The Impact of Censoring in the Uncertainty of a Parametric Exponential Survival Model

Grant Schissler and Alexia Spoon

Abstract

Censoring is a vital part of survival analysis, where the event in question is not observed within the scope of the study for a given individual. This project aims to assess the exact impact that censoring plays in quantifying the uncertainty of a parametric exponential survival model through the 95% confidence interval. As the rate of censoring increases from 0% to 90%, the 95% confidence interval of the parametric model widens, indicating that there is an increase in uncertainty and variability associated with the increase in censored observations. This increase in uncertainty can be credited to the bias of the maximum likelihood estimator for the exponential distribution as censoring is introduced.

Background

- Parametric models are commonly used in survival analysis, as they tend to give more precise estimates when the model fits the data well
- Much is known about the role censoring plays in nonparametric models, i.e. Kaplan-Meier estimate
- Less is known about the role random censoring plays in fully parametric survival models
- Simulated data will allow for complete control over all aspects of data to determine the contribution of censoring
- Previous studies mainly focus on whether random censoring has any effect on model uncertainty, not the reasoning behind the contribution

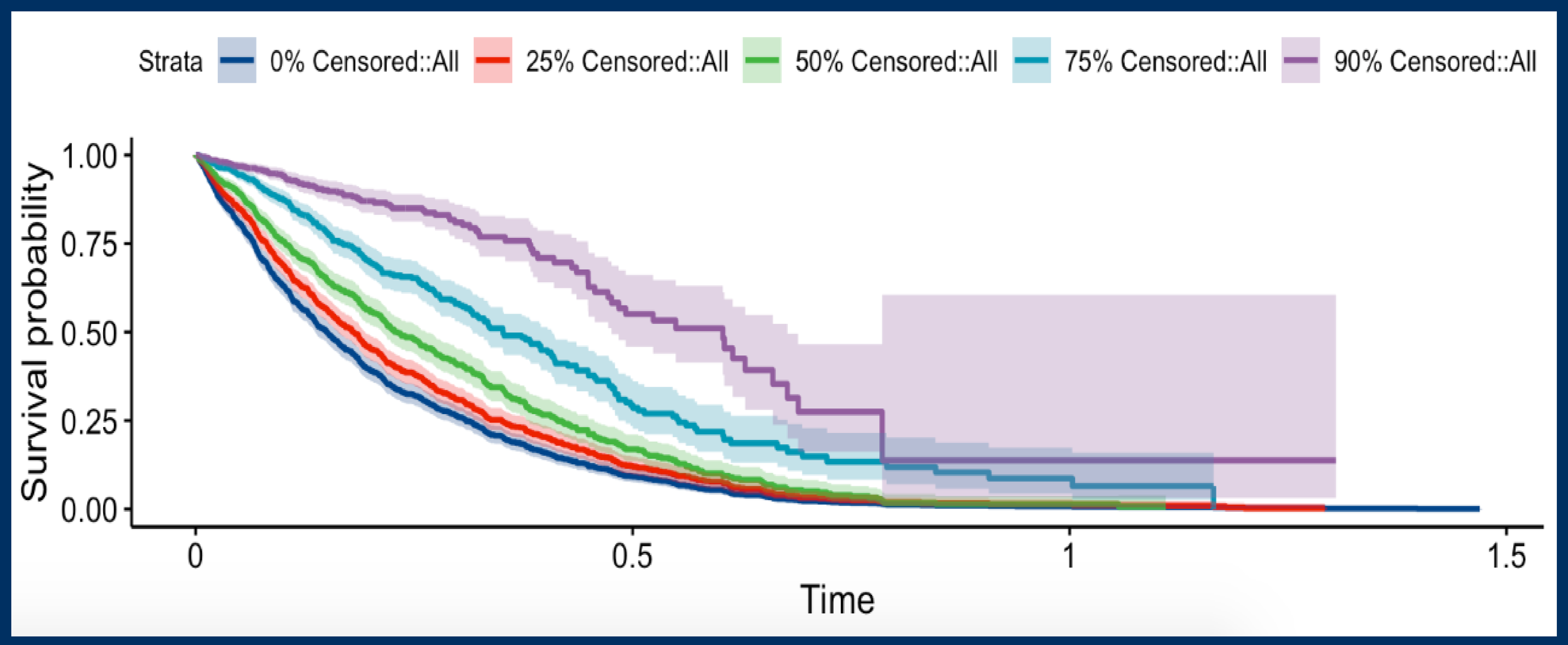

Figure 1.
Figure 1 depicts the nonparametric survival functions with 95% confidence intervals with increasing rates of censoring.

