Analysis of Left Truncated and Right Censored Competing Risks Data

Ayon Ganguly
Department of Mathematics
Indian Institute of Technology Guwahati, India
aganguly@iitg.ernet.in

December 22, 2016

Paper

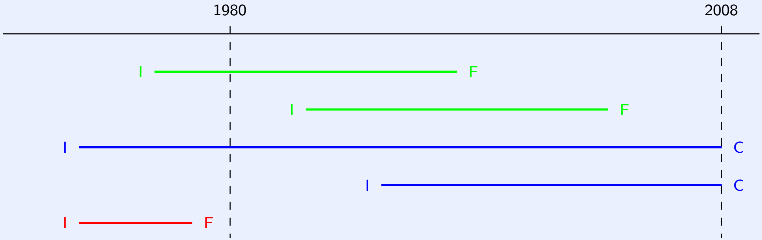
The talk is based on the following paper.

- Kundu, D., Mitra, D., and Ganguly, A. (2016), Analysis of Left Truncated Right Censored Competing Risks Data, *Computational Statistics and Data Analysis* (Accepted for publication).

Main Sections

Description of Data

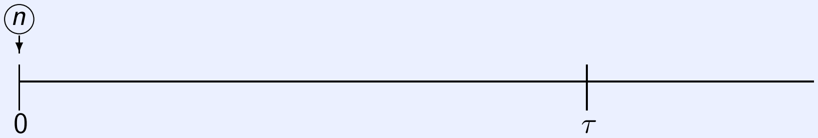
- Hong, Meeker, and McCalley (2009)[1].
- Approximately 150,000 high-voltage power transformers.



[1] Hong Y., Meeker W. Q., and 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.

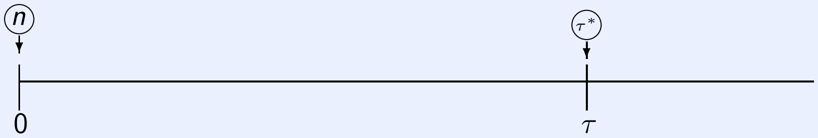
Type-I Censoring

- Known as right censoring.
- n **:** Number of items put on the test.
- τ **:** Pre-fixed time.
- $\tau^* = \tau$ **:** Experiment termination time.



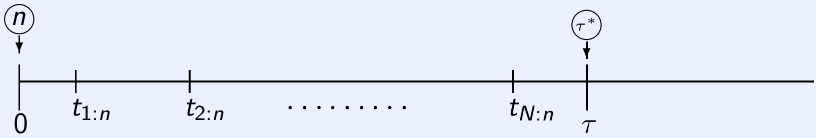
Type-I Censoring

- Known as right censoring.
- n **:** Number of items put on the test.
- τ **:** Pre-fixed time.
- $\tau^* = \tau$ **:** Experiment termination time.



Type-I Censoring

- Known as right censoring.
- n **:** Number of items put on the test.
- τ **:** Pre-fixed time.
- $\tau^* = \tau$ **:** Experiment termination time.

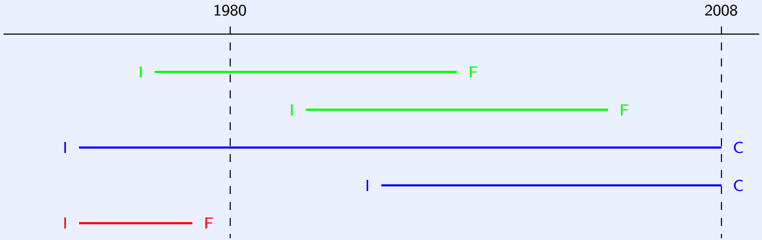


Truncation

- Failure times are observed only when they take on values in a particular range.
- No information about failure times out side this range.
- T has CDF $F(\cdot)$.
- Range is (τ, ∞), where τ is a known constant.
- The CDF of truncated random variable T^* is given by $(F(t) - F(\tau))/(1 - F(\tau))$, if $t \geq \tau$.

Left Truncation and Right Censored Data

- Hong, Meeker, and McCalley (2009)[2].
- Approximately 150,000 high-voltage power transformers.



[2]Hong Y., Meeker W. Q., and 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.

Competing Risks

- Suppose that a each transformer fails due to one of the causes, (i) excessive load or (ii) excessive heating.
- The cause of the failure is known along with the failure time.
- This is known as competing risks in statistical literature.

Competing Risks

- There are mainly two ways to analysis the competing risks data.
  - Latent failure times model approach as suggested by Cox (1959)[3].
  - Cause specific hazard function model as suggested by Prentice et al. (1978)[4].

[3] Cox D. R. (1959), The analysis of exponentially distributed lifetimes with two types of failures, *Journal of the Royal Statistical Society Series B*, 21:411–421.

[4] Prentice R. L., Kalbfleish J. D., Peterson Jr. A. V., Flurnoy N., Farewell V. T., Breslow N. E. (1978), The analysis of failure times in presence of competing risks, *Biometrics*, 34:541–554.

