Estimating the reliability of components in a series systems from masked failure data

Alex Towell

Abstract

We estimate the reliability of series systems from masked failure data consisting of right-censored system lifetimes and masked component cause of failures. Under a minimal set of conditions that permit us to ignore how the component cause of failures are masked, we estimate the sampling distribution of the MLE in a variety of situations. We find that as long as the masking does not always include a particular component as a potential cause of failure whenever another component is also a potential cause, the MLE is unique and consistent. However, this is often not a realistic situation in industrial settings. For example, a system may be repaired by replacing an entire circuit board consisting of multiple components, in which case whenever one of the components on that circuit board fails, all of the components on the circuit board are potential component causes of failure. In this case, we have a non-unique MLE, which provides less information about the reliability of the series system and its components.

Contents

1 Introduction

Accurate reliability specifications of components of a multi-component system, such as mean time to failure (MTTF), can be quite valuable. For instance, such information may be used to help identify the component cause of system failure. However, this information is often not available, or if it is available, it may not be very accurate. In this case, we may try to estimate the reliability of each component from system failure data.

We are interested in estimating the reliability of *series systems*. In a series system, the system fails whenever any of its components fail. The famous statistician George Box once opined that all models are wrong, but some are useful, meaning that a theoretical model of real phenomenon is incapable of representing its exact behaviour, but it may still be useful. Indeed, sometimes it is even more useful, since a simplified model is more understandable and only includes the most salient features.

Since many real-world systems are greatly impaired whenever a set of its *critical* components fail, the series system model is often a very *useful* approximation of its internal structure. For example, people in a particular experiment may be regarded as series systems where the components are some relevant set of vital organs. This model is often a very useful abstraction that averages over many complicated, unobserved details that may only have a small effect on an average person's lifespan.

In this paper, the experimental unit is a particular type of series system that consists of m components and we seek to estimate the reliability of the components from *masked data* where we observe only partial information about the system, e.g., the component lifetimes are masked (not observed in the sample).

We imagine that we conduct an experiment where we obtain the lifetimes of the m components, which has a sample space $(0, \infty)^m$. An observed result is called an *outcome* of the experiment, e.g., $(1.2, 0.5, 1.1)$. However, we assume that this data is masked in the following ways:

1. The experiments may be right-censored, which means that the experiment or observation may be suspended before the system fails.

2. We can only observe the lifetime of the component with the minimum lifetime. This constraint is imposed by the fact that the system is arranged in series and thus stops working whenever one of the components fails.

3. It may not be easy or possible to identify the failed component. We consider the case where we can only observe a set of components that includes the failed component. We refer to this set of components as the candidate set.[1]

   This constraint may be due to a variety of reasons. A canonical example is given by supposing that when the series system fails, it is repaired by simultaneously replacing some subset of the components, thus preventing isolating the specific failed component.

We take a random sample of n such experiments, where the right-censored system lifetimes are i.i.d. when conditioned on the right-censoring time, but the candidate sets are not assumed to be independent but not necessarily identically distributed. In later sections, we consider a set of realistic conditions that allow us to ignore the distribution of candidate sets, which allows us to construct a reduced likelihood function for the masked failure data. Finally, we find the MLE $\hat{\boldsymbol{\theta}}$ that maximizes the reduced likelihood function, which is the same MLE for "full" likelihood function where we know the distribution of candidate sets.

2 Statistical model

Consider a system with m components, indexed by $1, 2, \ldots, m$. Each component only has two possible states, *failed* and *functioning*. The state of the system also only has two possible states, *failed* and *functioning*. The

---

[1]If the candidate set has only a single component, then we know the exact component cause of failure.

lifetime of a system is the elapsed time from when the new, functioning system is put into operation until it fails for the *first time*. We are not interested in what happens to a system after it fails, e.g., we do not consider repairing systems and putting them back into operation under further observation.

A system that is in a functioning state if and only if at least k of the m components are in a functioning state is denoted a k-out-of-m system. We narrow our scope to m-out-of-m systems, also known as *series systems*. A series system is functioning if and only if every component is functioning. Consequently, in a series system, precisely one component is the cause of failure.