Data and Design

- Simulated 1000 data points from an exponential distribution with rate 5, which will serve at the event times for the survival model. The probability density function and survival function are:

$$f(x) = \lambda e^{-\lambda x} \text{ and } S(x) = e^{-\lambda x}$$

- Random Censoring: Sample of observations chosen as “censored”, with the censoring time being a random value from time 0 to event time for the given observation.
- Maximum likelihood estimate is used in each fully parametric survival model. Using the density function and survival function above, the likelihood will have the form:

$$L(\lambda) = \prod_{i \in D} \lambda e^{-\lambda x_i} \prod_{i \in R} e^{-\lambda C_r},$$

which yields the closed form:

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} x_i}$$

- Confidence intervals for each survival curve constructed to determine uncertainty of model with given rate of censoring

Results

- Figure 1 shows the nonparametric models with increasing rates of censoring
  - The variance of the Kaplan Meier estimate has the form:

$$\hat{V}\left[\hat{S}(t)\right] = \hat{S}(t)^2 \sum_{t_i \le t} \frac{d_i}{Y_i(Y_i - d_i)},$$

which will increase when the number at risk remains large with high rates of censoring

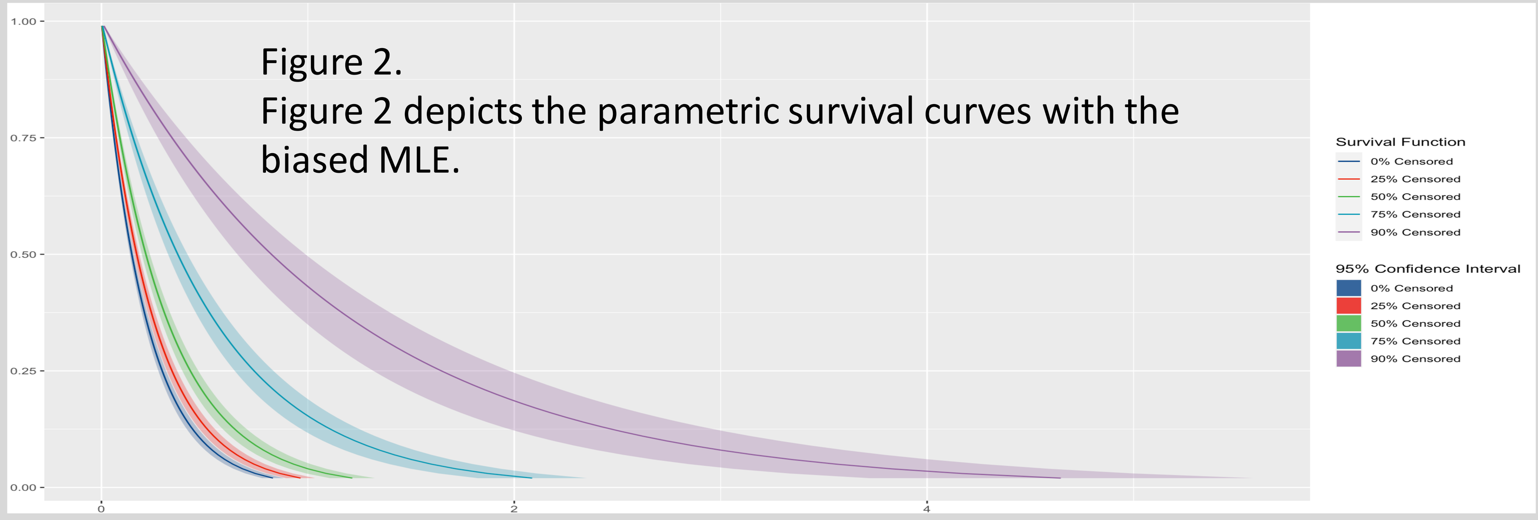

Figure 2.
Figure 2 depicts the parametric survival curves with the biased MLE.

