Bootstrapping confidence intervals (BCa) of the maximum likelihood estimator of components in a series systems from masked failure data

Alex Towell

Abstract

We estimate the parameters of a series system with Weibull component lifetimes from relatively small samples consisting of right-censored system lifetimes and masked component cause of failure. Under a set of conditions that permit us to ignore how the component cause of failures are masked, we assess the bias and variance of the estimator. Then, we assess the accuracy of the boostrapped variance and calibration of the confidence intervals of the MLE under a variety of scenarios.

Contents

0.0.1 Assessing the Bootstrapped Confidence Intervals

Our primary interest is in assessing the performance of the BCa confidence intervals for the MLE. We will assess the performance of the BCa confidence intervals by computing the coverage probability of the confidence intervals. Under a variety of scenarios, we will bootstrap a 95%-confidence interval for $\boldsymbol{\theta}$ using the BCa method, and we will evaluate its calibration by computing the coverage probability and its precision by assessing the width of the confidence interval.

The coverage probability is defined as the proportion of times that the true value of $\boldsymbol{\theta}$ falls within the confidence interval. We will compute the coverage probability by generating R datasets from the Data Generating Process (DGP) and computing the coverage probability for each dataset. We will then aggregate this information across all R datasets to estimate the coverage probability.

0.1 Sensitivity of the MLE to Simulation Parameters

Estimating component parameters in the presence of masking, censoring, and other issues like poorly designed systems can be challenging, e.g., multiple local maxima, ridges, and flat regions on the surface of the likelihood function.

0.1.1 Issues Influencing the MLE

1. **Masking Probability**: These aspects significantly impact the MLE's accuracy and precision. By varying the shape or scale parameter of a single component, we explore how they affect the sensitivity of the sampling distribution of the MLE. Sections 0.3.3 and **??** delve into these areas.

2. **Masking Probability**: Candidate sets can be constructed in many ways, and our likelihood model is not robust to all of them.

Identifiability refers to the unique mapping of the model parameters to the likelihood function. Lack of identifiability can lead to multiple sets of parameters that explain the data equally well, making inference about the true parameters challenging. This can arise in our likelihood model in some situations.

For instance, if an analyst has a procedure that identifies when a larger component has failed, but not which of the smaller components failed in that larger component, then the parameters of the components in that larger component are interchangeable.

when candidate sets are constructed in a way that makes the likelihood function flat or nearly flat. For example, if the candidate set

his can lead to non-identifiability in the likelihood model, where the parameters of some components are indiscernible. Although such occurrences might arise by chance in our Bernoulli candidate set model, they can introduce bias in the MLE, especially as the masking probability p increases.

3. **Right-Censoring**: While we do not explore this empirically, the presence of aggressive right-censoring can lead to bias in the MLE by pushing it to estimate larger MTTF (Mean Time To Failure) for the system components. The phenomenon is well-documented, and readers can refer to Klein and Moeschberger (2005) for further insights. This potential bias is something to keep in mind during experimental design.

0.1.2 Addressing Non-Identifiability and Convergence

A flat or nearly flat likelihood function can cause numerical methods like Newton-Raphson to take a long time to converge. We mostly disregarded identifiability issues in our simulation study, only discarding datasets that failed to converge within 125 iterations. Extreme masking probabilities were the exceptions.

0.1.3 Effect of Masking the Component Cause of Failure

Masking the component cause of failure can introduce biases in the MLE. The effect can be more pronounced for components frequently appearing in candidate sets. Details on how this bias affects the MLE for the shape and scale parameters of a Weibull component are provided in this section, including the interplay between masking and right-censoring effects.

0.2 Other Considerations

These aspects mentioned above comprise some of the challenges in estimating the parameters of the components in our likelihood model. We must also consider the impact of simulation parameters like sample size, which will be further discussed in the relevant parts of the simulation study. The broader implications, combined with subsequent bootstrapping of confidence intervals, shape our assessment of the performance of the BCa confidence intervals and the overall method.

0.3 Sensitivity of the MLE to Simulation Parameters

0.3.1 Challenges and Issues Explored in Our Simulation Study

In our simulation study, we explore the sensitivity of the sampling distribution of the MLE, not just to the sample size, but to the masking probability p and the presence of a least reliable component by varying the shape parameter or scale parameter of a single component.

