00546	0.00459	5.0

Model selection by AIC

Generate data from distributions

↓

The AIC value is computed for each model

↓

We select the model that has smallest AIC

Table 7 (Lognormal)

(μ, σ)	Trunc. (%) Cen. (%)	Sample size	Percentage (%)			Average AIC		
			LN	WB	EP	LN	WB	EP
(3.5, 0.5)	30; 62.3	50	76.9	20.4	0	172.2	173.8	190.6
		100	86.0	14.0	0	344.4	347.6	381.4
		200	91.6	8.4	0	684.9	691.3	759.7
	60; 51.1	50	79.9	20.1	0	216.6	218.2	232.9
		100	85.6	14.4	0	431.5	434.6	465.3
		200	92.4	7.6	0	860.8	866.7	928.1

Table 8 (Weibull)

(μ, σ)	Trunc. (%) Cen. (%)	Sample size	Percentage (%)			Average AIC		
			LN	WB	EP	LN	WB	EP
(3.55, 0.33)	30; 56.8	50	33.0	67.0	0	193.0	191.1	214.1
		100	17.9	82.1	0	383.3	378.7	424.0
		200	8.5	91.5	0	763.8	753.7	847.3
	60; 42.3	50	31.6	68.4	0	240.3	238.0	262.4
		100	17.5	82.5	0	481.7	475.9	526.0
		200	7.2	92.8	0	959.1	948.2	1049.1

Table 9 (Exponential)

η	Trunc. (%) Cen. (%)	Sample size	Percentage (%)			Average AIC		
			LN	WB	EP	LN	WB	EP
30	30; 43.9	50	19.0	11.3	69.7	250.8	249.2	248.3
		100	13.9	12.2	73.9	503.6	499.3	498.3
		200	8.7	12.8	78.5	989.2	967.3	966.8
	60; 41.6	50	22.0	10.5	67.5	260.0	258.7	257.7
		100	17.6	11.8	70.6	518.3	515.0	514.0
		200	9.0	12.2	78.8	1036.4	1028.3	1027.3

Data analysis

1. Transformer lifetime data (Hong et al. 2009)

Original data: 710 observations with 62 failures.

Subset of the data: 286 observations with 39 failures.

If a transformer were installed before 1980 and failed after 1980, we call the sample a truncated unit. If a transformer still in service in March 2008, we call the sample as a censored unit.

- Based on the dataset, we can estimate the parameters of the lognormal, Weibull, and exponential distributions. Then, the AIC for the respective models can be computed.

	Lognormal	Weibull	Exponential
AIC	470.04	472.29	470.67

Table 10

Method					
EM1		EM2		NR	
Step	$(\hat{\mu}, \hat{\sigma})$	Step	$(\hat{\mu}, \hat{\sigma})$	Step	$(\hat{\mu}, \hat{\sigma})$
1	(3.065, 0.968)	1	(3.065, 0.968)	1	(3.065, 0.968)
2	(3.675, 1.209)	2	(4.255, 0.659)	2	(13.156,-3.395)
3	(4.015, 1.292)	3	(4.196, 0.769)		
4	(4.221, 1.339)	4	(4.164, 0.873)		
5	(4.355, 1.379)	5	(4.161, 0.966)		
6	(4.446, 1.418)	6	(4.182, 1.045)		
⋮	⋮	⋮	⋮		
45	(4.961, 1.875)	61	(4.957, 1.872)		
46	(4.962, 1.876)	62	(4.959, 1.873)		
47	(4.963, 1.877)	63	(4.960, 1.874)		
48	(4.964, 1.878)	64	(4.961, 1.875)		

- 2. Data from Balakrishnan and Mitra (2012)

We use the data is given by Table A.1 of Balakrishnan and Mitra (2012). Sample size=100

Truncation percentage=40%.

The data is generated from the Weibull distribution with the parameter vector $(\mu, \sigma) = (3.55, 0.33)$.