- Figure 2 shows the parametric models with increasing rates of censoring using random censoring
  - As the rate of censoring increases, the maximum likelihood estimate for the parametric survival model will decrease
  - The mean and standard deviation of the model will increase, which will cause the 95% confidence interval to widen
- The maximum likelihood estimate is greatly biased as the rate of censoring increases
- Table 1 shows the simulated bias from the simulated maximum likelihood estimate with the defined rate parameter of 5

Table 1. Bias of Simulated Maximum Likelihood Estimate from Parameter Rate = 5

	0% Censoring	25% Censoring	50% Censoring	75% Censoring	90% Censoring
MLE Estimate	4.707233	4.054999	3.218483	1.874614	0.8417043
Bias	-0.292767	-0.9450014	-1.781517	-3.125386	-4.158296

Discussion

- Figure 3 shows the bias corrected point estimates of the survival curves for the special case where the censoring is at the exponentially distributed event times
  - Bias corrected λ estimate in this case would be $\widehat{\lambda_{BC}} = \left({}^{n-1}/_{\sum_{i=1}^{n} \delta_i}\right)\hat{\lambda}$
  - In this case, all survival curves are now stacked on top of one another, showing that the bias correction fully eliminates bias
- We can confirm from this that there is heavy bias in parametric survival models as censoring increases
- Conclusions:
  - As the rate of censoring increases, the parametric survival model becomes increasingly biased, yielding larger confidence intervals
  - It is important to take care when interpreting survival function estimates when censoring is high as these estimates will be biased and not fully informative
- Further research must be conducted in order to find the exact bias when the random censoring follows a uniform distribution from 0 to the observation's event time.
  - The closed form of the MLE now includes a combination of exponential times and uniform times in the denominator

References

- Klein, John, and Melvin L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data.* New York: Springer, 2003. Print.
- Pettitt, A. N., et al. "Bias Correction for Censored Data with Exponential Lifetimes." *Statistica Sinica*, vol 8, no. 3, Institute of Statistical Science, Academia Sinica, 1998, pp. 941-963, https://www.jstor.org/stable/24306472.
- Zhao, Yuchao, and H. Christpoher Frey. "Quantification of Variability and Uncertainty for Censored Data Sets and Application to Air Toxic Emission Factors." *Risk Analysis*, vol 24, no. 4, 2004, pp.1019-1034., https://doi.org/10.1111/j.0272-4332.2004.00504.x.