Briefly, we consider some examples which may arise that make it difficult to estimate the parameters of the components of a series system under our likelihood model:

1. **Candidate Sets Construction**: Candidate sets are constructed such that component 1 is present if and only if component 2 is present. In this case, the parameters of components 1 and 2 are not identifiable in our likelihood model. This may occur if an analyst can only identify a larger failed component without specifying the smaller components within it.[1]

   In our Bernoulli candidate set model, such constructions for candidate sets can arise only by chance. In Section 0.3.3, we explore this issue by assessing the effect of varying the masking probability p in the Bernoulli candidate set model on the MLE. It is not the pathological case discussed above, where the parameters of the components are not identifiable, but as the masking probability p increases, the

[1] It may be possible to combine the components into a single larger component to make the model identifiable.

MLE becomes less accurate and precise since there is less information in the data set about component causes of failure.

2. **Least Reliable Component**: If a series system has a significantly less reliable component that causes every system failure, the data may only contain information about that component's parameters.

   We explore this issue in Section **??** by assessing the effect of varying the reliability of a single component on the MLE.

3. **Aggressive Right-Censoring**: If the right-censoring time τ is too short, the data may not contain enough information to estimate the parameters of any of the components. This is a problem with the design of the experiment, not with the likelihood model.

   We do not explore this issue, but it is something to keep in mind when designing experiments. We do comment on the expected effect of right-censoring on the MLE in Section 0.3.2.

Due to censoring and masking, by chance it may be the case that the likelihood Function conditioned on particular data sets is flat (non-identifiable) or nearly flat, which can cause numerical methods like Newton-Raphson to take a long time to converge to a solution. We largely ignored identifiability issues in our simulation study, with the exception that we discarded any data sets that did not converge to a solution after 125 iterations.[2] A failure to converge within 125 iterations could be seen as evidence of potential identifiability issues, in which case one might argue that this data set is not informative enough about the parameters of the components in our likelihood model.

Nonetheless, such scenarios occurred infrequently, usually less than 1%, except in extreme cases where the masking probability was upwards of 75%. However, during the bootstrapping of confidence intervals, we included all solutions, whether they converged to a solution or not. This worst-case analysis approach was adopted because a primary objective was to assess the performance of the BCa confidence intervals. We were concerned that if we took any additional steps, we may unintentionally bias the results in favor of producing narrow BCa confidence intervals with good coverage.

0.3.2 Effect of Right-Censoring

Right-censoring introduces a source of bias in the MLE. Right-censoring has the effect of pushing the MLE to estimate a larger MTTF for each of the components, so that the series system has a larger MTTF. This is because when we observe a right-censoring event, we know that the system failed after the censoring time, but we do not know precisely when it will fail. This uncertainty has the effect of pushing the MLE to estimate a larger MTTF for the system so that it is more likely to fail after the censoring time. See Klein and Moeschberger (2005) for more information on this phenomenon.

To increase the MTTF of a series system, the mean time to failure (MTTF) of each component is increased. By Equation (**??**), The MTTF for the j^{th} is given by

$$\text{MTTF}_j = \lambda_j \Gamma(1 + 1/k_j),$$

therefore, in order to increase the MTTF of the components, lower values for the shape parameters are chosen and higher values for the scale parameters are chosen. Depending on the values of the shape and scale parameters, either the shape or scale parameter may have a larger effect on the MTTF.

0.3.3 Effect of Masking the Component Cause of Failure

When we observe a system failure, we know that one of the components in the candidate set caused the system to fail, but we do not know which one. This uncertainty has the effect of pushing the MLE to estimate a smaller MTTF for each of the components in the candidate set. For components that are frequently in candidate sets but proportionally not more likely to be a component cause of failure, the effect is more pronounced, which may introduce a source of bias in the MLE for such components.

[2]The choice of 125 iterations was driven by the computational demands of the simulation study combined with the subsequent bootstrapping of the confidence intervals.

In our Bernoulli candidate set model, the masking probability p determines how commonly each non-failed component is in the candidate set, and so we expect that as p increases, this will become a more pronoucned source of bias.[3] However, note that the effect of masking, which pushes the MLE to estimate a smaller MTTF, has opposite effect to that of right-censoring, which pushes the MLE to estimate a larger MTTF. As these two sources of bias compete with each other, it is not clear which one will dominate

In what follows, we explain how the bias induced by masking the component cause of failure effects the MLE for the shape and scale parameters of a Weibull component. Assessing Equation (**??**), we see that the MTTF of a Weibull component is proportional to its scale parameter λ_j, which means when we decrease the scale parameter λ_j (keeping the shape parameter k_j constant), the MTTF decreases. Therefore, if the j^{th} component is in the candidate set, to make it more likely to appear in the candidate set, its scale parameter should be decreased, potentially biasing the MLE for the scale parameter downwards.