Our sample consists of lifetime data for i.i.d. series systems. Since the component lifetimes are subject to chance variations, we denote the lifetime of the jth component of the ith system by the random variable T_{ij}. We assume that the m component lifetimes, $T_{i1}, \ldots, T_{im}$, are statistically independent and non-identical. In a series system, whenever any component fails, the system fails. Thus, the lifetime of the ith system, T_i, is given by the component with the smallest lifetime,

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

We assume that the component lifetimes may be adequately modeled by some parametric probability distribution. In what follows, matrices and vectors are denoted in boldface, e.g., $\boldsymbol{x}$ is a vector. The ith column and (i, j)-th element of a matrix $\boldsymbol{A}$ is denoted respectively by $\boldsymbol{A_j}$ and $\boldsymbol{A_{ij}}$. We let $\boldsymbol{\theta_j}$ denote the parameter vector of the jth component, and each $\boldsymbol{\theta_j}$ for $j = 1, \ldots, m$ may be different sizes. We define the parameter vector of the entire series system by

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

The cumulative distribution function (cdf) for a random varible T is given by

$$F_T(t) = \Pr\{T \le t\}, \tag{2.1}$$

where $\Pr\{T \in E\}$ denotes the probability that a random variable T realizes a value in E. The probability density function (pdf) for a continuous random varible T is given by

$$f_T(t) = \frac{d}{dt} F_T(t). \tag{2.2}$$

If a random variable X is discrete, then its probability mass function (pmf) f is given by $f(x) = \Pr\{X = x\}$.

If a random variable T has a parametric distribution indexed by a parameter vector $\boldsymbol{\beta}$, we denote its pdf by $f_T(t; \boldsymbol{\beta})$ and likewise for other distribution functions, e.g., the ith series system has a cdf denoted by $F_{T_i}(t; \boldsymbol{\theta})$. In the case of the lifetime distribution functions of components, we subscript the distribution functions by component index instead of the symbol for the random variable, e.g., the cdf of the jth component for the i-th system is denoted by $F_j(t; \boldsymbol{\theta_j})$ instead of $F_{T_{ij}}(t; \boldsymbol{\theta_j})$. If it is clear from the context which random variable a distribution function is for, we may drop the subscripts, e.g., $F(t)$ instead of $F_T(t)$. Finally, as an abuse of notation, we often write a function as $f(t)$ when we really mean that f is a function of variable t.

There are two particularly important functions in survival analysis, the survival function and the hazard function.

Definition 2.1. *The survival function, $R_T(t)$, of a random lifetime T is the probability that it realizes a value larger than some specified duration of time t,*

$$\begin{aligned} R_T(t) &= \Pr\{T > t\} \\ &= 1 - F_T(t). \end{aligned} \tag{2.3}$$

In other words, $R_T(t)$ denotes the probability that T survives longer than t.

The hazard function is a bit more subtle. For a random lifetime T, the probability that a failure occurs between t and Δt given that no failure occurs before time t is given by

$$\Pr\{T \le t + \Delta t | T > t\} = \frac{\Pr\{t < T < t + \Delta t\}}{\Pr\{T > t\}}.$$

The *failure rate* is given by the above divided by the length of the time interval, Δt:

$$\frac{\Pr\{t < T_{ij} < t + \Delta t\}}{\Delta t} \frac{1}{\Pr\{T_{ij} > t\}} = \frac{R_j(t; \theta_j) - R_j(t + \Delta t; \theta_j)}{R_j(t)}.$$

3

Definition 2.2. *The hazard function $h_T(t)$ for a continuous random variable T is the instantaneous failure rate at time t, which is given by*

$$\begin{aligned} h_T(t) &= \lim_{\Delta t \to 0} \frac{\Pr\{t < T < t + \Delta t\}}{\Delta t} \frac{1}{\Pr\{T > t\}} \\ &= \frac{f_T(t)}{R_T(t)}. \end{aligned} \tag{2.4}$$

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 $f(x_1, \ldots, x_m)$.

2.1 Masked data

The object of interest is the (unknown) parameter value $\boldsymbol{\theta}$. To estimate $\boldsymbol{\theta}$, we need *data*. In general, there are three types of life data: exact lifetime, interval lifetime (where the lifetime is known to have occurred between two points in time), and right censored data, where the suspension time of the experiment is reached before the system fails.