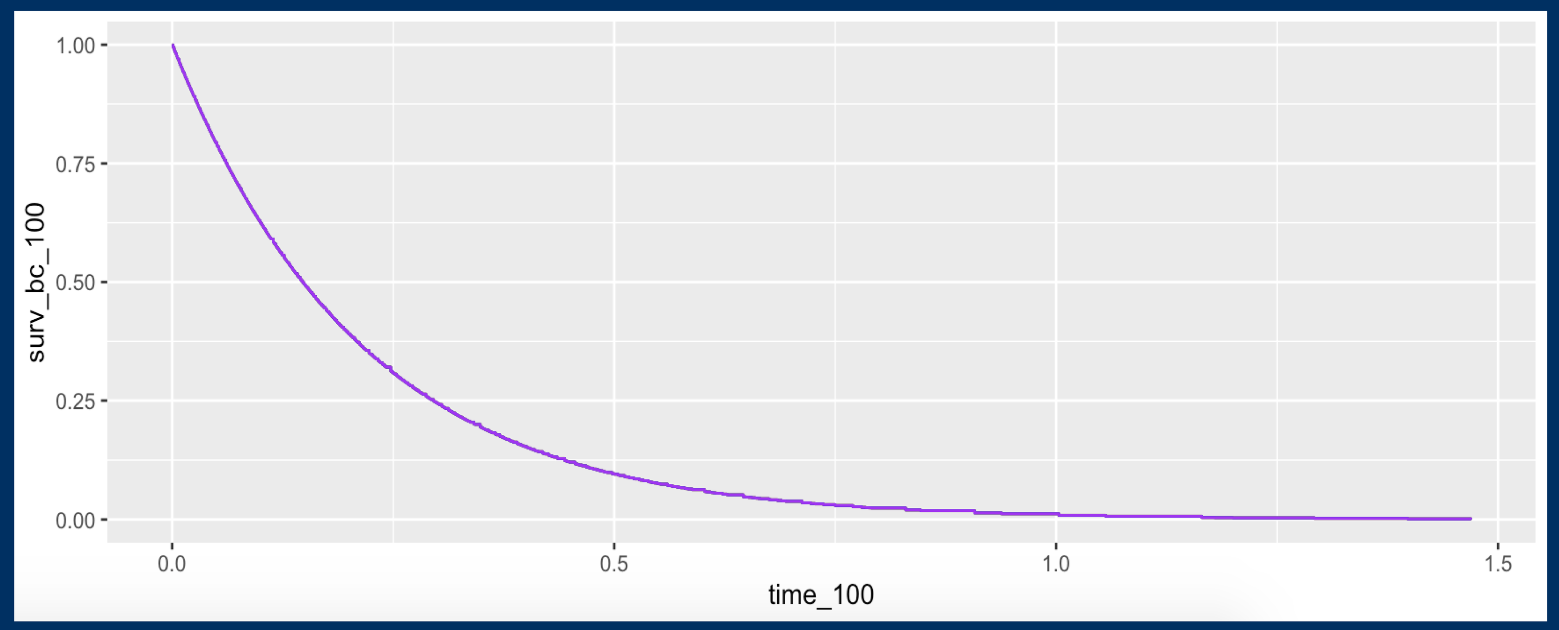

Figure 3.
Figure 3 depicts the point estimates of the survival curves when censoring is at exact exponential event times with bias correction.

Figure 1.
Figure 1 depicts the nonparametric survival functions with 95% confidence intervals with increasing rates of censoring.

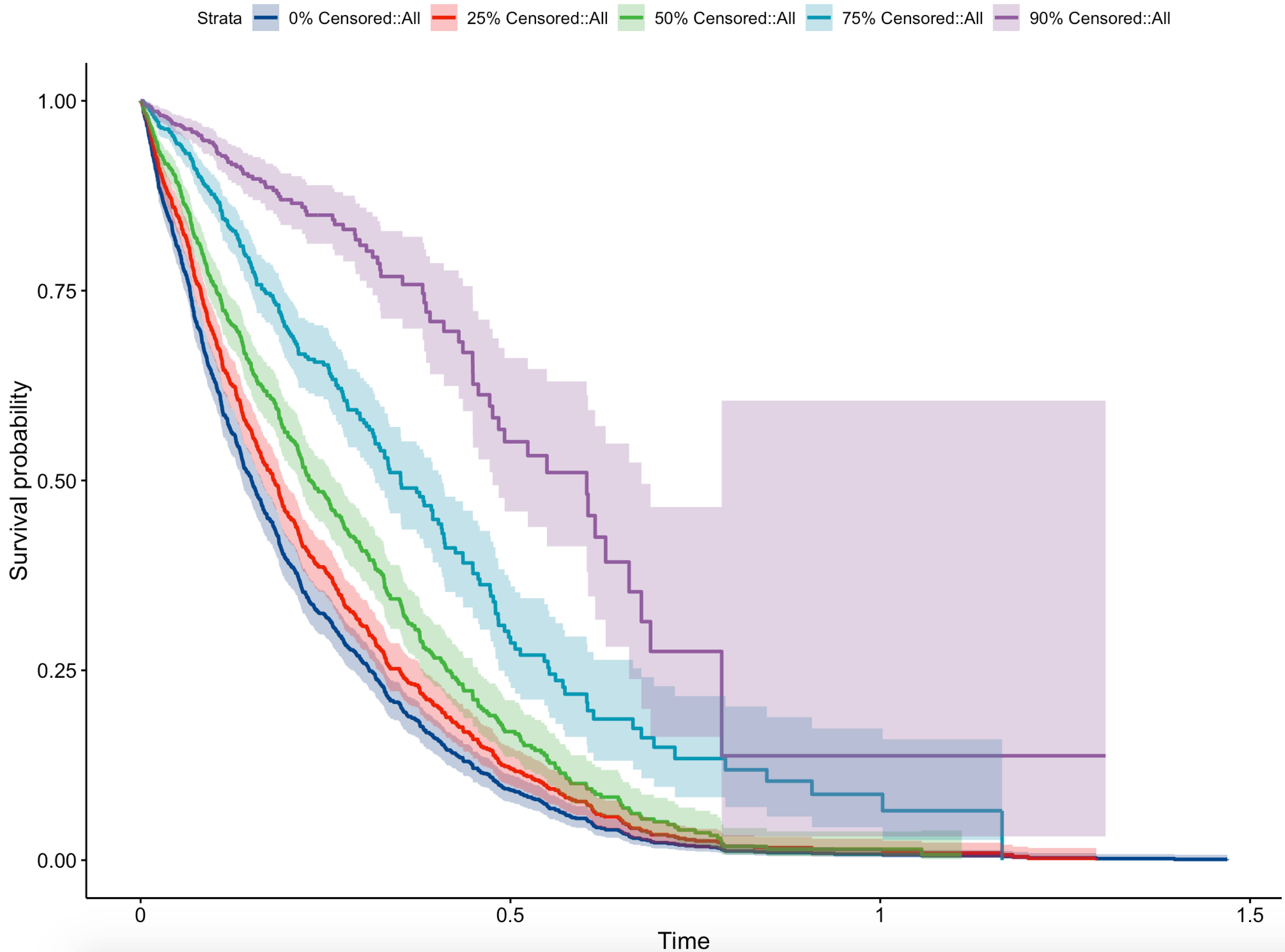

Figure 2.

