Bootstrapping statistics 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

1 Introduction

Accurately estimating the reliability of individual components in multi-component systems is an important problem in many engineering domains. However, component lifetimes and failure causes are often not directly observable. In a series system, only the system-level failure time may be recorded along with limited information about which component failed. Such *masked* data poses challenges for estimating component reliability.

In this paper, we develop a maximum likelihood approach to estimate component reliability in series systems using right-censored lifetime data and candidate sets that contain the failed component. The key contributions are:

1. Deriving a likelihood model that accounts for right-censoring and masked failure causes through candidate sets. This allows the available masked data to be used for estimation.
2. Validating the accuracy, precision, and robustness of the maximum likelihood estimator through an extensive simulation study under different sample sizes, masking probabilities, and censoring levels.
3. Demonstrating that bootstrapping provides well-calibrated confidence intervals for the MLEs even with small samples.

Together, these contributions provide a statistically rigorous methodology for learning about latent component properties from series system data. The methods are shown to work well even when failure information is significantly masked. This capability expands the range of applications where component reliability can be quantified from limited observations.

The remainder of this paper is organized as follows. First, we detail the series system and masked data models. Next, we present the likelihood construction and maximum likelihood theory. We then describe the bootstrap approach for variance and confidence interval estimation. Finally, we validate the methods through simulation studies under various data scenarios and sample sizes.

2 Series System Model

Consider a system composed of m components arranged in a series configuration. Each component and system has two possible states, functioning or failed. We have n systems whose lifetimes are independent and identically distributed (i.i.d.). The lifetime of the i^{th} system denoted by the random variable T_i. The lifetime of the j^{th} component in the i^{th} system is denoted by the random variable T_{ij}. We assume the component lifetimes in a single system are statistically independent and non-identically distributed. Here, lifetime is defined as the elapsed time from when the new, functioning component (or system) is put into operation until it fails for the first time. A series system fails when any component fails, thus the lifetime of the i^{th} system is given by the component with the shortest lifetime,

$$T_i = \min\{T_{i1}, T_{i2}, \ldots, T_{im}\}.$$

There are three particularly important distribution functions in survival analysis: the survival function, the probability density function, and the hazard function. The survival function, $R_{T_i}(t)$, is the probability that the i^{th} system has a lifespan larger than a duration t,

$$R_{T_i}(t) = \Pr\{T_i > t\} \tag{2.1}$$

The probability density function (pdf) of T_i is denoted by $f_{T_i}(t)$ and may be defined as

$$f_{T_i}(t) = -\frac{d}{dt} R_{T_i}(t).$$

Next, we introduce the hazard function. The probability that a failure occurs between t and Δt given that no failure occurs before time t is given by

$$\Pr\{T_i \leq t + \Delta t | T_i > t\} = \frac{\Pr\{t < T_i < t + \Delta t\}}{\Pr\{T_i > t\}}.$$

The failure rate is given by the dividing this equation by the length of the time interval, Δt:

$$\frac{\Pr\{t < T < t + \Delta t\}}{\Delta t} \frac{1}{\Pr\{T > t\}} = \frac{R_T(t) - R(t + \Delta t)}{R_T(t)}.$$

The hazard function $h_{T_i}(t)$ for T_i is the instantaneous failure rate at time t, which is given by

$$\begin{aligned} h_{T_i}(t) &= \lim_{\Delta t \to 0} \frac{\Pr\{t < T_i < t + \Delta t\}}{\Delta t} \frac{1}{\Pr\{T_i > t\}} \\ &= \frac{f_{T_i}(t)}{R_{T_i}(t)}. \end{aligned} \tag{2.2}$$

\end{definition}

The lifetime of the j^{th} component is assumed to follow a parametric distribution indexed by a parameter vector $\boldsymbol{\theta_j}$. The parameter vector of the overall system is defined as

$$\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_m}).$$

When a random variable T is parameterized by a particular $\boldsymbol{\theta}$, we denote the reliability function by $R_T(t; \boldsymbol{\theta})$, and the same for other distribution functions. If it is clear from the context which random variable a distribution function is for, we drop the subscripts, e.g., $R(t)$ instead of $R_T(t)$. As a special case, we denote the pdf of the j^{th} component by $f_j(t; \boldsymbol{\theta_j})$ and its reliability function by $R_j(t; \boldsymbol{\theta_j})$.

Two random variables X and Y have a joint pdf $f_{X,Y}(x, y)$. Given the joint pdf $f(x, y)$, the marginal pdf of X is given by

$$f_X(x) = \int_{\mathcal{Y}} f_{X,Y}(x, y) dy,$$

where $\mathcal{Y}$ is the support of Y. (If Y is discrete, replace the integration with a summation over $\mathcal{Y}$.)

The conditional pdf of Y given $X = x$, $f_{Y|X}(y|x)$, is defined as

$$f_{X|Y}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

We may generalize all of the above to more than two random variables, e.g., the joint pdf of $X_1, \ldots, X_m$ is denoted by $f(x_1, \ldots, x_m)$.

Next, we dive deeper into these concepts and provide mathematical derivations for the reliability function, pdf, and hazard function of the series system. We begin with the reliability function of the series system, as given by the following theorem.

Theorem 1. *The series system has a reliability function given by*

$$R(t; \boldsymbol{\theta}) = \prod_{j=1}^{m} R_j(t; \boldsymbol{\theta_j}). \tag{2.3}$$

Proof. The reliability function is defined as

$$R(t; \boldsymbol{\theta}) = \Pr\{T_i > t\}$$

which may be rewritten as

$$R(t; \boldsymbol{\theta}) = \Pr\{\min\{T_{i1}, \ldots, T_{im}\} > t\}.$$

For the minimum to be larger than t, every component must be larger than t,

$$R(t;\boldsymbol{\theta}) = \Pr\{T_{i1} > t, \ldots, T_{im} > t\}.$$

Since the component lifetimes are independent, by the product rule the above may be rewritten as

$$R(t;\boldsymbol{\theta}) = \Pr\{T_{i1} > t\} \times \cdots \times \Pr\{T_{im} > t\}.$$

By definition, $R_j(t;\boldsymbol{\theta}) = \Pr\{T_{ij} > t\}$. Performing this substitution obtains the result

$$R(t;\boldsymbol{\theta}) = \prod_{j=1}^{m} R_j(t;\boldsymbol{\theta_j}).$$

□

Theorem 1 shows that the system's overall reliability is the product of the reliabilities of its individual components. This property is inherent to series systems and will be used in the subsequent derivations.

Next, we turn our attention to the pdf of the system lifetime, described in the following theorem.

Theorem 2. *The series system has a pdf given by*

$$f(t;\boldsymbol{\theta}) = \sum_{j=1}^{m} f_j(t;\boldsymbol{\theta_j}) \prod_{\substack{k=1 \\ k\neq j}}^{m} R_k(t;\boldsymbol{\theta_j}). \tag{2.4}$$

Proof. By definition, the pdf may be written as

$$f(t;\boldsymbol{\theta}) = -\frac{d}{dt} \prod_{j=1}^{m} R_j(t;\boldsymbol{\theta_j}).$$

By the product rule, this may be rewritten as

$$\begin{aligned} f(t;\boldsymbol{\theta}) &= -\frac{d}{dt} R_1(t;\boldsymbol{\theta_1}) \prod_{j=2}^{m} R_j(t;\boldsymbol{\theta_j}) - R_1(t;\boldsymbol{\theta_1}) \frac{d}{dt} \prod_{j=2}^{m} R_j(t;\boldsymbol{\theta_j}) \\ &= f_1(t;\boldsymbol{\theta}) \prod_{j=2}^{m} R_j(t;\boldsymbol{\theta_j}) - R_1(t;\boldsymbol{\theta_1}) \frac{d}{dt} \prod_{j=2}^{m} R_j(t;\boldsymbol{\theta_j}). \end{aligned}$$

Recursively applying the product rule $m-1$ times results in

$$f(t;\boldsymbol{\theta}) = \sum_{j=1}^{m-1} f_j(t;\boldsymbol{\theta_j}) \prod_{\substack{k=1 \\ k\neq j}}^{m} R_k(t;\boldsymbol{\theta_k}) - \prod_{j=1}^{m-1} R_j(t;\boldsymbol{\theta_j}) \frac{d}{dt} R_m(t;\boldsymbol{\theta_m}),$$

which simplifies to

$$f(t;\boldsymbol{\theta}) = \sum_{j=1}^{m} f_j(t;\boldsymbol{\theta_j}) \prod_{\substack{k=1 \\ k\neq j}}^{m} R_k(t;\boldsymbol{\theta_k}).$$

□

Theorem 2 shows the pdf of the system lifetime as a function of the pdfs and reliabilities of its components.

We continue with the hazard function of the system lifetime, defined in the next theorem.