In this paper, we limit our focus to exact lifetimes and right censored data. In particular, 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 fixed right-censoring time by τ. Whether the ith system is right-censored is a random variable defined as $\delta_i = 1_{\{\tau_i < T_i\}}$ where $\delta_i = 1$ indicates right-censoring. The right-censored time of the ith system in the sample, S_i, is given by

$$S_i = \min\{T_i, \tau_i\} \tag{2.5}$$

and the indicator for right censoring, δ_i, is given by

$$\delta_i = 1_{T_i > \tau_i}. \tag{2.6}$$

We also have masked data about the component cause of failure in the form of candidate sets. A candidate set consists of some subset of component indices that plausibly contain the index of the failed component. The sample space of candidate sets are all subsets of $\{1, \ldots, m\}$, with the exception of the empty set (thus there are $2^m - 1$ possible outcomes in the sample space). If a series system is right-censored ($\delta_i = 1$), then all of the components are in the candidate set. We typically denote the candidate set corresponding to the ith system in the sample by c_i. Since the data generating process for candidate sets may be subject to chance variations, we model it as a random set $\mathcal{C}_i$. We may also view $\mathcal{C}_i$ as a random Boolean vector $X_{i1}, \ldots, X_{im}$ where $X_{ij} = 1$ denotes that the jth component is in the observed candidate set c_i.

We let $\{a_i\}_{i \leq k}$ denote the sequence $a_1, \ldots, a_k$. If it is clear from the context, we may also use $\{a_i\}$. The random sample of masked data is give by

$$\{(S_i, \delta_i, \mathcal{C}_i)\}_{i \leq n},$$

where $\mathcal{C}_i$ is the random candidate set, S_i is the right-censored lifetime, and δ_i is the indicator for right-censoring for the ith series system. We show a graphical model of how the masked data is generated in Figure 1.
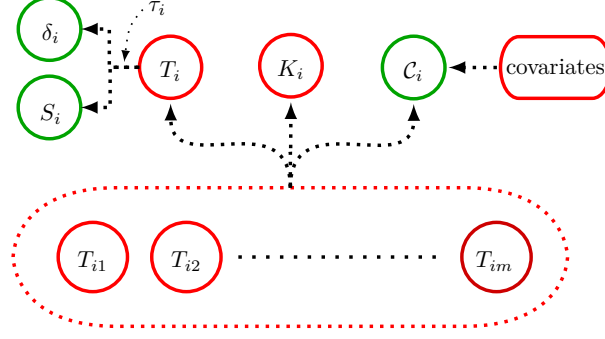
Figure 1: Dependency graph of the generative model for $(S_i, \delta_i, \mathcal{C}_i)$. The elements in green are observed in the sample, while the elements in red are unobserved. We see that $\mathcal{C}_i$ is a related to both the unobserved component lifetimes $T_{i1}, \ldots, T_{im}$ and other unknown and unobserved covariates, like ambient temperature or the particular diagnostician who generated the candidate set. These two complications for $\mathcal{C}_i$ are why we in later sections seek a way to construct a reduced likelihood function that is not a function of the distribution of $\mathcal{C}_i$.

An example of masked data for exact, right-censored system failure times with candidate sets that mask the component cause of failure can be seen in Table 1 for an 3-out-of-3 series system.

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

Right-censored time (S_i)	Right censored (δ_i)	Candidate set (C_i)
4.3	0	$\{1,2\}$
1.3	0	$\{2\}$
5.4	1	$\{\}$
2.6	0	$\{2,3\}$
3.7	0	$\{1,2,3\}$
10	1	$\{\}$

3 Series system lifetime

The reliability function of the series system is given by the following theorem.