Figure 2 depicts the parametric survival curves with the biased MLE.

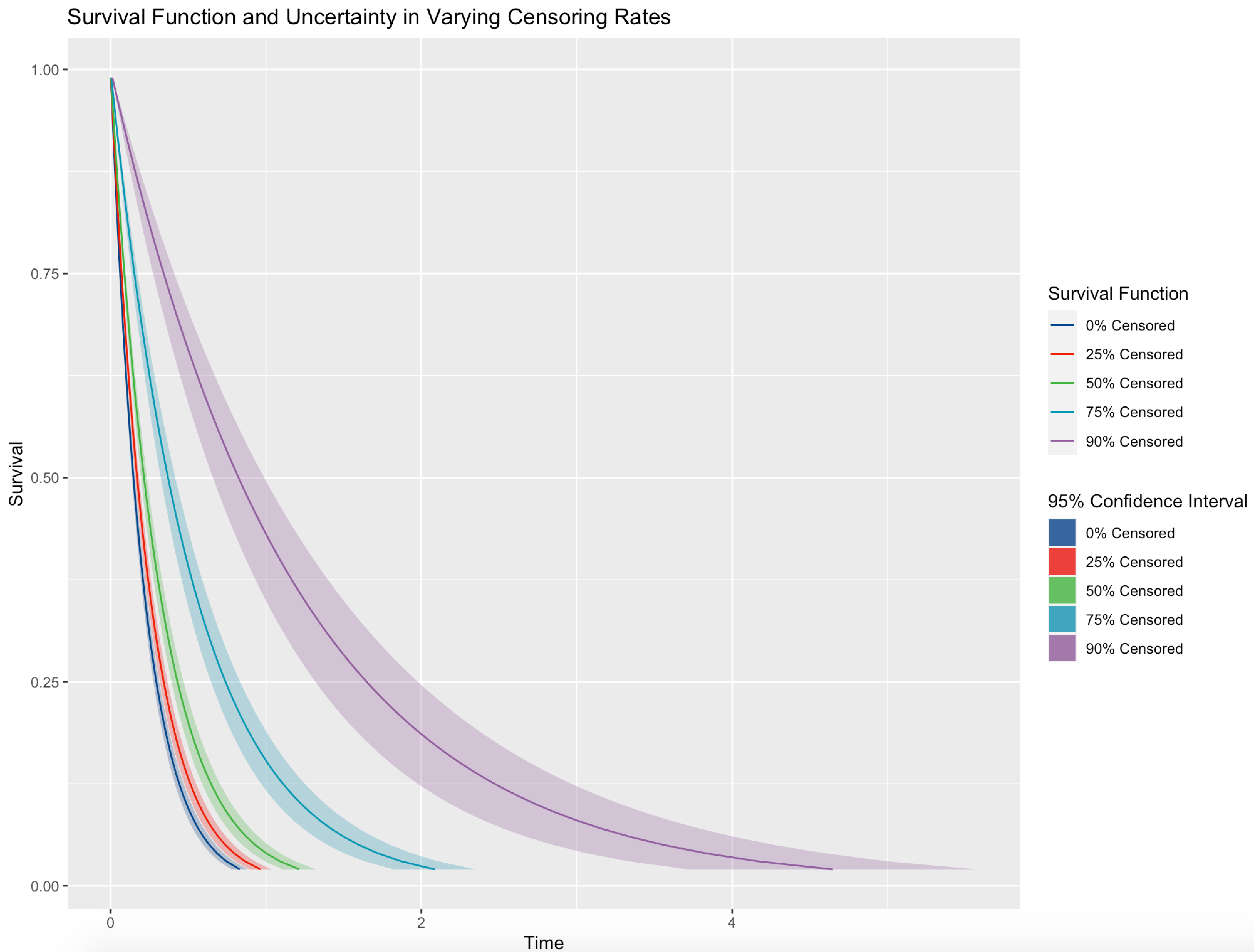

Figure 3.
Figure 3 depicts the point estimates of the survival curves when censoring is at exact exponential event times with bias correction.