Theorem 3. *The series system has a hazard function given by*

$$h(t;\boldsymbol{\theta}) = \sum_{j=1}^{m} h_j(t;\boldsymbol{\theta_j}). \tag{2.5}$$

Proof. By Equation (2.2), the i^{th} series system lifetime has a hazard function defined as

$$h(t;\boldsymbol{\theta}) = \frac{f_{T_i}(t;\boldsymbol{\theta})}{R_{T_i}(t;\boldsymbol{\theta})}.$$

Plugging in expressions for these functions results in

$$h(t;\boldsymbol{\theta}) = \frac{\sum_{j=1}^{m} f_j(t;\boldsymbol{\theta_j}) \prod_{\substack{k=1 \\ k\neq j}}^{m} R_k(t;\boldsymbol{\theta_k})}{\prod_{j=1}^{m} R_j(t;\boldsymbol{\theta_j})},$$

which can be simplified to

$$h_{T_i}(t;\boldsymbol{\theta}) = \sum_{j=1}^{m} \frac{f_j(t;\boldsymbol{\theta_j})}{R_j(t;\boldsymbol{\theta_j})} = \sum_{j=1}^{m} h_j(t;\boldsymbol{\theta_j}).$$

□

Theorem 3 reveals that the system's hazard function is the sum of the hazard functions of its components. By definition, the hazard function is the ratio of the pdf to the reliability function,

$$h(t;\boldsymbol{\theta}) = \frac{f(t;\boldsymbol{\theta})}{R(t;\boldsymbol{\theta})},$$

and we can rearrange this to get

$$\begin{aligned} f(t;\boldsymbol{\theta}) &= h(t;\boldsymbol{\theta})R(t;\boldsymbol{\theta}) \\ &= \left\{\sum_{j=1}^{m} h_j(t;\boldsymbol{\theta_j})\right\}\left\{\prod_{j=1}^{m} R_j(t;\boldsymbol{\theta_j})\right\}, \end{aligned} \tag{2.6}$$

which we sometimes find to be a more convenient form than Equation (2.4).

In this section, we derived the mathematical forms for the system's reliability function, pdf, and hazard function. Next, we build upon these concepts to derive distributions related to the component cause of failure.

2.1 Component Cause of Failure

Whenever a series system fails, precisely one of the components is the cause. We model the component cause of the series system failure as a random variable.

Definition 1. *The component cause of failure of a series system is denoted by the random variable K_i whose support is given by $\{1,\ldots,m\}$. For example, $K_i = j$ indicates that the component indexed by j failed first, i.e.,*

$$T_{ij} < T_{ij'}$$

for every j' in the support of K_i except for j. Since we have series systems, K_i is unique.

The system lifetime and the component cause of failure has a joint distribution given by the following theorem.

Theorem 4. *The joint pdf of the component cause of failure K_i and series system lifetime T_i is given by*

$$f_{K_i,T_i}(j,t;\boldsymbol{\theta}) = h_j(t;\boldsymbol{\theta_j})R_{T_i}(t;\boldsymbol{\theta}), \tag{2.7}$$

where $h_j(t;\boldsymbol{\theta_j})$ is the hazard function of the j^{th} component and $R_{T_i}(t;\boldsymbol{\theta})$ is the reliability function of the series system.

Proof. Consider a series system with 3 components. By the assumption that component lifetimes are mutually independent, the joint pdf of T_{i1}, T_{i2}, T_{i3} is given by

$$f(t_1, t_2, t_3; \boldsymbol{\theta}) = \prod_{j=1}^{3} f_j(t; \boldsymbol{\theta_j}).$$

The first component is the cause of failure at time t if $K_i = 1$ and $T_i = t$, which may be rephrased as the likelihood that $T_{i1} = t$, $T_{i2} > t$, and $T_{i3} > t$. Thus,

$$\begin{aligned} f_{K_i,T_i}(j; \boldsymbol{\theta}) &= \int_t^\infty \int_t^\infty f_1(t; \boldsymbol{\theta_1}) f_2(t_2; \boldsymbol{\theta_2}) f_3(t_3; \boldsymbol{\theta_3}) dt_3 dt_2 \\ &= \int_t^\infty f_1(t; \boldsymbol{\theta_1}) f_2(t_2; \boldsymbol{\theta_2}) R_3(t; \boldsymbol{\theta_3}) dt_2 \\ &= f_1(t; \boldsymbol{\theta_1}) R_2(t; \boldsymbol{\theta_2}) R_3(t_1; \boldsymbol{\theta_3}). \end{aligned}$$

Since $h_1(t; \boldsymbol{\theta_1}) = f_1(t; \boldsymbol{\theta_1}) / R_1(t; \boldsymbol{\theta_1})$,

$$f_1(t; \boldsymbol{\theta_1}) = h_1(t; \boldsymbol{\theta_1}) R_1(t; \boldsymbol{\theta_1}).$$

Making this substitution into the above expression for $f_{K_i,T_i}(j, t; \boldsymbol{\theta})$ yields

$$\begin{aligned} f_{K_i,T_i}(j, t; \boldsymbol{\theta}) &= h_1(t; \boldsymbol{\theta_1}) \prod_{l=1}^{m} R_l(t; \boldsymbol{\theta_l}) \\ &= h_1(t; \boldsymbol{\theta_1}) R(t; \boldsymbol{\theta}). \end{aligned}$$

Generalizing from this completes the proof. □

3 Likelihood Model for Masked Data

The object of interest is the (unknown) parameter value $\boldsymbol{\theta}$. To estimate this $\boldsymbol{\theta}$, we need *data*. In our case, we call it *masked data* because we do not necessarily observe the event of interest, say a system failure, directly. We consider two types of masking: masking the system failure lifetime and masking the component cause of failure.

We generally encounter three types of system failure lifetime masking:

1. A system failure is observed at a particular point in time.
2. A system failure is observed to occur within a particular interval of time.
3. A system failure is not observed, but we know that the system survived at least until a particular point in time. This is known as *right-censoring* and can occur if, for instance, an experiment is terminated while the system is still functioning.

We generally encounter two types of component cause of failure masking:

1. The component cause of failure is observed.
2. The component cause of failure is not observed, but we know that the failed component is in some set of components. This is known as *masking* the component cause of failure.

Thus, the component cause of failure masking will take the form of candidate sets. A candidate set consists of some subset of component labels that plausibly contains the label of the failed component. The sample space of candidate sets are all subsets of $\{1, \ldots, m\}$, thus there are 2^m possible outcomes in the sample space.

In this paper, we limit our focus to observing *right censored* lifetimes and exact lifetimes but with masked component cause of failures. We consider a sample of n i.i.d. series systems, each of which is put into operation at some time and and observed until either it fails or is right-censored. We denote the right-censoring time

of the i^{th} system by τ_i. We do not directly observe the system lifetime, T_i, but rather, we observe the right-censored lifetime, S_i, which is given by

$$S_i = \min\{\tau_i, T_i\}, \tag{3.1}$$

We also observe a right-censoring indicator, δ_i, which is given by

$$\delta_i = 1_{T_i < \tau_i} \tag{3.2}$$

where $1_{\text{condition}}$ is an indicator function that outputs 1 if *condition* is true and 0 otherwise. Here, $\delta_i = 1$ indicates the event of interest, a system failure, was observed.

If a system failure lifetime is observed, then we also observe a candidate set that contains the component cause of failure. We denote the candidate set for the i^{th} system by $\mathcal{C}_i$, which is a subset of $\{1, \ldots, m\}$. Since the data generating process for candidate sets may be subject to chance variations, it as a random set.

Consider we have an independent and identically distributed (i.i.d.) random sample of masked data, $D = \{D_1, \ldots, D_n\}$, where each D_i contanis the following:

- S_i, the system lifetime of the i^{th} system.
- δ_i, the right-censoring indicator of the i^{th} system.
- $\mathcal{C}_i$, the set of candidate component causes of failure for the i^{th} system.

The masked data generation process is illustrated by Figure 1.

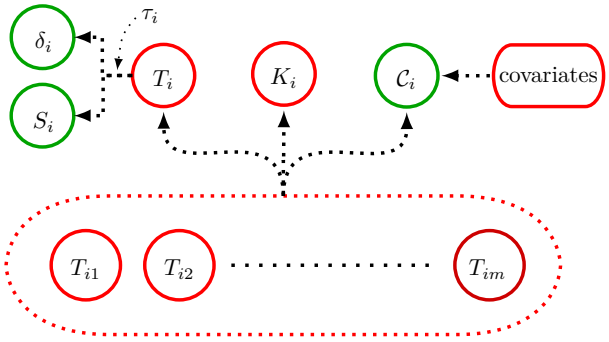

Figure 1: This figure showcases a dependency graph of the generative model for $D_i = (S_i, \delta_i, \mathcal{C}_i)$. The elements in green are observed in the sample, while the elements in red are unobserved (latent). We see that $\mathcal{C}_i$ is related to both the unobserved component lifetimes $T_{i1}, \ldots, T_{im}$ and other unknown and unobserved covariates, like ambient temperature or the particular diagnostician who generated the candidate set. These two complications for $\mathcal{C}_i$ are why seek a way to construct a reduced likelihood function in later sections that is not a function of the distribution of $\mathcal{C}_i$.

An example of masked data D for exact, right-censored system failure times with candidate sets that mask the component cause of failure can be seen in Table 1 for a series system with $m = 3$ components.

Table 1: Right-censored lifetime data with masked component cause of failure.