Theorem 3.1. *The series system has a reliability function given by*

$$R(t; \boldsymbol{\theta}) = \prod_{j=1}^{m} R_j(t; \boldsymbol{\theta_j}). \tag{3.1}$$

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

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

$\square$

Theorem 3.2. *The random system lifetime T_i has a pdf given by*

$$f(t; \boldsymbol{\theta}) = \sum_{j=1}^{m} f_j(t; \boldsymbol{\theta_j}) \prod_{\substack{k=1 \\ k \neq j}}^{m} R_k(t; \boldsymbol{\theta_j}). \tag{3.2}$$

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

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

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

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

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

$\square$

Theorem 3.3. *The system lifetime T_i has a hazard function function given by*

$$h(t; \boldsymbol{\theta}) = \sum_{j=1}^{m} h_j(t; \boldsymbol{\theta_j}). \tag{3.3}$$

Proof. The ith 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

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

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

$\square$

By definition,

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

and thus

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

$$= \{\sum_{j=1}^{m} h_j(t; \boldsymbol{\theta_j})\}\{\prod_{j=1}^{m} R_j(t; \boldsymbol{\theta_j})\}, \qquad (3.4)$$

which we find to be a more convenient form than Equation (3.2).

4   Component cause of failure

If a series system failed at time t, precisely one of the components failed at time t. We model the component cause of the series system failure as a random variable.

Definition 4.1. *The component cause of failure of the ith 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.

Note that a more succinct way to define K_i is given by

$$K_i = \operatorname{argmin}_j \{T_{ij} : j \in \{1, \ldots, m\}\}.$$

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

Theorem 4.1. *The joint pdf of the component cause of failure K_i and series system lifetime T_i is given by*

$$f_{K_i, T_i}(j, t; \boldsymbol{\theta}) = h_j(t; \boldsymbol{\theta_j}) R_{T_i}(t; \boldsymbol{\theta}), \qquad (4.1)$$

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

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

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

7

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. $\qquad\square$

5 Masked data models

We consider two types of masked data, *right-censored* system lifetime data (masking the system lifetime) and *masking the component cause of failure*. In the following subsections, we derive the distribution functions for each of these types of masking.

5.1 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 joint distribution of $\mathcal{C}_i$ and T_i has the joint pdf that may be written as

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

We assume we know the pdf $f_{T_i}(t_i; \boldsymbol{\theta})$ but we do not know $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i\}$, i.e., the data generating process for candidate sets is not known. However, in order for the masked data $\mathcal{C}_i$ to provide information about $\boldsymbol{\theta}$, $\mathcal{C}_i$ must be correlated with the ith system in some useful way, in which case the conditional distribution of $\mathcal{C}_i$ given $T_i = t_i$ may be important even though the object of statistical interest is the series system rather than the candidate sets, which may be statistically problematic, e.g., $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_n$ may not be identically distributed or there may be many important unobserved covariates.

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 ith system is given by assuming the following condition.

Condition 5.1. *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 ith system.

Assuming Condition 5.1, $\mathcal{C}_i$ 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 5.1, 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 4.1, 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 jth 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

$$\begin{aligned}
f_{T_i, K_i, \mathcal{C}_i}(t_i, j, c_i; \boldsymbol{\theta}) &= f_{T_i, K_i}(t_i, k; \boldsymbol{\theta}) \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j\} \\
&= h_j(t_i; \boldsymbol{\theta_j}) R_{T_i}(t_i; \boldsymbol{\theta}) \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j\}.
\end{aligned} \tag{5.1}$$

We are going to need the joint pdf of T_i and $\mathcal{C}_i$, which may be obtained by summing over the support $\{1, \dots, m\}$ of K_i in Equation (5.1),

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

By Condition 5.1, $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j\} = 0$ 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

$$f_{T_i, \mathcal{C}_i}(t_i, c_i; \boldsymbol{\theta}) = R_{T_i}(t_i; \boldsymbol{\theta}) \sum_{j \in c_i} \left\{ h_j(t_i; \boldsymbol{\theta_j}) \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j\} \right\}. \tag{5.2}$$

When we try to find an MLE of $\boldsymbol{\theta}$ (see Section 6), 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 5.2. *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 [1], in many industrial problems, masking generally occurred due to time constraints and the expense of failure analysis. In this setting, Condition 5.2 generally holds.

Assuming Conditions 5.1 and 5.2, $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i | T_i = t_i, K_i = j\}$ may be factored out of the summation in Equation (5.2), 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 5.3. *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 5.1, 5.2, and 5.3 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 from the ith observation,

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

This derivation of l_i is the one used in many papers on masked series lifetime data (see [2]).