Conversely, we see that the MTTF decreases as we increase the shape parameter k_j. Therefore, if the j^{th} component is in the candidate set, to make it more likely to appear in in the candidate set, its shape parameter should be increased, potentially biasing the MLE for the shape parameter upwards.

> NOTE TO SELF: The right-censoring has an effect best seen by the MTTF of the series system, and consequently the components. The masking probability, on the other hand, has an effect best seen by the probability of component failure. Additionally, when looking at the plots in the sim study where we vary k_3, we see that the effect of masking is more pronounced for ... finish these thoughts after looking at the plots.

0.3.4 Assessing the Impact of Different Component Reliabilities

In Figure 1, the top row shows two components with roughly the same shape and scale parameters, and the bottom row shows two components with different shape and scale parameters. Every component has the same MTTF, but they have vastly different hazard and survival functions (and vastly different probabilities of being the cause of a system failure). The top system is well-designed, with no weak link in the chain, while the bottom system is poorly designed and difficult to understand.

In this section, we vary the scale or shape parameter of one of the components to examine the effect this has on the MLE and its sampling distribution.

References

Klein, J. P. and Moeschberger, M. L. (2005). *Survival analysis: techniques for censored and truncated data.* Springer Science & Business Media.

[3] In a more complicated candidate set model, it is possible that masking could introduce a significant source of bias for some components, and none at all for others.

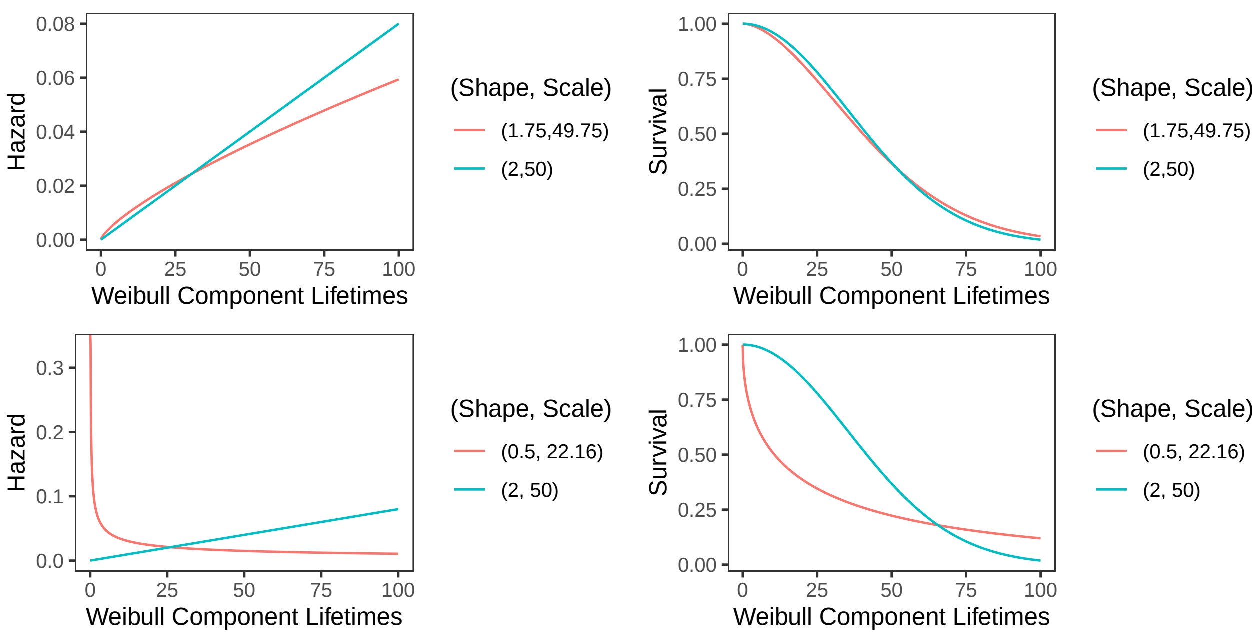

Figure 1: Two components (in series configuration). All components have the same MTTF. On the top, both components have a similiar aging process. On the bottom, the red component has a burn-in process and the blue component has an aging process.