System	Right-censoring time (S_i)	Right censoring indicator (δ_i)	Candidate set ($\mathcal{C}_i$)
1	4.3	1	$\{1, 2\}$
2	1.3	1	$\{2\}$
3	5.4	0	$\emptyset$
4	2.6	1	$\{2, 3\}$
5	3.7	1	$\{1, 2, 3\}$
6	10	0	$\emptyset$

In our model, we assume the data is governed by a pdf, which is determined by a specific parameter, represented as $\boldsymbol{\theta}$ within the parameter space $\boldsymbol{\Omega}$. The joint pdf of the data D can be represented as follows:

$$f(D;\boldsymbol{\theta}) = \prod_{i=1}^{n} f(s_i, \delta_i, c_i; \boldsymbol{\theta}),$$

where s_i is the observed system lifetime of the i^{th} system, δ_i is the observed right-censoring indicator of the i^{th} system, and c_i is the observed candidate set of the i^{th} system.

This joint pdf tells us how likely we are to observe the particular data, D, given the parameter $\boldsymbol{\theta}$. When we keep the data constant and allow the parameter $\boldsymbol{\theta}$ to vary, we obtain what is called the likelihood function L, defined as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} L_i(\boldsymbol{\theta})$$

where

$$L_i(\boldsymbol{\theta}) = f(s_i, \delta_i, c_i; \boldsymbol{\theta})$$

is the likelihood contribution of the i^{th} system. In other words, the likelihood function quantifies how likely different parameter values $\boldsymbol{\theta}$ are, given the observed data.

For each type of data, right-censored data and masked component cause of failure data, we will derive the *likelihood contribution* L_i, which refers to the part of the likelihood function that this particular piece of data contributes to.

We present the following theorem for the likelihood contribution model.

Theorem 5. *The likelihood contribution of the i-th system is given by*

$$L_i(\boldsymbol{\theta}) = \begin{cases} R_{T_i}(s_i;\boldsymbol{\theta}) & \text{if } \delta_i = 0 \\ \beta_i R_{T_i}(s_i;\boldsymbol{\theta}) \sum_{j\in c_i} h_j(s_i;\boldsymbol{\theta_j}) & \text{if } \delta_i = 1, \end{cases} \tag{3.3}$$

where $\delta_i = 0$ indicates the i^{th} system is right-censored at time s_i and $\delta_i = 1$ indicates the i^{th} system is observed to have failed at time s_i and the component cause of failure is masked by the candidate set is c_i.

In the follow subsections, we prove this result for each type of masked data, right-censored system lifetime data ($\delta_i = 0$) and masking of the component cause of failure ($\delta_i = 1$).

3.1 Masked Component Cause of Failure

Suppose a diagnostician is unable to identify the precise component cause of the failure, e.g., due to cost considerations he or she replaced multiple components at once, successfully repairing the system but failing to precisely identity the failed component. In this case, the cause of failure is said to be *masked*.

The unobserved component lifetimes may have many covariates, like ambient operating temperature, but the only covariate we observe in our masked data model are the system's lifetime and additional masked data in the form of a candidate set that is somehow correlated with the unobserved component lifetimes.

The key goal of our analysis is to estimate the parameters, $\boldsymbol{\theta}$, which maximize the likelihood of the observed data, and to estimate the precision and accuracy of this estimate using the Bootstrap method.

To achieve this, we first need to assess the joint distribution of the system's continuous lifetime, T_i, and the discrete candidate set, $\mathcal{C}_i$, which can be written as

$$f_{T_i,\mathcal{C}_i}(t_i, c_i; \boldsymbol{\theta}) = f_{T_i}(t_i;\boldsymbol{\theta}) \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i\},$$

where $f_{T_i}(t_i;\boldsymbol{\theta})$ is the pdf of T_i and $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i\}$ is the conditional pmf of $\mathcal{C}_i$ given $T_i = t_i$.

We assume the pdf $f_{T_i}(t_i;\boldsymbol{\theta})$ is known, but we do not have knowledge of $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i\}$, i.e., the data generating process for candidate sets is unknown.

However, it is critical that the masked data, $\mathcal{C}_i$, is correlated with the i^{th} system. This way, the conditional distribution of $\mathcal{C}_i$ given $T_i = t_i$ may provide information about $\boldsymbol{\theta}$, despite our Statistical interest being primarily in the series system rather than the candidate sets.

To make this problem tractable, we assume a set of conditions that make it unnecessary to estimate the generative processes for candidate sets. The most important way in which $\mathcal{C}_i$ is correlated with the i^{th} system is given by assuming the following condition.

Condition 1. *The candidate set $\mathcal{C}_i$ contains the index of the the failed component, i.e.,*

$$\Pr_{\boldsymbol{\theta}}\{K_i \in \mathcal{C}_i\} = 1$$

where K_i is the random variable for the failed component index of the i^{th} system.

Assuming Condition 1, $\mathcal{C}_i$ must contain the index of the failed component, but we can say little else about what other component indices may appear in $\mathcal{C}_i$.

In order to derive the joint distribution of $\mathcal{C}_i$ and T_i assuming Condition 1, we take the following approach. We notice that $\mathcal{C}_i$ and K_i are statistically dependent. We denote the conditional pmf of $\mathcal{C}_i$ given $T_i = t_i$ and $K_i = j$ as

$$\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j\}.$$

Even though K_i is not observable in our masked data model, we can still consider the joint distribution of T_i, K_i, and $\mathcal{C}_i$. By Theorem 4, the joint pdf of T_i and K_i is given by

$$f_{T_i,K_i}(t_i, j; \boldsymbol{\theta}) = h_j(t_i; \boldsymbol{\theta_j}) R_{T_i}(t_i; \boldsymbol{\theta}),$$

where $h_j(t_i; \boldsymbol{\theta_j})$ is the hazard function for the j^{th} component and $R_{T_i}(t_i; \boldsymbol{\theta})$ is the reliability function of the system. Thus, the joint pdf of T_i, K_i, and $\mathcal{C}_i$ may be written as

$$\begin{aligned} f_{T_i,K_i,\mathcal{C}_i}(t_i, j, c_i; \boldsymbol{\theta}) &= f_{T_i,K_i}(t_i, k; \boldsymbol{\theta}) \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j\} \\ &= h_j(t_i; \boldsymbol{\theta_j}) R_{T_i}(t_i; \boldsymbol{\theta}) \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j\}. \end{aligned} \tag{3.4}$$

We are going to need the joint pdf of T_i and $\mathcal{C}_i$, which may be obtained by summing over the support $\{1, \ldots, m\}$ of K_i in Equation (3.4),

$$f_{T_i,\mathcal{C}_i}(t_i, c_i; \boldsymbol{\theta}) = R_{T_i}(t_i; \boldsymbol{\theta}) \sum_{j=1}^{m} \left\{ h_j(t_i; \boldsymbol{\theta_j}) \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j\} \right\}.$$

By Condition 1, $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j\} = 0$ when $K_i = j$ and $j \notin c_i$, and so we may rewrite the joint pdf of T_i and $\mathcal{C}_i$ as

$$f_{T_i,\mathcal{C}_i}(t_i, c_i; \boldsymbol{\theta}) = R_{T_i}(t_i; \boldsymbol{\theta}) \sum_{j \in c_i} \left\{ h_j(t_i; \boldsymbol{\theta_j}) \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j\} \right\}. \tag{3.5}$$

When we try to find an MLE of $\boldsymbol{\theta}$ (see Section 4), we solve the simultaneous equations of the MLE and choose a solution $\hat{\boldsymbol{\theta}}$ that is a maximum for the likelihood function. When we do this, we find that $\hat{\boldsymbol{\theta}}$ depends on the unknown conditional pmf $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j\}$. So, we are motivated to seek out more conditions (that approximately hold in realistic situations) whose MLEs are independent of the pmf $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j\}$.

Condition 2. *Any of the components in the candidate set has an equal probability of being the cause of failure. That is, for a fixed $j \in c_i$,*

$$\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j'\} = \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j\}$$

for all $j' \in c_i$.

According to (Guess et al., 1991), in many industrial problems, masking generally occurred due to time constraints and the expense of failure analysis. In this setting, Condition 2 generally holds.

Assuming Conditions 1 and 2, $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j\}$ may be factored out of the summation in Equation (3.5), and thus the joint pdf of T_i and $\mathcal{C}_i$ may be rewritten as

$$f_{T_i,\mathcal{C}_i}(t_i, c_i; \boldsymbol{\theta}) = \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j'\} R_{T_i}(t_i; \boldsymbol{\theta}) \sum_{j \in c_i} h_j(t_i; \boldsymbol{\theta_j})$$

where $j' \in c_i$.

If $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j'\}$ is a function of $\boldsymbol{\theta}$, the MLEs are still dependent on the unknown $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j'\}$. This is a more tractable problem, but we are primarily interested in the situation where we do not need to know (nor estimate) $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j'\}$ to find an MLE of $\boldsymbol{\theta}$. The last condition we assume achieves this result.

Condition 3. *The masking probabilities conditioned on failure time T_i and component cause of failure K_i are not functions of $\boldsymbol{\theta}$. In this case, the conditional probability of $\mathcal{C}_i$ given $T_i = t_i$ and $K_i = j'$ is denoted by*

$$\beta_i = \Pr\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j'\}$$

where β_i is not a function of $\boldsymbol{\theta}$.