Suppose we have jointly observed a candidate set and a series system failure time and we are interested in the probability that a particular component is the cause of the observed system failure.

Theorem 5.1. *Assuming Conditions 5.1 and 5.2, the conditional probability of K_i given $\mathcal{C}_i = c_i$ and $T_i = t_i$ is given by*

$$\Pr\{K_i = j | T_i = t_i, \mathcal{C}_i = c_i\} = \frac{h_j(t_i | \boldsymbol{\theta_j})}{\sum_{l \in c_i} h_l(t_i | \boldsymbol{\theta_l})} 1_{\{j \in c_i\}}. \tag{5.3}$$

Proof. The conditional probability $\Pr\{K_i = j | T_i = t_i, \mathcal{C}_i = c_i\}$ may be written as

$$\Pr\{K_i = j | T_i = t_i, \mathcal{C}_i = c_i\} = \frac{\Pr_{\boldsymbol{\theta}}\{C_= c_i | K_i = j, T_i = t_i\} f_{K_i, T_i}(j, t_i; \boldsymbol{\theta})}{\sum_{j=1}^{m} \Pr_{\boldsymbol{\theta}}\{C_= c_i | K_i = j, T_i = t_i\} f_{K_i, T_i}(j, t_i; \boldsymbol{\theta})}.$$

By Theorem 4.1, $f_{K_i, T_i}(j, t_i; \boldsymbol{\theta}) = h_j(t_i; \boldsymbol{\theta}) R_{T_i}(t_i; \boldsymbol{\theta})$. We may make this substitution in the above equation and cancel the common factors $R_{T_i}(t_i; \boldsymbol{\theta})$ in the numerator and denominator, yielding

$$\Pr\{K_i = j | T_i = t_i, \mathcal{C}_i = c_i\} = \frac{\Pr_{\boldsymbol{\theta}}\{C_= c_i | K_i = j, T_i = t_i\} h_j(t_i; \boldsymbol{\theta})}{\sum_{j=1}^{m} \Pr_{\boldsymbol{\theta}}\{C_= c_i | K_i = j, T_i = t_i\} h_j(t_i; \boldsymbol{\theta})}.$$

Assuming Condition 5.1, we may rewrite the above as

$$\Pr\{K_i = j | T_i = t_i, \mathcal{C}_i = c_i\} = \frac{\Pr_{\boldsymbol{\theta}}\{C_= c_i | K_i = j, T_i = t_i\} h_j(t_i; \boldsymbol{\theta})}{\sum_{l \in c_i} \Pr_{\boldsymbol{\theta}}\{C_= c_i | K_i = l, T_i = t_i\} h_j(t_i; \boldsymbol{\theta})}.$$

Assuming Condition 5.2, we may rewrite the above as

$$\Pr\{K_i = j | T_i = t_i, \mathcal{C}_i = c_i\} = \frac{\Pr_{\boldsymbol{\theta}}\{C_= c_i | K_i = j', T_i = t_i\} h_j(t_i; \boldsymbol{\theta_j})}{\Pr_{\boldsymbol{\theta}}\{C_= c_i | K_i = j', T_i = t_i\} \sum_{l \in c_i} h_l(t_i; \boldsymbol{\theta_l})},$$

where $j' \in c_i$. Finally, we may cancel the common probability masking factors in the numerator and denominator, obtaining the result

$$\Pr\{K_i = j | T_i = t_i, \mathcal{C}_i = c_i\} = \frac{h_j(t_i; \boldsymbol{\theta_j})}{\sum_{l \in c_i} h_l(t_i; \boldsymbol{\theta_l})}.$$

$\square$

Note that we do not require Condition 5.3 in deriving the conditional distribution of K_i given T_i and $\mathcal{C}_i$.

