ELEVENTH ANNUAL

STUDENT SHOWCASE

VIRTUAL

APRIL 24, 2020

coloradomesa.edu/showcase

DEPARTMENT:

Statistics

FACULTY SPONSOR:

Dr. Rick Ott

STUDENT(S):

Kendell Visser

PROJECT TITLE:

Kaplan-Meier Survival Curves in Three Samples

Kaplan-Meier Survival Curves in Three-Samples

Author: Kendell Visser
Research Mentor: Dr. Rick Ott

Motivating Question

Can we estimate three survival curves and uphold stochastic ordering?

Introduction

The Kaplan-Meier (KM) Estimator is the most well-known survival estimator in survival analysis; however, using data from a Stanford University study in 1977 comparing degree 1 and degree 2 Carcinoma Cancer patients, there is a violation of an assumption called stochastic ordering. We expect a degree 1 Carcinoma Cancer patient to live longer than a degree 2 patient due to the severity of the disease. We will look at an estimator by Rojo (2004) to force stochastic ordering as we evaluate the Carcinoma data as well as a theoretical three-sample case.

Definitions

- Cumulative Distributive Function (CDF):

the probability that the variable takes a value less than or equal to t, time:

$$F(t) \leq P(X \leq t)$$

For a continuous distribution, this can be expressed mathematically as:

$$F(t) = \int_{-\infty}^{t} f(x)\,dx$$

- Survival Function and the Kaplan-Meier:

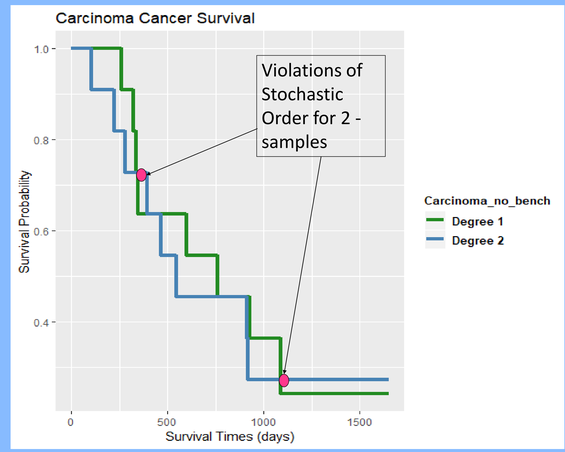
- The survival function, S(t) = 1 – F(t).
The survival function looks at the "Probability of surviving beyond time t".

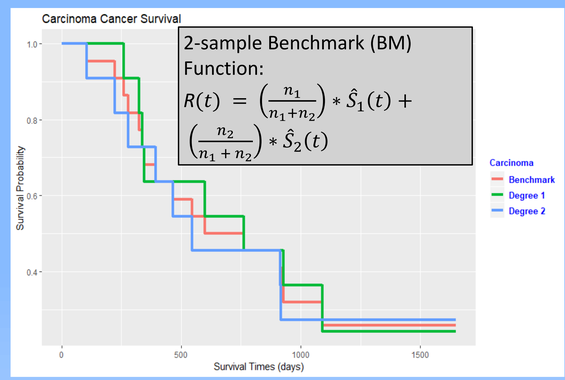
Let t_j = times of death; d_j = deaths at time t_j; n_j = number of participants remaining in the study at time t_j.
For $t \in [t_j, t_{j+1}), j = 1, 2, 3, \ldots,$ we have…

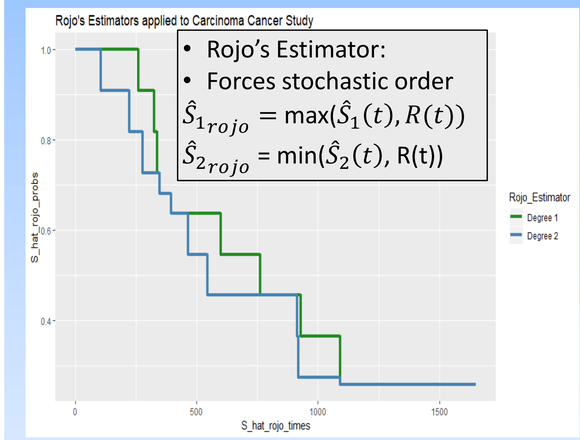
Kaplan-Meier Estimator: $\hat{S}(t) = \Pi_{i=1}^{j}\left(1 - \frac{d_i}{n_i}\right)$.

- Stochastic Ordering: Let $X_1(t) \sim F_1(t)$ and $X_2(t) \sim F_2(t)$ and $F_1(t) \leq F_2(t)$ for all t. In terms of respective survival functions :$S_2(t) \leq S_1(t)$ for all t.