When Conditions 1, 2, and 3 are satisfied, the joint pdf of T_i and $\mathcal{C}_\rangle$ is given by

$$f_{T_i,\mathcal{C}_i}(t_i, c_i; \boldsymbol{\theta}) = \beta_i R_{T_i}(t_i; \boldsymbol{\theta}) \sum_{j \in c_i} h_j(t_i; \boldsymbol{\theta_j}).$$

When we fix the sample and allow $\boldsymbol{\theta}$ to vary, we obtain the contribution to the likelihood L from the i^{th} observation when the system lifetime is exactly known (i.e., $\delta_i = 1$) but the component cause of failure is masked by a candidate set c_i:

$$L_i(\boldsymbol{\theta}) = R_{T_i}(t_i; \boldsymbol{\theta}) \sum_{j \in c_i} h_j(t_i; \boldsymbol{\theta_j}). \tag{3.6}$$

To summarize this result, assuming Conditions 1, 2, and 3, if we observe an exact system failure time for the i-th system ($\delta_i = 1$), but the component that failed is masked by a candidate set c_i, then its likelihood contribution is given by Equation (3.6).

3.2 Right-Censored Data

As described in Section **??**, we observe realizations of $(S_i, \delta_i, \mathcal{C}_i)$ where $S_i = \min\{T_i, \tau_i\}$ is the right-censored system lifetime, $\delta_i = 1_{\{T_i < \tau_i\}}$ is the right-censoring indicator, and $\mathcal{C}_i$ is the candidate set.

In the previous section, we discussed the likelihood contribution from an observation of a masked component cause of failure, i.e., $\delta_i = 1$. We now derive the likelihood contribution of a *right-censored* observation ($\delta_i = 0$) in our masked data model.

Theorem 6. *The likelihood contribution of a right-censored observation ($\delta_i = 0$) is given by*

$$L_i(\boldsymbol{\theta}) = R_{T_i}(s_i; \boldsymbol{\theta}). \tag{3.7}$$

Proof. When right-censoring occurs, then $S_i = \tau_i$, and we only know that $T_i > \tau_i$, and so we integrate over all possible values that it may have obtained,

$$L_i(\boldsymbol{\theta}) = \Pr_{\boldsymbol{\theta}}\{T_i > s_i\}.$$

By definition, this is just the survival or reliability function of the series system evaluated at s_i,

$$L_i(\boldsymbol{\theta}) = R_{T_i}(s_i; \boldsymbol{\theta}).$$

□

When we combine the two likelihood contributions, we obtain the likelihood contribution for the i^{th} system shown in Theorem 5,

$$L_i(\boldsymbol{\theta}) = \begin{cases} R_{T_i}(s_i; \boldsymbol{\theta}) & \text{if } \delta_i = 0 \\ \beta_i R_{T_i}(s_i; \boldsymbol{\theta}) \sum_{j \in c_i} h_j(s_i; \boldsymbol{\theta_j}) & \text{if } \delta_i = 1. \end{cases}$$

We use this result in the next section to derive the maximum likelihood estimator of $\boldsymbol{\theta}$.

4 Maximum Likelihood Estimation

In our analysis, we use maximum likelihood estimation (MLE) to estimate the series system parameter $\boldsymbol{\theta}$ from the masked data (Engelhardt, 1992; Casella and Berger, 2002). The MLE finds parameter values that maximize the likelihood of the observed data under the assumed model. The maximum likelihood estimate, $\hat{\boldsymbol{\theta}}$, is the solution of:

$$L(\hat{\boldsymbol{\theta}}) = \max_{\boldsymbol{\theta} \in \boldsymbol{\Omega}} L(\boldsymbol{\theta}), \tag{4.1}$$

where $L(\boldsymbol{\theta})$ is the likelihood function of the observed data. For computational and analytical simplicity, we work with the log-likelihood function, denoted as $\ell(\boldsymbol{\theta})$, instead of the likelihood function (Casella and Berger, 2002).

Theorem 7. *The log-likelihood function, $\ell(\boldsymbol{\theta})$, for our masked data model is the sum of the log-likelihoods for each observation,*

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \tag{4.2}$$

where $\ell_i(\boldsymbol{\theta})$ is the log-likelihood contribution for the i^{th} observation:

$$\ell_i(\boldsymbol{\theta}) = \sum_{j=1}^{m} \log R_j(s_i; \boldsymbol{\theta_j}) + \delta_i \log\big(\sum_{j \in c_i} h_j(s_i; \boldsymbol{\theta_j})\big). \tag{4.3}$$

Proof. The log-likelihood function is the logarithm of the likelihood function,

$$\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = \log \prod_{i=1}^{n} L_i(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log L_i(\boldsymbol{\theta}).$$

Substituting $L_i(\boldsymbol{\theta})$ from Equation (3.3) and separating the two cases of δ_i, we get

Case 1: If the i-th system is right-censored ($\delta_i = 0$),

$$\ell_i(\boldsymbol{\theta}) = \log R_{T_i}(s_i; \boldsymbol{\theta}) = \sum_{j=1}^{m} \log R_j(s_i; \boldsymbol{\theta_j}).$$

Case 2: If the i-th system's component cause of failure is masked but the failure time is known ($\delta_i = 1$),

$$\begin{aligned} \ell_i(\boldsymbol{\theta}) &= \log R_{T_i}(s_i; \boldsymbol{\theta}) + \log \beta_i + \log\big(\sum_{j \in c_i} h_j(s_i; \boldsymbol{\theta_j})\big) \\ &= \sum_{j=1}^{m} \log R_j(s_i; \boldsymbol{\theta_j}) + \log\big(\sum_{j \in c_i} h_j(s_i; \boldsymbol{\theta_j})\big)\Big). \end{aligned}$$

By Condition 3, we may ignore the term $\log \beta_i$ in the MLE since it does not depend on $\boldsymbol{\theta}$. This gives us the result in Theorem 7. □

The MLE, $\hat{\boldsymbol{\theta}}$, is often found by solving a system of equations derived from setting the derivative of the log-likelihood function to zero, i.e.,

$$\frac{\partial}{\partial \theta_j} \ell(\boldsymbol{\theta}) = 0, \tag{4.4}$$

for each component θ_j of the parameter $\boldsymbol{\theta}$ (Engelhardt, 1992). When there's no closed-form solution, we resort to numerical methods like the Newton-Raphson method.

Often, the MLE is found by setting the derivative of the log-likelihood function to zero, leading to a system of equations that is usually solved using numerical methods, such as the Newton-Raphson method [2, 3].

MLE has desirable asymptotic properties that underpin statistical inference, namely that it is asymptotically unbiased, unique, and normally distributed, with a variance given by the inverse of the Fisher Information Matrix (FIM) [2]. However, for smaller samples or complex models, these asymptotic properties may not yield accurate approximations. Hence, we propose to use the bootstrap method to offer an empirical approach for estimating the sampling distribution of the MLE.

5 Bootstrapping the Confidence Intervals of the MLE

We utilize the non-parametric bootstrap to approximate the sampling distribution of the MLE. In the non-parametric bootstrap, we resample from the observed data with replacement to generate a bootstrap sample. The MLE is then computed for the bootstrap sample. This process is repeated B times, giving us B bootstrap replicates of the MLE. The sampling distribution of the MLE is then approximated by the empirical distribution of the bootstrap replicates of the MLE.

Once we have a bootstrap distribution of the MLE, we can use it to construct confidence intervals for the parameters $\boldsymbol{\theta}$. The confidence interval for θ_j is given by

$$\left[\hat{\theta}_j - z_{\alpha/2}\hat{\sigma}_j, \hat{\theta}_j + z_{\alpha/2}\hat{\sigma}_j\right],$$

where $\hat{\theta}_j$ is the MLE of θ_j, $\hat{\sigma}_j$ is the

We will bootstrap the confidence interval using the percentile method and variance will be computed using the variance of the bootstrap replicates of the MLE.

In our simulation study, we will assess the performance of the bootstrapped confidence intervals by computing their widths and computing coverage probabilities, i.e., a 95% confidence interval should contain the true value 95% of the time.

Together, the width and the coverage probability, tell us how informative the confidence interval is. We say that a confidence interval has *good coverage* if its coverage probability is close to the nominal confidence level.

If the confidence intervals have good coverage, a small confidence interval width means we are more confident that the true value $\boldsymbol{\theta}$ is close to the MLE, and a large confidence interval width means we are less confident that the true value of $\boldsymbol{\theta}$ is close to the MLE. However, if the confidence interval has poor coverage, then the confidence interval width is not particularly informative. Thus, we see that the confidence interval is only useful if it has good coverage,

5.1 Issues with Resampling from the Observed Data

While the bootstrap method provides a robust and flexible tool for statistical estimation, its effectiveness can be influenced by several factors (Efron and Tibshirani, 1994).

Firstly, instances of non-convergence in our bootstrap samples were observed. Such cases can occur when the estimation method, like the MLE used in our analysis, fails to converge due to the specifics of the resampled data (Casella and Berger, 2002). This issue can potentially introduce bias or reduce the effective sample size of our bootstrap distribution.

Secondly, the bootstrap's accuracy can be compromised with small sample sizes, as the method relies on the law of large numbers to approximate the true sampling distribution. For small datasets, the bootstrap samples might not adequately represent the true variability in the data, leading to inaccurate results (Efron and Tibshirani, 1994).

Thirdly, our data involves right censoring and a Bernoulli candidate set model for component failure masking. These aspects can cause certain data points or trends to be underrepresented or not represented at all in our data, introducing bias in the bootstrap distribution (Klein and Moeschberger, 2005).

Despite these challenges, we found the bootstrap method useful in approximating our parameter distributions, taking care in interpreting the results. Further research could explore more sophisticated techniques that can handle these issues more effectively.