If we do not know a particular subset $\mathcal{C}_i$, then we let $\mathcal{C}_i$ contain all m componentsn, $\{1, \ldots, m\}$, in which case we have

$$\Pr\{K_i = j | T_i = t_i\} = \frac{h_j(t_i; \boldsymbol{\theta_j})}{h(t_i; \boldsymbol{\theta})}.$$

5.2 Right-censored data

In the previous two sections, we discussed the joint distribution of T_i and $\mathcal{C}_i$. However, in our data model, we do not directly observe T_i. As described in Section 2.1, we instead 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 our candidate set model, assuming Conditions 5.1, 5.2, and 5.3 hold, the candidate set includes the index of the failed component. However, when we have right censoring, there are no failed components and therefore, there is no masking for the cause of failure. In this case, we let the candidate set be the empty set $\emptyset$.

We now present a theorem for the joint distribution of $S_i, \delta_i, \mathcal{C}_i$ given a right-censoring time τ_i.

Theorem 5.2. *The joint distribution of $S_i, \delta_i, \mathcal{C}_i$, where δ_i indicates right-censoring, is given by*

$$f(s_i, \delta_i, c_i; \boldsymbol{\theta}) = R_{T_i}(s_i; \boldsymbol{\theta})(1_{\{c_i = \emptyset\}})^{\delta_i} \left(\beta_i \sum_{j \in c_i} h_j(s_i; \boldsymbol{\theta_j}) \right)^{1 - \delta_i}. \tag{5.4}$$

Proof. There are two cases to consider.

Case 1: If $\delta_i = 0$, then $S_i < \tau_i$. In this case, $S_i = T_i$, and so

$$f_{S_i, \mathcal{C}_i}(s_i, c_i; \boldsymbol{\theta}) = f_{T_i, \mathcal{C}_i}(s_i, c_i; \boldsymbol{\theta}) 1_{\{\delta_i = 0\}} \tag{5.5}$$

$$= \beta_i R_{T_i}(s_i; \boldsymbol{\theta}) \sum_{j \in c_i} h_j(s_i; \boldsymbol{\theta_j}) 1_{\{\delta_i = 0\}}. \tag{5.6}$$

Case 2: If $\delta_i = 1$, then $S_i = \tau_i$ and $c_i = \emptyset$. In this case, we only know that $T_i > \tau_i$, and so we integrate over all possible values that it may have obtained,

$$f_{S_i, \mathcal{C}_i}(s_i, c_i; \boldsymbol{\theta}) = \int_{\tau_i}^{\infty} f_{T_i}(t; \boldsymbol{\theta}) 1_{\{s_i = \tau_i \wedge \delta_i = 1 \wedge c_i = \emptyset\}} dt \tag{5.7}$$

$$= R_{T_i}(s_i; \boldsymbol{\theta}) 1_{\{\delta_i = 1 \wedge c_i = \emptyset\}}. \tag{5.8}$$

We can add these two cases together (since only one of them is active given the indicator function), and so we can write the joint pdf as

$$f(s_i, \delta_i, c_i; \boldsymbol{\theta}) = \beta_i R_{T_i}(s_i; \boldsymbol{\theta}) \sum_{j \in c_i} h_j(s_i; \boldsymbol{\theta_j}) 1_{\{\delta_i = 0\}} + R_{T_i}(s_i; \boldsymbol{\theta}) 1_{\{\delta_i = 1 \wedge c_i = \emptyset\}} \tag{5.9}$$

$$= R_{T_i}(s_i; \boldsymbol{\theta}) \left(\beta_i \sum_{j \in c_i} h_j(s_i; \boldsymbol{\theta_j}) 1_{\{\delta_i = 0\}} + 1_{\{\delta_i = 1 \wedge c_i = \emptyset\}} \right) \tag{5.10}$$

$$= R_{T_i}(s_i; \boldsymbol{\theta})(1_{\{c_i = \emptyset\}})^{\delta_i} \left(\beta_i \sum_{j \in c_i} h_j(s_i; \boldsymbol{\theta_j}) \right)^{1 - \delta_i}. \tag{5.11}$$

$\square$

6 Maximum likelihood estimation

We use maximum likelihood estimation (MLE) to estimate the series system parameter vector $\boldsymbol{\theta}$ given the masked failure data described in Section 2.1. This is achieved by maximizing the likelihood function $L(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ so that, under the assumed model, the observed data is most likely. The point in the parameter space Ω that maximizes the likelihood function is called the maximum likelihood estimate.