Our Model

- n : Number of items put on the test.
- Left truncated right censored data.
    - T_i : Lifetime of the i-th unit, $i = 1, 2, \ldots, n$.
    - τ_{iL} : Left truncation time for the i-th unit.
    - τ_{iR} : Right censoring time for the i-th unit.
    - ν_i : Truncation indicator. It is 1 if i-th unit is not truncated; 0 if it is truncated.

Our Model

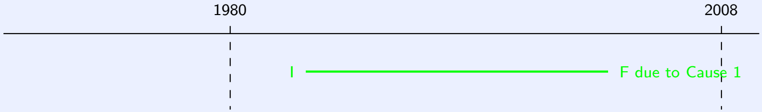
- Two competing risks.
  - Cox's latent failure time model [Cox (1959)[5]].
  - T_{ji} : Latent failure time of the i-th unit under cause j, $j = 1, 2$.
  - $T_i = \min\{T_{1i},\ T_{2i}\}$.
  - δ_i : Cause-censoring indicator. 1 if it fails from cause 1; 2 if it fails from cause 2; 0 if it is censored.

[5] Cox D. R. (1959) The analysis of exponentially distributed lifetimes with two types of failures, *Journal of the Royal Statistical Society Series B*, 21:411–421.

Our Model

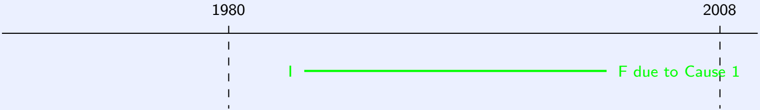
- Weibull distribution with the same shape parameter but different scale parameters.
  - $T_{1i} \underbrace{\sim}_{}^{i.i.d.}$ Weibull (α, λ_1) and $T_{2i} \overset{i.i.d.}{\sim}$ Weibull (α, λ_2).
  - $f(x;\, \alpha,\, \lambda) = \alpha \lambda x^{\alpha - 1} e^{-\lambda x^{\alpha}} \mathbb{1}_{(0,\,\infty)}(x)$.
  - T_{1i} and T_{2i} are independent.

Likelihood Function

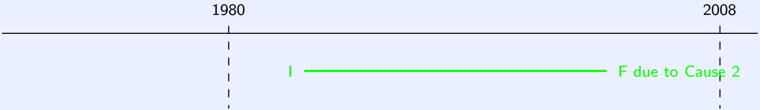


Likelihood contribution: $\alpha\lambda_1 e^{-\lambda_1 t_i^\alpha} \times e^{-\lambda_2 t_i^\alpha}$.

Likelihood Function

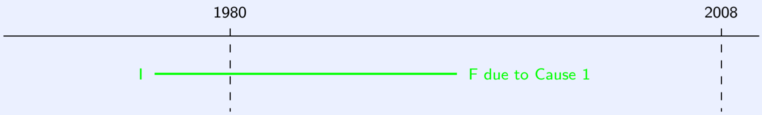


Likelihood contribution: $\alpha\lambda_1 e^{-\lambda_1 t_i^\alpha} \times e^{-\lambda_2 t_i^\alpha}$.



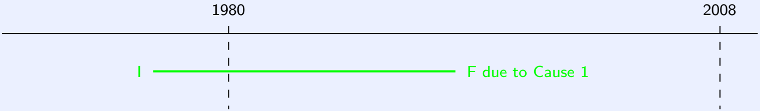
Likelihood contribution: $\alpha\lambda_2 e^{-\lambda_2 t_i^\alpha} \times e^{-\lambda_1 t_i^\alpha}$.

Likelihood Function

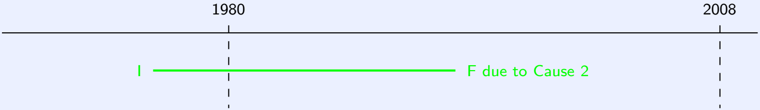


Likelihood contribution: $\frac{\alpha\lambda_1 e^{-\lambda_1 t_i^\alpha} \times e^{-\lambda_2 t_i^\alpha}}{e^{-(\lambda_1+\lambda_2)\tau_{iL}^\alpha}}$.

Likelihood Function

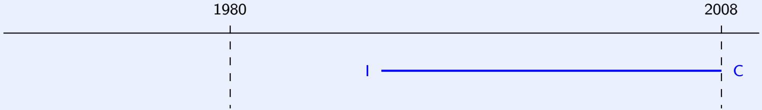


Likelihood contribution: $\dfrac{\alpha\lambda_1 e^{-\lambda_1 t_i^{\alpha}} \times e^{-\lambda_2 t_i^{\alpha}}}{e^{-(\lambda_1+\lambda_2)\tau_{iL}^{\alpha}}}$.



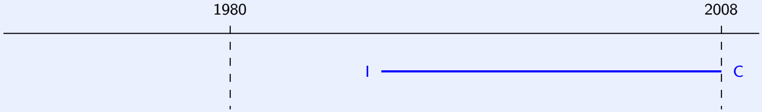
Likelihood contribution: $\dfrac{\alpha\lambda_2 e^{-\lambda_2 t_i^{\alpha}} \times e^{-\lambda_1 t_i^{\alpha}}}{e^{-(\lambda_1+\lambda_2)\tau_{iL}^{\alpha}}}$.