Parametric bootstrap

There is an alternative form of the bootstrap called the parametric bootstap, where the bootstrap samples are generated from a parametric distribution. However, this parametric bootstrap is not appropriate for our analysis because we do not assume a parametric form for the distribution of the candidate sets $\mathcal{C}_i$. In our simulation study, we use what we call the Bernoulli candidate set model (see Section **??**), which is a parametric model for the distribution of the candidate sets, but we want to assess the performance of the bootstrap method when the true data generating process for candidate sets is not known.

6 Series System with Weibull Components

In the real world, systems are quite complex:

1. They are not perfect series systems.
2. The components in a system are not independent.
3. The lifetimes of the components are not precisely modeled by any named probability distributions.
4. The components may depend on many other unobserved factors.

With these caveats in mind, we model the data as coming from a Weibull series system of $m = 5$ components, and other factors, like ambient temperature, are either negligible (on the distribution of component lifetimes) or are more or less constant.

The j^{th} component of the i^{th} has a lifetime distribution given by

$$T_{ij} \sim \text{WEI}(\boldsymbol{\theta_j})$$

where $\boldsymbol{\theta_j} = (k_j, \lambda_j)$ for $j = 1, \ldots, m$. Thus, $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_m})' = \big(k_1, \lambda_1, \ldots, k_m, \lambda_m\big)$. The random variable T_{ij} has a reliability function, pdf, and hazard function given respectively by

$$R_j(t; \lambda_j, k_j) = \exp\left\{-\left(\frac{t}{\lambda_j}\right)^{k_j}\right\}, \tag{6.1}$$

$$f_j(t; \lambda_j, k_j) = \frac{k_j}{\lambda_j}\left(\frac{t}{\lambda_j}\right)^{k_j - 1} \exp\left\{-\left(\frac{t}{\lambda_j}\right)^{k_j}\right\}, \tag{6.2}$$

$$h_j(t; \lambda_j, k_j) = \frac{k_j}{\lambda_j}\left(\frac{t}{\lambda_j}\right)^{k_j - 1} \tag{6.3}$$

where $t > 0$ is the lifetime, $\lambda_j > 0$ is the scale parameter and $k_j > 0$ is the shape parameter. The shape parameters $k_1, \ldots, k_m$ have the following interpretations:

$k_j < 1$ The hazard function decreases with respect to time. For instance, this may occur as a result of defective components being weeded out early. This is known as the *infant mortality* phase.

$k_j = 1$ The hazard function is constant with respect to time. This is an idealized case that is rarely observed in practice, but may be useful for modeling purposes.

$k_j > 1$ The hazard function increases with respect to time. For instance, this may occur as a result of components wearing out. This is known as the *aging* phase.

The lifetime of the series system composed of m Weibull components has a reliability function given by

$$R(t; \boldsymbol{\theta}) = \exp\left\{-\sum_{j=1}^{m}\left(\frac{t}{\lambda_j}\right)^{k_j}\right\}. \tag{6.4}$$

Proof. By Theorem 1,

$$R(t;\boldsymbol{\theta}) = \prod_{j=1}^{m} R_j(t;\lambda_j,k_j).$$

Plugging in the Weibull component reliability functions obtains the result

$$\begin{aligned} R(t;\boldsymbol{\theta}) &= \prod_{j=1}^{m} \exp\left\{-\left(\frac{t}{\lambda_j}\right)^{k_j}\right\} \\ &= \exp\left\{-\sum_{j=1}^{m}\left(\frac{t}{\lambda_j}\right)^{k_j}\right\}. \end{aligned}$$

□

The Weibull series system's hazard function is given by

$$h(t;\boldsymbol{\theta}) = \sum_{j=1}^{m} \frac{k_j}{\lambda_j}\left(\frac{t}{\lambda_j}\right)^{k_j-1}, \tag{6.5}$$

whose proof follows from Theorem 3.

The pdf of the series system is given by

$$f(t;\boldsymbol{\theta}) = \left\{\sum_{j=1}^{m} \frac{k_j}{\lambda_j}\left(\frac{t}{\lambda_j}\right)^{k_j-1}\right\} \exp\left\{-\sum_{j=1}^{m}\left(\frac{t}{\lambda_j}\right)^{k_j}\right\}. \tag{6.6}$$

Proof. By definition,

$$f(t;\boldsymbol{\theta}) = h(t;\boldsymbol{\theta})R(t;\boldsymbol{\theta}).$$

Plugging in the failure rate and reliability functions given respectively by Equations (6.4) and (6.5) completes the proof. □

6.1 Weibull Likelihood Model for Masked Data

In Section 3, we discussed two separate kinds of likelihood contributions, masked component cause of failure data (with exact system failure times) and right-censored data. The likelihood contribution of the i^{th} system is given by the following theorem.

Theorem 8. *Let δ_i be an indicator variable that is 1 if the i^{th} system fails and 0 (right-censored) otherwise. Then the likelihood contribution of the i^{th} system is given by*

$$L_i(\boldsymbol{\theta}) = \begin{cases} \exp\left\{-\sum_{j=1}^{m}\left(\frac{t_i}{\lambda_j}\right)^{k_j}\right\}\beta_i \sum_{j\in c_i} \frac{k_j}{\lambda_j}\left(\frac{t_i}{\lambda_j}\right)^{k_j-1} & \text{if } \delta_i = 1, \\ \exp\left\{-\sum_{j=1}^{m}\left(\frac{t_i}{\lambda_j}\right)^{k_j}\right\} & \text{if } \delta_i = 0. \end{cases} \tag{6.7}$$

Proof. By Theorem 5, the likelihood contribution of the i-th system is given by

$$L_i(\boldsymbol{\theta}) = \begin{cases} R_{T_i}(s_i;\boldsymbol{\theta}) & \text{if } \delta_i = 0 \\ \beta_i R_{T_i}(s_i;\boldsymbol{\theta}) \sum_{j\in c_i} h_j(s_i;\boldsymbol{\theta_j}) & \text{if } \delta_i = 1. \end{cases}$$

By Equation (6.4), the system reliability function R_{T_i} is given by

$$R_{T_i}(t_i;\boldsymbol{\theta}) = \exp\left\{-\sum_{j=1}^{m}\left(\frac{t_i}{\lambda_j}\right)^{k_j}\right\}.$$

and by Equation (6.3), the Weibull component hazard function h_j is given by

$$h_j(t_i;\boldsymbol{\theta_j}) = \frac{k_j}{\lambda_j}\left(\frac{t_i}{\lambda_j}\right)^{k_j-1}.$$

Plugging these into the likelihood contribution function obtains the result. □

Taking the log of the likelihood contribution function obtains the following result.

Corollary 1. *The log-likelihood contribution of the i-th system is given by*

$$\ell_i(\boldsymbol{\theta}) = -\sum_{j=1}^{m}\left(\frac{t_i}{\lambda_j}\right)^{k_j} + \delta_i \log\left(\sum_{j\in c_i}\frac{k_j}{\lambda_j}\left(\frac{t_i}{\lambda_j}\right)^{k_j-1}\right) \tag{6.8}$$

where we drop any terms that do not depend on $\boldsymbol{\theta}$ since they do not affect the MLE.

We find an MLE by solving (4.4), i.e., a point $\hat{\boldsymbol{\theta}} = (\hat{k}_1, \hat{\lambda}_1, \ldots, \hat{k}_m, \hat{\lambda}_m)$ satisfying $\nabla_\theta \ell(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, where $\nabla_{\boldsymbol{\theta}}$ is the gradient of the log-likelihood function (score) with respect to $\boldsymbol{\theta}$.

To solve this system of equations, we use the Newton-Raphson method, which requires the score and the Hessian of the log-likelihood function. We analytically derive the score since it is useful to have for the Newton-Raphson method, but we do not do the same for the Hessian of the log-likelihood for the following reasons:

1. The gradient is relatively easy to derive, and it is useful to have for computing gradients efficiently and accurately, which will be useful for numerically approximating the Hessian.
2. The Hessian is tedious and error prone to derive, and Newton-like methods often do not require the Hessian to be explicitly computed.

The following theorem derives the score function.

Theorem 9. *The score function of the log-likelihood contribution of the i-th Weibull series system is given by*

$$\nabla \ell_i(\boldsymbol{\theta}) = \left(\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial k_1}, \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \lambda_1}, \cdots, \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial k_m}, \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \lambda_m}\right)', \tag{6.9}$$

where

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial k_r} = -\left(\frac{t_i}{\lambda_r}\right)^{k_r}\log\left(\frac{t_i}{\lambda_r}\right) + \frac{\frac{1}{t_i}\left(\frac{t_i}{\lambda_r}\right)^{k_r}\left(1 + k_r \log\left(\frac{t_i}{\lambda_r}\right)\right)}{\sum_{j\in c_i}\frac{k_j}{\lambda_j}\left(\frac{t_i}{\lambda_j}\right)^{k_j-1}} 1_{\delta_i=1\wedge r\in c_i} \tag{6.10}$$

and

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \lambda_r} = \frac{k_r}{\lambda_r}\left(\frac{t_i}{\lambda_r}\right)^{k_r} - \frac{\left(\frac{k_r}{\lambda_r}\right)^2\left(\frac{t_i}{\lambda_r}\right)^{k_r-1}}{\sum_{j\in c_i}\frac{k_j}{\lambda_j}\left(\frac{t_i}{\lambda_j}\right)^{k_j-1}} 1_{\delta_i=1\wedge r\in c_i} \tag{6.11}$$

The result follows from taking the partial derivatives of the log-likelihood contribution of the i-th system given by Equation (6.7). It is a tedious calculation so the proof has been omitted, but the result has been verified by using a very precise numerical approximation of the gradient.