6.1 Likelihood function

Given an i.i.d. random sample of masked data

$$\{(S_i, \delta_i, \mathcal{C}_i)\}_{i \leq n}$$

parameterized by $\boldsymbol{\theta} \in \boldsymbol{\Omega}$, the joint pdf is given by

$$f\big(\{(s_i, \delta_i, \mathcal{C}_i)\}_{i \leq n}; \boldsymbol{\theta}\big) = \prod_{i=1}^{n} f(s_i, \delta_i, \mathcal{C}_i; \boldsymbol{\theta}).$$

The joint pdf computes the likelihood that the observed data, $\{s_i, \delta_i, c_i\}_{i \leq n}$, occurs. When we fix the observe data and instead allow the parameter $\boldsymbol{\theta}$ to vary, we obtain what is called the likelihood function,

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

Substituting the joint pdf of $S_i, \delta_i, \mathcal{C}_i$ given by Theorem 5.2, we obtain the likelihood function for our specified data,

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} R_{T_i}(s_i; \boldsymbol{\theta}) \left\{ \beta_i \sum_{j \in c_i} h_j(s_i; \boldsymbol{\theta_j}) \right\}^{1-\delta_i}. \tag{6.1}$$

The maximum likelihood estimator is given by the estimate of $\boldsymbol{\theta}$ that maximizes the likelihood function.

According to [3], a point $\hat{\boldsymbol{\theta}}$ in $\boldsymbol{\Omega}$ at which $L(\boldsymbol{\theta})$ is a maximum is called the *maximum likelihood estimate* (MLE) of $\boldsymbol{\theta}$. That is, $\hat{\boldsymbol{\theta}}$ is a value of $\boldsymbol{\theta}$ that satisfies

$$f\big(\{(s_i, \delta_i, \mathcal{C}_i)\}; \hat{\boldsymbol{\theta}}\big) = \max_{\boldsymbol{\theta} \in \boldsymbol{\Omega}} f\big(\{(s_i, \delta_i, \mathcal{C}_i)_{i \leq n}\}; \boldsymbol{\theta}\big). \tag{6.2}$$

Essentially, the MLE is the point in the parameter space that maximizes the likelihood of the observed data.

Any point that maximizes the likelihood function also maximizes the log-likelihood function. Thus, for computational reasons, we work with the log-likelihood function.

Theorem 6.1. *The log-likelihood function, $\ell(\boldsymbol{\theta})$, is given by the logarithm of the likelihood function for the masked failure data that satisfies Conditions 5.1, 5.1, and 5.1,*

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log R_{T_i}(s_i; \boldsymbol{\theta}) + \sum_{i=1}^{n} (1 - \delta_i) \log \left\{ \sum_{j \in c_i} h_j(s_i; \boldsymbol{\theta_j}) \right\}. \tag{6.3}$$

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

$$\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}).$$

Substituting the likelihood function given by Equation (6.1), we rewrite the log-likelihood function as

$$
\begin{aligned}
\ell(\boldsymbol{\theta}) &= \log \left\{ \prod_{i=1}^{n} R_{T_i}(s_i; \boldsymbol{\theta}) \left\{ \beta_i \sum_{j \in c_i} h_j(t_i; \boldsymbol{\theta_j}) \right\}^{1-\delta_i} \right\} \\
&= \log \prod_{i=1}^{n} R_{T_i}(s_i; \boldsymbol{\theta}) + \log \prod_{i=1}^{n} \left\{ \beta_i \sum_{j \in c_i} h_j(s_i; \boldsymbol{\theta_j}) \right\}^{1-\delta_i} \\
&= \sum_{i=1}^{n} \log R_{T_i}(s_i; \boldsymbol{\theta}) + \sum_{i=1}^{n} (1 - \delta_i) \log \left\{ \sum_{j \in c_i} h_j(s_i; \boldsymbol{\theta_j}) \right\} + \text{constant},
\end{aligned}
$$

where we aggregate all the β terms into the constant (with respect to $\boldsymbol{\theta}$) and so can be ignored in maximum likelihood estimation. $\square$

