A Flexible Model with Linear Approximation for Left Truncated Right Censored Data

Ayon Ganguly

Department of Mathematics
Indian Institute of Technology Guwahati
Email: aganguly@iitg.ac.in

International Indian Statistical Association (IISA) Conference 2022

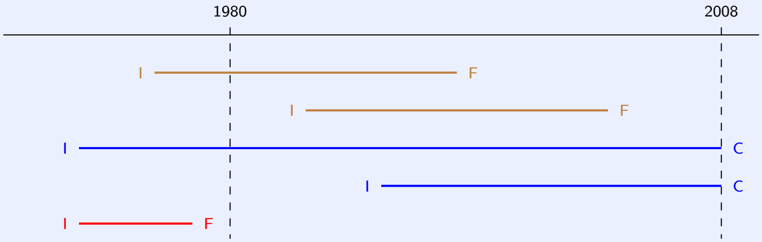
December 29, 2022

Main Sections

1. Left Truncation and Right Censoring
2. Our Model
3. Likelihood Inference
4. Robustness Study
5. Data Analysis
6. Conclusion

Left Truncation and Right Censored Data

- Left truncated and right censored data (LTRC).



LTRC Data: Another example

- Lifetimes were collected from a retirement centre, Channing House, in Palo Alto, California.
- Minimum age of to enter the centre: 60 years
- Individuals entered the centre at different ages after 60 years.
- Starting time of study: 1964.
- Ending time of study: July 01, 1975.
- Total individuals: 462 (F-365, M-97).
- No information before entering centre $\rightarrow$ Left truncation.
- Left truncation time: Age of entering.
- Alive individuals left house before ending time or still alive at the end of the study $\rightarrow$ Right censoring.
- Censoring time: Age at departure or at July 01, 1975.
- Censored individuals: 286 (F-235, M-51).

LTRC Data: Another example

- Available in R-package `boot`.
- First 6 rows of the data is given below.

Sex	Entry	Exit	Time	Cens
Male	782	909	127	1
Male	1020	1128	108	1
Male	856	969	113	1
Male	915	957	42	1
Male	863	983	120	1
Male	906	1012	106	1

Brief Literature Review

- Hong et al. (2009)[1] : Weibull Distribution.

[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.

Brief Literature Review

- Balakrishnan and Mitra (2012)[2]: Weibull Distribution, used EM algorithm.

[2] Balakrishnan N. and Mitra D. (2012). Left truncated and right censored Weibull data and likelihood inference with an illustration, *Computational Statistics and Data Analysis*, 56:4011–4025

Brief Literature Review

- Balakrishnan and Mitra (2014)[3] : Generalized gamma family of distributions, EM algorithm, model discrimination.
- Mitra et al. (2021)[4] : Lehmann family of distribution, stochastic EM, model discrimination.

---

[3] Balakrishnan, N. and Mitra, D. (2014). EM-based likelihood inference for some lifetime distributions based on left truncated and right censored data and associated model discrimination (with discussions). *South African Statistical Journal*, 48, 125–204.

[4] Mitra, D., Kundu, D. and Balakrishnan, N. (2021). Likelihood analysis and stochastic EM algorithm for left truncated right censored data and associated model selection from the Lehmann family of life distributions. *Japanese Journal of Statistics and Data Science*, 4, 1019–1048.

Brief Literature Review

- Emura and Michimae (2022)[5] : Review of parametric models and inferences.

[5] Emura, T., and Michimae, H. (2022). Left-truncated and right-censored field failure data: Review of parametric analysis for reliability. Quality and Reliability Engineering International, 38, 3919–3934.

Covariates

- Presence of co-variates.
- Channing House data: Male or Female.
- Cox's proportional hazards model:

$$\lambda(t, \boldsymbol{x}) = \lambda_0(t) \exp\left(\boldsymbol{x}'\beta\right).$$

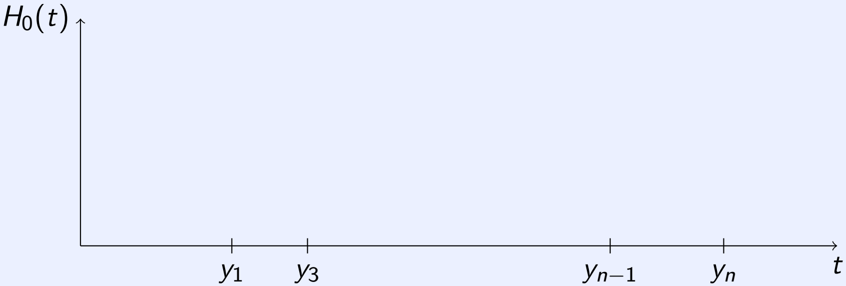
- In terms of cumulative hazard function:

$$\Lambda(t, \boldsymbol{x}) = \Lambda_0(t) \exp\left(\boldsymbol{x}'\beta\right).$$