By the linearity of differentiation, the gradient of a sum of functions is the sum of their gradients, and so the score function conditioned on the entire sample is given by

$$\nabla \ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \nabla \ell_i(\boldsymbol{\theta}). \tag{6.12}$$

7 Simulation Study

We derived the likelihood model for masked data for the Weibull series system in Section 6. In this section, we describe the design of our simulation study, and we assess the performance of the bootstrap method for estimating the sampling distribution of the MLE for a Weibull series system with $m = 5$ components under our proposed likelihood model.

7.1 Realistic System Designs

A series system is only as reliable as its least reliable component. In order to make the simulation study representative of real-world scenarios, at least for systems designed to be reliable, we choose parameter values that are representative of real-world systems where there is no single component that is much less reliable than the others.

One way to define reliability is by the mean time to failure (MTTF), which is the expected value of the lifetime, which for the Weibull distribution is given by

$$\text{MTTF} = \lambda\,\Gamma(1 + 1/k),$$

where Γ is the gamma function.

We consider the data from (Guo et al., 2013), which includes a study of the reliability of a series system with three Weibull components with shape and scale parameters given by

$$\begin{aligned} k_1 &= 1.2576 \quad \lambda_1 = 994.3661 \\ k_2 &= 1.1635 \quad \lambda_2 = 908.9458 \\ k_3 &= 1.1308 \quad \lambda_3 = 840.1141. \end{aligned} \tag{7.1}$$

Our approach is to extend this system to a five component system by adding two more components with shape and scale parameters given by

$$\begin{aligned} k_4 &= 1.1802 \quad \lambda_4 = 940.1141 \\ k_5 &= 1.3311 \quad \lambda_5 = 836.1123. \end{aligned} \tag{7.2}$$

Table 2: Meean Time To Failure of Weibull Components and Series System

	MTTF
Component 1	924.8693
Component 2	862.1568
Component 3	803.5639
Component 4	888.2181
Component 5	768.6793
Series System	223.0336

As shown by Table 2, there are no components that are significantly less reliable than any of the others. Note that a series system in which, say, one of the components does have a significantly shorter MTTF would also pose significant challenges to estimating the parameters of the system from our masked failure data, since the failure time of the series system would be dominated by the failure time of the least reliable component. See Section 7.3 for further discussion.

Verification

To verify that our likelihood model is correct, we load the Table 2 data from (Guo et al., 2013) and fit the Weibull series model to the data to see if we can recover the MLE they reported. When we fit the Weibull series model to this data by maximizing the likelihood function, we obtain the following fit for the shape and scale parameters given respectively by

$$\hat{k}_1 = 1.2576, \hat{k}_2 = 1.1635, \hat{k}_3 = 1.1308,$$

and

$$\hat{\lambda}_1 = 994.3661, \hat{\lambda}_2 = 908.9458, \hat{\lambda}_3 = 840.1141,$$

which is in agreement with the MLE they reported. Satisfied that our likelihood model is correct, we proceed with the simulation study.

7.2 Data Generating Process

In this section, we describe the data generating process for our simulation study. It consists of three parts: the series system, the candidate set model, and the right-censoring model.

Weibull Series System Lifetime

We generate data from a Weibull series system with $m = 5$ components. As described in Section 6, the j^{th} component of the i^{th} system has a lifetime distribution given by

$$T_{ij} \sim \text{WEI}(k_j, \lambda_j)$$

and the lifetime of the series system composed of m Weibull components is defined as

$$T_i = \min\{T_{i1}, \ldots, T_{im}\}.$$

To generate a data set, we first generate the m component failure times, by efficiently sampling from their respective distributions, and we then set the failure time t_i of the system to the minimum of the component failure times.

Right-Censoring Model

We employ a very simple right-censoring model, where the right-censoring time τ is fixed at some known value, e.g., an experiment is run for a fixed amount of time τ, and all systems that have not failed by the end of the experiment are right-censored. The censoring time S_i of the i^{th} system is thus given by

$$S_i = \min\{T_i, \tau\}.$$

So, after we generate the system failure time T_i, we generate the censoring time S_i by taking the minimum of T_i and τ.

Masking Model for Component Cause of Failure

We must generate data that satisfies the masking conditions described in Section 3.1. There are many ways to satisfying the masking conditions. We choose the simplest method, which we call the *Bernoulli candidate set model*. In this model, each non-failed component is included in the candidate set with a fixed probability p, independently of all other components and independently of $\boldsymbol{\theta}$, and the failed component is always included in the candidate set.

7.3 Issues with Convergence to the MLE

Identifiability

When estimating the parameters, we must be careful to ensure that the parameters are identifiable such that the likelihood function is maximized at a unique point. If the likelihood function is not maximized at a unique point, then the MLE is not unique, and a lot of the theory we have developed so far breaks down, particularly when the likelihood surface is flat or has multiple local maxima (**?**)

One way in which this problem may arise is if the data is not informative enough. For example, if we have a series system and in the observed masked data component 1 is in the candidate set if and only if component 2 is in candidate set, then we do not have enough information to estimate the parameters of component 1 and component 2 separately.

Another way is if the the series system has a component that is the least reliable by a significant margin and is most likely the component cause of failure. In this case, our data is not informative enough to estimate the parameters of the other components. We constructed a quick experiment to demonstrate this phenomenon in Figure 2.

We simply made the MTTF of the first component much smaller than the others, and for even large samples there was not generally enough information to estimate the parameters of the other components. In this case, the MLE is not unique, and the likelihood surface is flat, as shown.

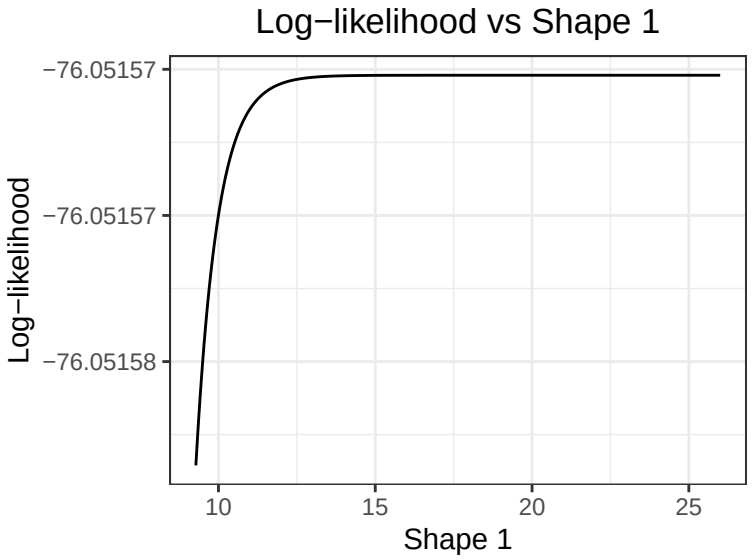

Figure 2: Log-likelihood profile of a flat surface (non-unique MLE)

We encountered this issue in our simulation study for small samples due to right-censored and masked component cause of failure data, despite parameterizing a series system with components that have similar MTTFs. Our decision was to simply exclude these data sets from our analysis, since they are not informative enough to estimate the parameters of the system.

Parameter rescaling

When the parameters under investigation span different orders of magnitude, parameter rescaling can significantly improve the performance and reliability of optimization algorithms. Parameter rescaling gives an optimizer a sense of the typical size of each parameter, enabling it to adjust its steps accordingly. This is crucial in scenarios like ours, where shape and scale parametes are a few orders of magnitude apart. Without rescaling, the optimization routine may struggle, taking numerous small steps for larger parameters and overshooting for smaller ones.

Speed of convergence was particularly important in our case, since in our simulation study, we employ the Bootstrap method to estimate the sampling distribution of the MLE, which requires us to estimate the MLE for many data sets. We found that parameter rescaling significantly improved the speed of convergence, which allowed us to run our simulation study in a tractable amount of time.

7.4 Assessing the Bootstrapped Confidence Intervals

Once we analyze the bias, variance, and MSE of the MLE, using a simulation study, we then assess the bootstrapped variance, bias, MSE, and confidence intervals. This is useful for understanding the performance of the bootstrap method for estimating the sampling distribution of the MLE and constructing confidence intervals.

We will assess the bootstrapped variance, bias, MSE, and confidence intervals by comparing them to the bias, variance, MSE, and confidence intervals obtained from the empirical sampling distribution. In particular, if the variance is too large or small, then the coverage probability of the confidence intervals will be too small or large, respectively.

Under a variety of scenarios, we will bootstrap a 95%-confidence interval for $\boldsymbol{\theta}$ using the percentile method, and we will evaluate its accuracy by computing the coverage probability.

We want the coverage probability to be close to the nominal level, 95%, because if the coverage probability is too low, then we will be too confident in the precision and accuracy of the MLE, and if the coverage probability is too high, then we will not be confident enough in the precision and accuracy of the MLE.

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.