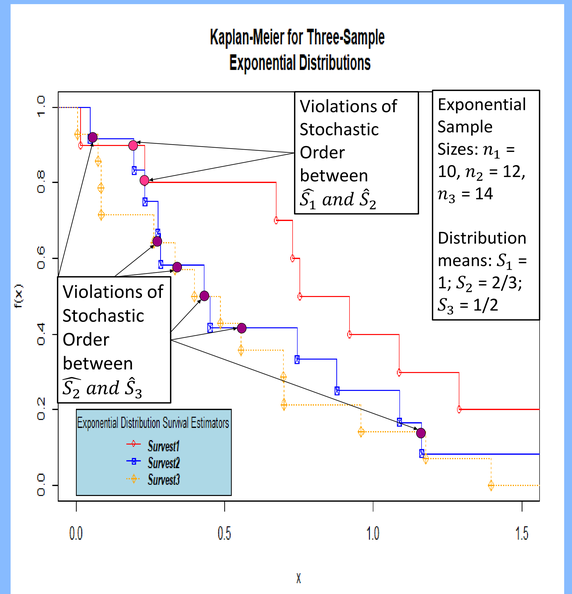
Carcinoma Cancer Patients – 2 sample Kaplan-Meier Survival Curves

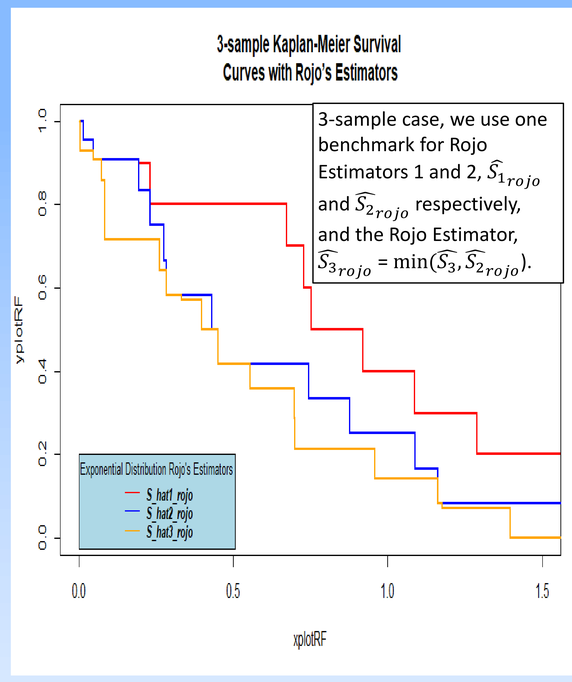






3-sample Exponential Kaplan-Meier Survival Curves





Future Statistical Research and Conclusion

Conclusions:

- We were successfully able to estimate three survival curves that uphold Stochastic order, where $\widehat{S_1}$ and $\widehat{S_2}$ by the benchmark to set Rojo's Estimators.
- We achieved getting $\widehat{S_{3_{rojo}}}$ by taking the $\min\left(\widehat{S_{2_{rojo}}}, \widehat{S_3}\right)$ where $\widehat{S_3}$ is the Kaplan – Meier survival curve.

Future Research:

- We can find the Benchmark (weighted average) between $\widehat{S_{2_{rojo}}}$ and $\widehat{S_{3_{rojo}}}$ instead and then find the $\widehat{S_1}$ $\max\left(\widehat{S_{1_{Rojo}}}, \widehat{S_1}\right)$ where $\widehat{S_1}$ is the Kaplan – Meier survival estimator.
- Finding the bias, mean squared error, and root mean squared errors for the three-sample case sampling from various distributions such as Weibull and Gamma. This can be done for the technique when the benchmark is between $\widehat{S_1}$ and $\widehat{S_2}$, as well as the case when the benchmark is found between $\widehat{S_2}$ and $\widehat{S_3}$.
- Apply similar model to up and coming data on CO-vid 19.

Acknowledgments

The preliminary work was supported through NSA grant number H98230-19-1-0016 and NSF grant DMS-1731082, to Dr. Javier Rojo PI, through the RUSIS @ Oregon State University during an REU in the Summer of 2019.
Additionally, I would like to thank Dr. Rojo for the inspiration of this project and Dr. Friedman for the opportunity to further my research in Survival Analysis and I would also like to thank Dr. Ott for advising the research within this project! Also, thank you to Dr. King for helping with the legends of the graphs in ggplot.

References

"1.3.6.2 Related Distributions." *Engineering Statistics Handbook*, NIST and SEMATECH
Fischer, Ismor. *Estimation: Kaplan-Meier Product-Limit Formula*. 29 May 2012
Gandara, Denise and Hardt, Marie and Visser, Kendell. *Estimation of Survival Functions Under Stochastic Order* 22 August 2019
Kabacoff, Robert I. "Line Charts." *Data Camp: Quick-R*
Lukas J A Stalpers & Edward L Kaplan (2018) Edward L. Kaplan and the Kaplan-Meier Survival Curve, BSHM Bulletin: Journal of the British Society for the History of Mathematics, 33:2, 109-135, DOI: 10.1080/17498430.2018.1450055
Martin, Laura J. "Types of Carcinoma: Basal Cell, Squamous Cell, and Adenocarcinoma." *WebMD*, WebMD, 15 Jan. 2018
Rojo, J. (2004). On the estimation of survival functions under a stochastic order constraint. Lecture Notes-Monograph Series, 44:37{61.
Rojo, J. and Jiang, F. (2013). Estimation of survival functions under stochastic ordering constraint. Unpublished manuscript, Statistics Department, Rice University, 2013.
Rojo, J. and Ma, Z. (1996). On the estimation of stochastically ordered survival functions. Journal of Statistical Computation and Simulation, 55:1-21.
Zabor, Emily C. "Survival Analysis in R." *Survival Analysis in R*