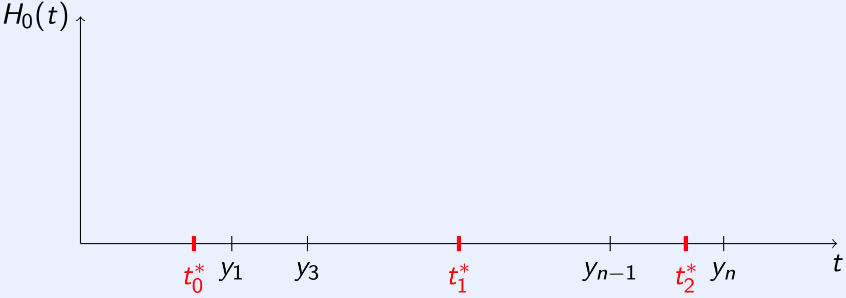
- Parametric or non-parametric estimate.
- Houwelingen and Stijnen (2014)[6] .

[6] van Houwelingen, H. C. and Stijnen, T. (2014). Cox regression model. In: Handbook of Survival Analysis, Eds: Klein, J. P., van Houwelingen, H. C., Ibrahim, J. G., and Scheike, T. H. CRC Press, Boca Raton, FL.

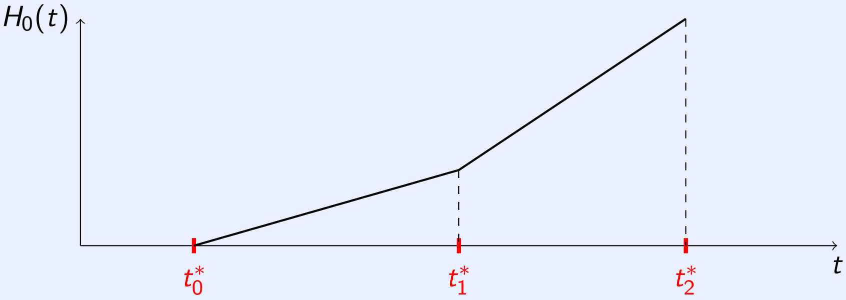
Piecewise Linear Approximation



Piecewise Linear Approximation



Piecewise Linear Approximation



Piecewise Linear Approximation

- Lifetimes: $y_1, y_2, \ldots, y_n$ (observed failures and right censored).
- Truncation times: $\tau_1, \tau_2, \ldots, \tau_n$.
- Cut-points: $t_0^*, t_1^*, \ldots, t_n^*$.
- $t_0^* < t_1^* < \ldots < t_n^*$, $t_0^* \leq \min\{y_{min}, \tau_{min}\}$ and $t_n^* \leq y_{max}$.
- Approximate $\Lambda_0(\cdot)$ by $H_0(\cdot)$:

$$H_0(t) = \sum_{k=1}^{N} (a_k + b_k t) I\left(t_{k-1}^* \leq t < t_k^*\right).$$

- $H_0(t_0^*) = 0 \implies a_1 = 0$.
- $H_0(\cdot)$ increasing $\implies b_k > 0$ for all k.
- $H_0(\cdot)$ continuous at t_k^* for all $k \implies a_k$ can be expressed in terms of b_k's.

Likelihood Inference

- Approximate cumulative hazard function:

$$H(t, \boldsymbol{x}) = H_0(t) \exp\left(\boldsymbol{x}'\beta\right).$$

- Obtain log-likelihood function.
- Number of parameters to estimate:
  Number of cut-points - 1 + Number of co-variates.
- Maximize numerically to obtain MLEs.
- Asymptotic confidence intervals of parameters.

Choice of Cut-points

- Looking at Breslow estimate of cumulative baseline hazard.
- Some suitable quantiles of data.
- AIC / BIC.

Robustness Study

- Simulation based study.
- To check if proposed model fits well to data generated from different parametric models.
- Generate data from a model.
- Estimate based on proposed model.
- Compare closeness.

Robustness Study

- Absolute integrated error (AIE):

$$AIE(x) = \frac{1}{R} \sum_{k=1}^{R} \frac{1}{y_{max,k} - y_{min,k}} \int_{y_{min,k}}^{y_{max,k}} \left| \widehat{H}_{PLA,k}(t,x) - \Lambda(t,x) \right| dt.$$

- Two parent models are considered:
  - Weibull distribution:

$$\Lambda(t, x) = (\lambda t)^{\gamma} e^{-\beta x}.$$

  - PH model with a mixture of Weibull distributions as the cumulative baseline hazard:

$$\Lambda(t, x) = (\lambda_1 t^{\alpha_1} + \lambda_2 t^{\alpha_2}) e^{-\beta x}.$$

Robustness Study

- Single covariate with two levels: 0 and 1.
- Different censoring percentage: 20%, 40%
- Different truncation percentage: 10%, 30%
- Different values of parameters.
- AIE values are not very large (< 0.5).
- PLA-based model is less affected by truncation percentage.
- PLA-based model is more affected by censoring percentage.