7.5 Simulation Scenarios

We parameterize τ by quantiles of the series system, e.g., if $q = 0.8$, then $\tau(0.8)$ is the 80% quantile of the series system such that 80% of the sytems are expected to fail before time $\tau(0.8)$ and 20% of the series systems are expected to be right-censored.

We define a simulation scenario to be some comination of n, p, and q. For the sake of brevity, we will refer to a simulation scenario by its sample size n, masking probability p, and right-censoring probability q. We are interested in choosing a small number of scenarios that are representative of real-world scenarios and that are interesting to analyze. There are many ways to choose scenarios, but we choose the following three.

1. Ideal case with no masking of the component cause of failure ($p = 0$) and no right-censoring ($q = 1$, or equivalently, $\tau = \infty$).
2. Moderate masking of the component cause of failure ($p = 0.2$) and moderate right-censoring ($q = 0.9$).
3. Significant masking of the component cause of failure ($p = 0.4$) and significant right-censoring ($q = 0.75$).

For each of these scenarios, we will consider sample sizes n from small samples ($n = 30$) to relatively large samples ($n = 300$).

We will analyze the performance of the MLE under these various scenarios by computing the coverage probability and width of the 95%-confidence intervals.

Here is an outline of the simulation study analysis:

1. Choose a scenario (sample size n, masking probability p, and right-censoring probability q).
2. Generate R datasets from the Data Generating Process (DGP). The DGP should be compatible with the assumptions in our likelihood model. In our case, we use:
   - Right-censored series lifetimes with $m = 5$ Weibull components.
   - Bernoulli candidate set model for the masking of component cause of failure.
3. For each of these R datasets, calculate the Maximum Likelihood Estimator (MLE).
4. For each of these R datasets, perform bootstrap resampling B times to create a set of bootstrap samples.
5. Calculate the MLE for each of these bootstrap samples. This generates an empirical distribution of the MLE, which is used to construct a Confidence Interval (CI) for the MLE.
6. Repeat steps 4 and 5 for each of the R datasets.
7. For each dataset, determine whether the true parameter value falls within the computed CI. Aggregate this information across all R datasets to estimate the coverage probability of the CI.
8. Interpret the results and discuss the performance of the MLE estimator under various scenarios.

For how we generate a scenario, see Appendix A.

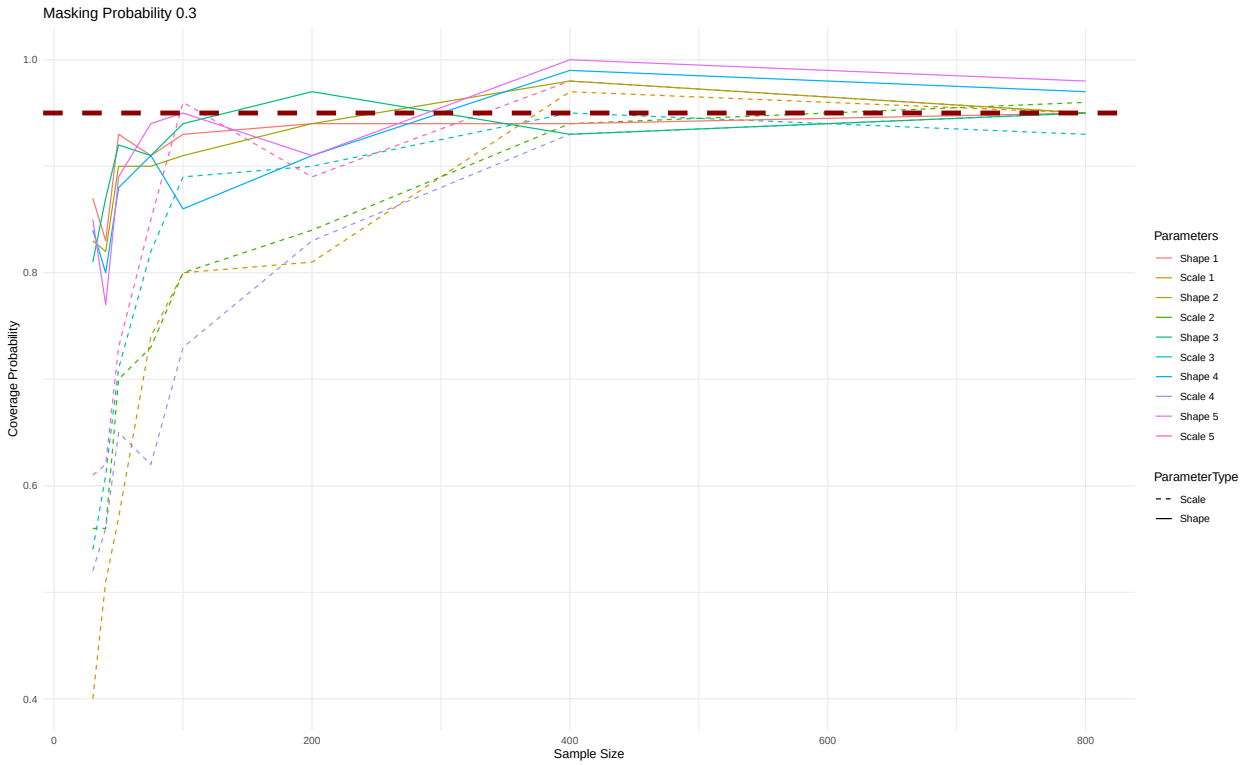

Figure 3: Coverage probability vs. sample size for masking probability 0.3

7.6 Coverage probability vs. sample size with a fixed masking probability and no right-censoring

We want to isolate the effect of the coverage probability as a function of the sample size n. We fix the masking probability to $p = 0.2$ and without right-censoring ($\tau = \infty$) and vary the sample size from $n = 30$ to $n = 800$. See Figure 3.

Here are some key observations:

1. It is immediately obvious that the scale parameters (dashed lines) have a much lower coverage probability than the shape parameters (solid lines), particularly for small sample sizes less than $n = 200$.

   In general, the scale parameters appear to be more difficult to estimate than the shape parameters.

2. As the sample size increases, the coverage probability for the shape parameters and scale parameters approaches the nominal level, 95%.

   This suggests that the sampling distribution of the MLE is converging in distribution to a multivariate normal distribution with mean $\boldsymbol{\theta}$ and variance-covariance given by the inverse of the FIM, consistent with the asymptotic theory.

7.7 Coverage probability vs. sample size and masking probability without right-censoring

We want to get a larger picture of how the coverage probability depends on the sample size n and masking probability p. We fix the right-censoring time to $\tau = \infty$ and vary the sample size from $n = 30$ to $n = 800$ and vary the masking probability from $p = 0$ (no masking) to $p = 0.4$ and then compute the coverage probability for each combination of sample size n and masking probability p.

The results of this analysis are summarized by Figure 4. Here are some key observations:

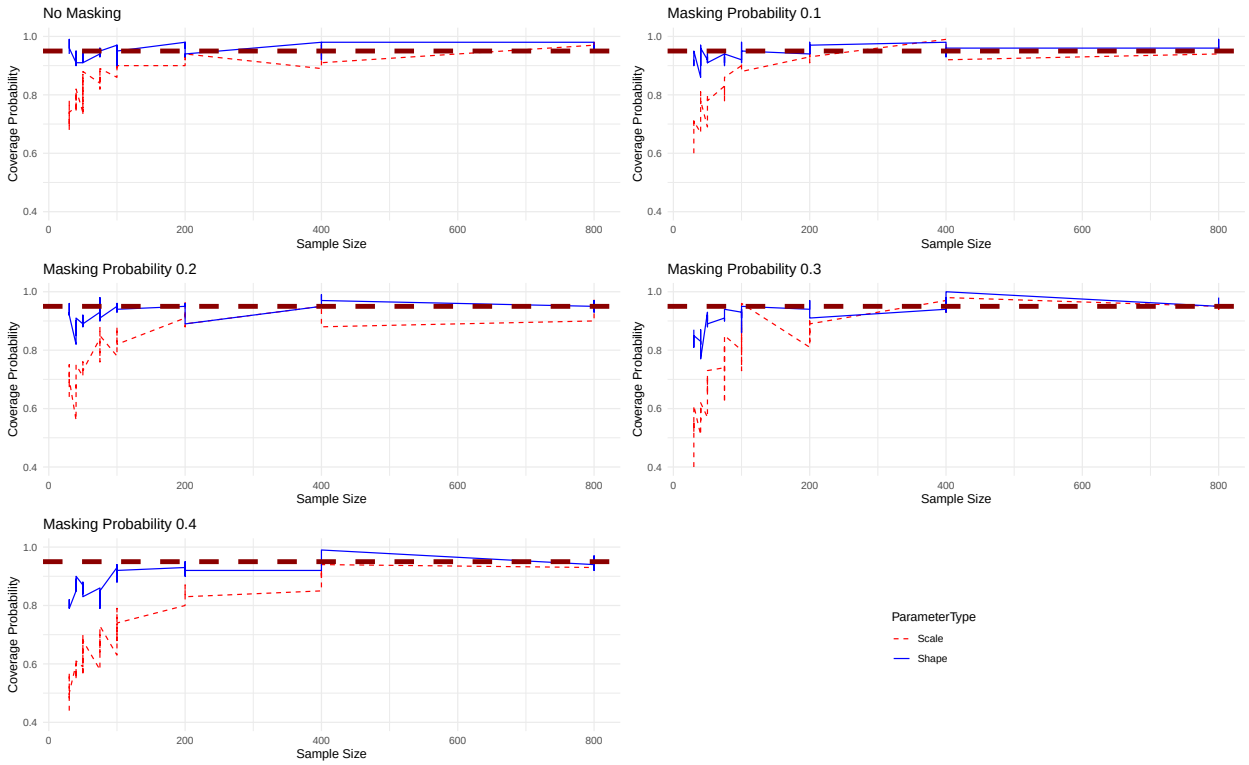

Figure 4: Coverage probability vs. sample size

1. For sample sizes $n \leq 100$, the coverage probability for the shape parameters is close to the nominal level, 95%, only for small masking probabilities. However, as the sample size increases, the coverage probability for the shape parameters quickly approaches the nominal level, 95%, for all masking probabilities reported here.