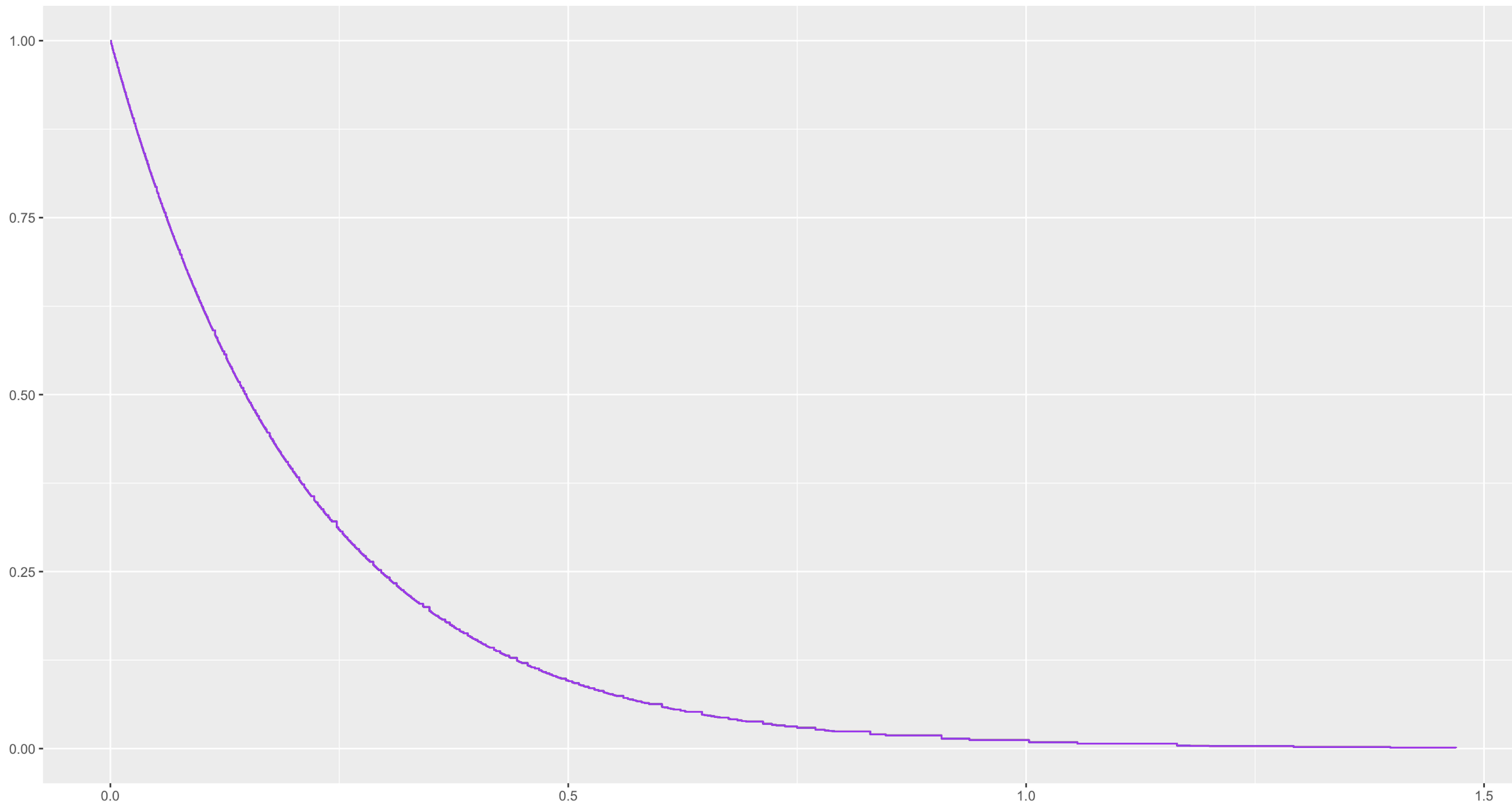

Table 1. Bias of Simulated Maximum Likelihood Estimate from Parameter Rate = 5					
	0% Censoring	25% Censoring	50% Censoring	75% Censoring	90% Censoring
MLE Estimate	4.707233	4.054999	3.218483	1.874614	0.8417043
Bias	-0.292767	-0.9450014	-1.781517	-3.125386	-4.158296