Analysis of Channing House Data

- The first 6 rows of data:

Sex	Entry	Exit	Time	Cens
Male	782	909	127	1
Male	1020	1128	108	1
Male	856	969	113	1
Male	915	957	42	1
Male	863	983	120	1
Male	906	1012	106	1

- Sex is used as covariate (Female: $x = 1$, Male: $x = 0$).
- Time transformation: $\frac{t-720}{100}$.

Analysis of Channing House Data

Cut Points	b_1	b_2	b_3	b_4	β	AIC
t_0^*, $q_{0.2}$, $q_{0.5}$, $q_{0.7}$	0.298	0.387	1.161	1.353	-0.306	542.859
t_0^*, $q_{0.2}$, $q_{0.5}$, $q_{0.8}$	0.298	0.387	1.204	1.391	-0.307	542.917
t_0^*, $q_{0.3}$, $q_{0.5}$, $q_{0.7}$	0.333	0.353	1.163	1.354	-0.307	544.014
t_0^*, $q_{0.3}$, $q_{0.5}$, $q_{0.8}$	0.333	0.354	1.205	1.392	-0.309	544.071
t_0^*, $q_{0.2}$, $q_{0.5}$	0.298	0.387	1.273	–	-0.307	541.443
t_0^*, $q_{0.3}$, $q_{0.5}$	0.333	0.353	1.275	–	-0.308	542.597
t_0^*, $q_{0.5}$, $q_{0.7}$	0.340	1.163	1.354	–	-0.308	542.065
t_0^*, $q_{0.5}$, $q_{0.8}$	0.340	1.206	1.393	–	-0.309	542.122

Analysis of Channing House Data

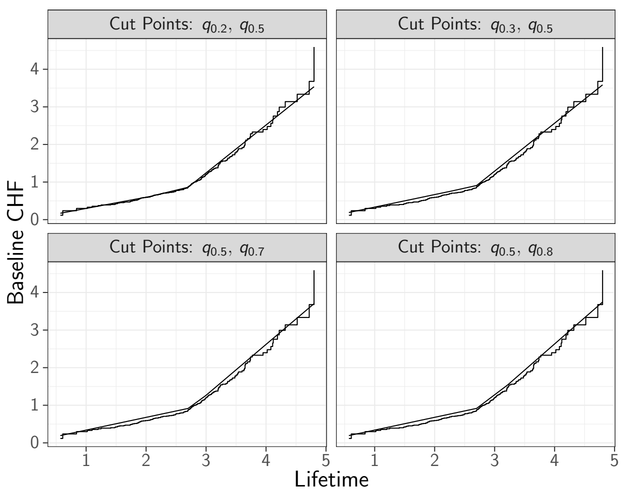


Figure: Plot of estimated baseline CHF with two cut points

Analysis of Channing House Data

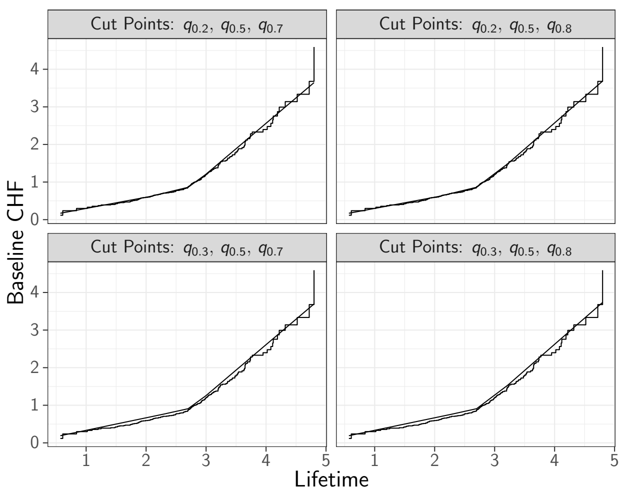


Figure: Plot of estimated baseline CHF with three cut points

Concluding Remark

- PLA based model for predictions of future failures.
  - Breslow estimator is constructed over the observed range of data.
  - PLA based model can be extended beyond the observed data range for making predictions.
- PLA-based model for infant failures.
  - Significant mass at numerical zero.
  - a_1 (the intercept of the first linear piece) > 0.
  - Estimate a_1 based on data.

Thank You