	Lognormal	Weibull	Exponential
AIC	425.25	418.83	466.39

Table 11

ε	Method	Step	$(\hat{\mu}, \hat{\sigma})$	Method	Step	$(\hat{\mu}, \hat{\sigma})$
0.001	NR	1	(3.5384, 0.2712)	EM	1	(3.3396, 0.2712)
		2	(3.5382, 0.3566)		2	(3.6653, 0.1297)
		3	(3.5375, 0.3423)		3	(3.6128, 0.1741)
		4	(3.5375, 0.3419)		⋮	⋮
		5	(3.5375, 0.3419)		12	(3.5343, 0.3390)
					13	(3.5354, 0.3400)
					14	(3.5361, 0.3406)
0.000001	NR	1	(3.5384, 0.2712)	EM	1	(3.3396, 0.2712)
		2	(3.5382, 0.3566)		2	(3.6653, 0.1297)
		3	(3.5375, 0.3423)		3	(3.6128, 0.1741)
		4	(3.5375, 0.3419)		⋮	⋮
		5	(3.5375, 0.3419)		12	(3.5343, 0.3390)
		6	(3.5375, 0.3419)		13	(3.5354, 0.3400)
					14	(3.5361, 0.3406)
					⋮	⋮
					28	(3.5375, 0.3419)
					29	(3.5375, 0.3419)
					30	(3.5375, 0.3419)

Conclusion and Discussion

- Lognormal distribution: EM algorithm converges but the NR method diverges frequently.
- Weibull distribution: one-dimensional Newton-Raphson method provide faster convergence rate than the EM algorithms.
- From the simulation, we can know that AIC is a suitable method for model selection.
- Fan and Wang (2011) develop the EM algorithm for the Weibull analysis under a very general step-stress life testing with masked data.

References

- Akaike, H., 1974. A new look at the statistical model identification, IEEE Transactions on Automatic Control **19** (6): 716–723.
- Balakrishnan, N., Cohen, A.C., 1991. Order Statistics and Inference: Estimation Methods. Academic Press, Boston
- Balakrishnan, N., Mitra, D., 2011. Likelihood inference for lognormal data with left truncation and right censoring with an illustration. Journal of Statistical Planning and Inference **141**, 3536-3553.
- Balakrishnan, N., Mitra, D., 2012. Left truncated and right censored Weibull data and likelihood inference with an illustration. Computational Statistics & Data analysis **56**, 4011-4025.
- Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed.).
- Cohen, A.C., 1991. Truncated and Censored Samples. Marcel Dekker, New York.
- Crow, E.L., Shimizu, K., 1988. Lognormal Distributions: Theory and Applications. Marcel Dekker, New York.

- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B **39**, 1-38.
- Emura, T. and Wang, W., 2012, Nonparametric maximum likelihood estimation for dependent truncation data based on copulas, Journal of Multivariate Analysis **110**, 171-188.
- Fan, T.H., Wang, W.L. 2011. Accelerated Life Tests for Weibull Series Systems With Masked Data. IEEE Transaction on Reliability **60**, No.3.
- Hong, Y., Meeker, W.Q., McCalley, J.D., 2009. Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. The Annals of Applied Statistics **3**, 857-879.
- Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distribution-vol. 1, second ed. John Wiley & Sons, New York.
- Lange, K., 1995. A gradient algorithm locally equivalent to the EM algorithm. Journal of the Royal Statistic Society. Series B **57**, 425-437.
- Meeker, W.Q., Escobar, L.A., 1998. Statistical Methods for Reliability Data. John Wiley & Sons, New York.

- Ng, H.K.T., Chan, P.S., Balakrishnan, N, 2002. Estimation of parameters from progressively censored data using EM algorithm. Computational Statistics & Data Analysis **39** (2002) 371 – 386.
- Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6 (2), 461–464.
- Yang, Y., 2005. Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation, Biometrika (2005), **92**, 4, 937–950.

Thanks for your attention.