Likelihood Function

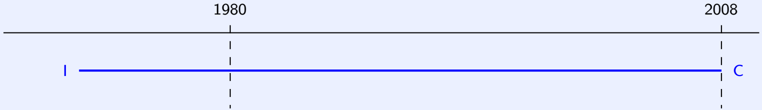


Likelihood contribution: $e^{-(\lambda_1+\lambda_2)t_i^\alpha}$.

Likelihood Function



Likelihood contribution: $e^{-(\lambda_1+\lambda_2)t_i^{\alpha}}$.



Likelihood contribution: $\dfrac{e^{-(\lambda_1+\lambda_2)t_i^{\alpha}}}{e^{-(\lambda_1+\lambda_2)\tau_{iL}^{\alpha}}}$.

Likelihood Function

- Log-likelihood function (ignoring the constant term):

$$l(\alpha,\, \lambda_1,\, \lambda_2) = m \ln \alpha + m_1 \ln \lambda_1 + m_2 \ln \lambda_2 + \alpha w_1 - (\lambda_1 + \lambda_2) w_2(\alpha).$$

  - $w_1 = \sum_{i \in I_1 \cup I_2} \ln t_i$.
  - $w_2(\alpha) = \sum_{i=1}^{n} t_i^{\alpha} - \sum_{i=1}^{n} (1 - \nu_i) \tau_{iL}^{\alpha}$.
  - I_j : Set of indices of failures due to cause j, $j = 1, 2$.
  - $m_j = |I_j|$ and $m = m_1 + m_2$.

MLEs: α known

- Log-likelihood function (ignoring the constant term):

$$l(\alpha, \lambda_1, \lambda_2) = m \ln \alpha + m_1 \ln \lambda_1 + m_2 \ln \lambda_2 + \alpha w_1 - (\lambda_1 + \lambda_2) w_2(\alpha).$$

- The MLEs of λ_1 and λ_2 exist and unique if $m_1 > 0$ and $m_2 > 0$.
- $\widehat{\lambda}_1(\alpha) = \dfrac{m_1}{w_2(\alpha)}$ and $\widehat{\lambda}_2(\alpha) = \dfrac{m_2}{w_2(\alpha)}$.

MLEs: α unknown

- Log-likelihood function (ignoring the constant term):

$$l(\alpha,\, \lambda_1,\, \lambda_2) = m \ln \alpha + m_1 \ln \lambda_1 + m_2 \ln \lambda_2 + \alpha w_1 - (\lambda_1 + \lambda_2) w_2(\alpha).$$

- For fixed α, $\widehat{\lambda}_1(\alpha) = \dfrac{m_1}{w_2(\alpha)}$ and $\widehat{\lambda}_1(\alpha) = \dfrac{m_2}{w_2(\alpha)}$.
- Profile log-likelihood function (ignoring the constant term):

$$p(\alpha) = m \ln \alpha - m \ln w_2(\alpha) + \alpha w_1.$$

- The function $p(\alpha)$ is unimodal.
- MLE of α can be found by maximizing the profile log-likelihood function using some numerical technique, like Newton-Raphson method or fixed point equation.
- $\widehat{\lambda}_1 = \widehat{\lambda}_1(\widehat{\alpha})$ and $\widehat{\lambda}_2 = \widehat{\lambda}_2(\widehat{\alpha})$.

Confidence Interval

- Due to the absence of the closed form MLEs, it is difficult to find the exact distributions of the MLEs and hence exact confidence intervals of the parameters.
- Parametric bootstrap confidence intervals of unknown parameters can be constructed in routine manner.

Conclusions

- Our conclusions and recommendations are based on simulation study.

Conclusions

- Our conclusions and recommendations are based on simulation study.
- The biases and RMSEs for all the parameters decrease as sample size increases.

Conclusions

- Our conclusions and recommendations are based on simulation study.
- The biases and RMSEs for all the parameters decrease as sample size increases.
- The truncation percentage has more effect on the performance of the estimates of shape parameter than on scale parameters.

Conclusions

- Our conclusions and recommendations are based on simulation study.
- The biases and RMSEs for all the parameters decrease as sample size increases.
- The truncation percentage has more effect on the performance of the estimates of shape parameter than on scale parameters.
- The Bayesian estimation can be done.

Conclusions

- Our conclusions and recommendations are based on simulation study.
- The biases and RMSEs for all the parameters decrease as sample size increases.
- The truncation percentage has more effect on the performance of the estimates of shape parameter than on scale parameters.
- The Bayesian estimation can be done.
- The inferential procedures can be easily extended to more than two competing risks.

Conclusions

- Our conclusions and recommendations are based on simulation study.
- The biases and RMSEs for all the parameters decrease as sample size increases.
- The truncation percentage has more effect on the performance of the estimates of shape parameter than on scale parameters.
- The Bayesian estimation can be done.
- The inferential procedures can be easily extended to more than two competing risks.
- One needs a little modification of the inferential procedures if shape parameter are assumed to be different.

Thank You