2. For the scale parameters, the coverage probability is too low for all sample sizes $n < 200$ for all masking probabilities reported here. For small sample sizes, the confience intervals particularly for the scale parameters, should probably be taken with a grain of salt.

In Section 5, we explore an alternative way to construct confidence intervals using the bootstrap method, which is generally a more accurate way to compute confidence intervals. Unlike the inverse of the observed FIM, it does not assume that the sampling distribution of the MLE is asymptotically normal, and so it is more robust to violations of this assumption.

8 Future Work

9 Conclusion

We have developed a likelihood model for series systems with latent components and right-censoring. We have provided evidence that, as long as certain regularity conditions are met, the MLE is asymptotically unbiased and consistent.

References

Please see below for a full list of references.

10 Appendix

10.1 Data

Simulation Code

```
############################################################
# Simulation data generating process for specified scenario #
# (n, p, q), where:                                        #
#    - n is a vector of sample sizes                       #
#     - n is a vector of sample sizes                      #
#     - p is a vector of masking probabilities             #
#     - q is a vector of right-censoring quantiles of the  #
#       Weibull series distribution.                       #
############################################################

# here is the R libary we developed for this project
library(wei.series.md.c1.c2.c3)

# for parallel processing
library(parallel)

# you can set a seed for reproducibility of the experimental run
# however, if you use parallel processing, this simple approach will not work.
# set.seed(1234)

################################################
# Here is an example of how to run a scenario #
################################################

# set the simulation name to be used in the file names
sim.name <- "sim-2"
# set the sample sizes
ns <- c(30, 40, 50, 75, 100, 200, 400, 800)
# set the masking probabilities
ps <- seq(0, 0,1, 0.2, 0.3, 0.4)
# set the right-censoring quantiles
qs <- c(0.5, 0.6, 0.7, 0.8, 0.9, 0.95)
# set the number of replicates
R <- 100
# set the number of CPU cores to use
ncores <- 4

# true parameter values
theta <- c(shape1 = 1.2576, scale1 = 994.3661,
           shape2 = 1.1635, scale2 = 908.9458,
           shape3 = 1.1308, scale3 = 840.1141,
           shape4 = 1.1802, scale4 = 940.1141,
           shape5 = 1.3311, scale5 = 836.1123)

shapes <- theta[seq(1, length(theta), 2)]
scales <- theta[seq(2, length(theta), 2)]

# helps the MLE optimization routine converge more quickly and reliably
```

```r
# by scaling the parameters to be of similar magnitude
parscale <- c(1, 1000, 1, 1000, 1, 1000, 1, 1000, 1, 1000)

sim.run <- function(sim.name, n, p, q, R = 1000) {
    mles <- list()
    problems <- list()

    tau <- wei.series.md.c1.c2.c3::qwei_series(
        p = q, scales = scales, shapes = shapes)

    cat("n =", n, ", p =", p, ", q = ", q, ", tau = ", tau, "\n")

    for (r in 1:R) {
        result <- tryCatch({

            df <- wei.series.md.c1.c2.c3::generate_guo_weibull_table_2_data(
                shapes = shapes,
                scales = scales,
                n = n,
                p = p,
                tau = tau)

            sol <- wei.series.md.c1.c2.c3::mle_nelder_wei_series_md_c1_c2_c3(
                df = df,
                theta0 = theta,
                reltol = 1e-7,
                parscale = parscale,
                maxit = 2000L)
            mles <- append(mles, list(sol))

            if (r %% 10 == 0) {
                cat("r = ", r, ": ", sol$par, "\n")
            }

        }, error = function(e) {
            cat("Error at iteration", r, ":")
            print(e)
            problems <- append(problems, list(list(
                error = e, n = n, p = p, q = q, tau = tau, df = df)))
        })
    }

    if (length(mles) != 0) {
        saveRDS(list(n = n, p = p, q = q, tau = tau, mles = mles),
            file = paste0("./results/", sim.name, "/results_", n, "_", p, "_", q, ".rds"))
    }

    if (length(problems) != 0) {
        saveRDS(list(n = n, p = p, q = q, tau = tau, problems = problems),
            file = paste0("./problems/", sim.name, "/problems_", n, "_", p, "_", q, ".rds"))
    }
}
```

```r
params <- expand.grid(n = ns, p = ps, q = qs)
result <- mclapply(
    1:nrow(params),
    function(i) sim.run(sim.name, params$n[i], params$p[i], params$q[i], R),
    mc.cores = ncores)
```

Appendix B: Simulation of scenarios using the Bootstrap method

```r
################################################################
# in this scenario, we want to see how we can use the bootstrap
# method to estimate the confidence intervals more precisely (better calibration
# of confidence intervals) for small sample sizes.
# we'll use it to construct a 95% confidence interval for the estimator. we'll
# compare this result to the asymptotic theory confidence interval.
# finally, we'll generate CIs by each method, asymptotic (inverse FIM) and
# bootstrap (cov), and compare the coverage probabilities.
################################################################

library(boot)
library(parallel)
library(wei.series.md.c1.c2.c3)

theta <- c(shape1 = 1.2576, scale1 = 994.3661,
           shape2 = 1.1635, scale2 = 908.9458,
           shape3 = 1.1308, scale3 = 840.1141,
           shape4 = 1.1802, scale4 = 940.1141,
           shape5 = 1.3311, scale5 = 836.1123)

shapes <- theta[seq(1, length(theta), 2)]
scales <- theta[seq(2, length(theta), 2)]

# number of CPU cores to use in bootstrap for parallel processing
ncores <- 4

# helps the MLE optimization routine converge more quickly and reliably
parscale <- c(1, 1000, 1, 1000, 1, 1000, 1, 1000, 1, 1000)

#set.seed(134849131)

# sample sizes
ns <- c(30, 50, 100, 200, 400)
# masking probabilities, no masking and 21.5% masking
ps <- c(0, 0.215)
# quantiles of weibull series distribution, no right-censoring and 25% right-censoring
qs <- c(1, 0.75)

sim.name <- "sim-1-boot"

sim.boot.run <- function(n, p, q, R = 1000) {

    problems <- list()
```

```
    tau <- wei.series.md.c1.c2.c3::qwei_series(
        p = q, scales = scales, shapes = shapes)

    cat("n =", n, ", p =", p, ", q = ", q, ", tau = ", tau, "\n")

    result <- tryCatch({
        df <- wei.series.md.c1.c2.c3::generate_guo_weibull_table_2_data(
            shapes = shapes,
            scales = scales,
            n = n,
            p = p,
            tau = tau)

        sol <- wei.series.md.c1.c2.c3::mle_nelder_wei_series_md_c1_c2_c3(
            df = df,
            theta0 = theta,
            reltol = 1e-7,
            parscale = parscale,
            maxit = 2000L)

        cat("mle: ", sol$par, "\n")

        sol.boot <- boot(df, function(df, i) {
            sol <- wei.series.md.c1.c2.c3::mle_nelder_wei_series_md_c1_c2_c3(
                df = df[i, ],
                theta0 = sol$par,
                reltol = 1e-7,
                parscale = parscale,
                maxit = 1000L)
            cat("boot: ", sol$par, "\n")
            sol$par
        }, ncpus = ncores, R = R)

        saveRDS(list(n = n, p = p, q = q, tau = tau, mle = sol, mle.boot = sol.boot),
            file = paste0("./results/", sim.name, "/results_", n, "_", p, "_", q, ".rds"))

        }, error = function(e) {
            print(e)
            problems <- append(problems, list(list(
                error = e, n = n, p = p, q = q, tau = tau, df = df)))
        })

    if (length(problems) != 0) {
        saveRDS(list(n = n, p = p, q = q, tau = tau, problems = problems),
                file = paste0("./problems/", sim.name, "/problems_", n, "_", p, "_", q, ".rds"))
    }
}

params <- expand.grid(n = ns, p = ps, q = qs)
result <- mclapply(
    1:nrow(params),
    function(i) sim.boot.run(sim.name, params$n[i], params$p[i], params$q[i]),
    mc.cores = ncores)
```

References

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury Advanced Series.

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Engelhardt, B. (1992). Introduction to probability and mathematical statistics. page 294.

Guess, F. M., Hodgson, T. J., and Usher, J. S. (1991). Estimating system and component reliabilities under partial information on cause of failure. *Journal of Statistical Planning and Inference*, 29:75–85.

Guo, H., Niu, P., and Szidarovszky, F. (2013). Estimating component reliabilities from incomplete system failure data. *Annual Reliability and Maintainability Symposium (RAMS)*, pages 1–6.

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