6.2 Solving the maximum likelihood equations

According to [3], if $\boldsymbol{\Omega}$ is a Cartesian product of l intervals, partial derivatives of $L(\boldsymbol{\theta})$ exist, and the MLEs do not occur on the boundary of $\boldsymbol{\Omega}$, then the MLEs will be solutions of the simultaneous equations

$$\frac{\partial}{\partial \theta_j} \ell(\boldsymbol{\theta}) = 0 \tag{6.4}$$

for $j = 1, \ldots, p$ where p is the number of components in $\boldsymbol{\theta}$. We call these equations the *maximum likelihood equations*. If multiple solutions to Equation (6.4) exist, then a maximum of these solutions with respect to the likelihood function should be selected.

If there is no closed-form solution to the maximum likelihood equations (see (6.4)), we may use iterative root-finding methods to numerically approximate a solution. A general approach is, if we have a guess $\boldsymbol{\theta}^{(n)}$, take a step in a "promising" direction $\boldsymbol{d}^{(n)}$ to obtain the next guess,

$$\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} + \alpha^{(n)} \boldsymbol{d}^{(n)}, \tag{6.5}$$

where $\boldsymbol{\theta}^{(0)}$ is the initial guess, e.g., a random point in the parameter space Ω.

Assuming that at $\boldsymbol{\theta}^{(n)}$, a sufficiently small step in the direction $\boldsymbol{d}^{(n)}$ results in an improvement with respect to the log-likelihood function, we say that we *overshoot* if

$$\ell(\boldsymbol{\theta}^{(n+1)}) < \ell(\boldsymbol{\theta}^{(n)}).$$

The value $\alpha^{(n)}$ in Equation (6.5) is a positive real number chosen by a *line search* method so that we do not overshot, which has an optimal value given by

$$\alpha^{(n)} = \mathrm{argmax}_\alpha \, \ell(\boldsymbol{\theta}^{(n)} + \alpha \boldsymbol{d}^{(n)}).$$

However, this may be too computationally expensive to compute, and so we use a less optimal but faster method known as *backtracking*. In the backtracking line search method, we determine $\alpha^{(n)}$ by initially letting $\alpha^{(n)} = 1$ and then, if we overshoot, redo the update with $\alpha^{(n)} \leftarrow r\alpha^{(n)}$, $0 < r < 1$, repeating until we do not overshoot.

Note that this is not necessarily the best course of action for finding global maximums, since depending on our initial guess $\boldsymbol{\theta}^{(0)}$, we may get stuck in a local maximum. There are many other methods that are more likely to find a global maximum, such as sometimes moving in a direction that decreases the likelihood of the guess. However, for our simulations, we found that we were always able to find a global maximum using the iterative method described in Equation (6.5).

We do as many iterations in Equation (6.5) as necessary to satisfy some *stopping condition*, which is usually something simple like the distance between $\boldsymbol{\theta}^{(n)}$ and $\boldsymbol{\theta}^{(n+1)}$ being sufficiently small. Under the right conditions, for sufficiently large n, $\boldsymbol{\theta}^{(n)} \approx \hat{\boldsymbol{\theta}}$.

Two popular iterative techniques are Newton-Raphson method and gradient ascent, which respectively are defined by letting

$$\boldsymbol{d}^{(n)}_{\text{Newton-Raphson}} = J^{-1}(\boldsymbol{\theta}^{(n)}) \nabla \ell(\boldsymbol{\theta}^{(n)})$$

and

$$\boldsymbol{d}^{(n)}_{\text{gradient ascent}} = \nabla \ell(\boldsymbol{\theta}^{(n)}),$$

where $J(\boldsymbol{\theta}^{(n)})$ and $\nabla \ell(\boldsymbol{\theta}^{(n)})$ are respectively the observed Fisher information matrix (Hessian of the log-likelihood function) and the score (gradient of the log-likelihood function), each evaluated at $\boldsymbol{\theta}^{(n)}$.

Depending upon the nature of the log-likelihood function, one or the other may work better in practice.[2] The R code for iterative methods is in Appendix C.

6.3   Properties of the MLE

According to [3], if certain regularity conditions are satisfied, then solutions of the maximum likelihood equation (6.4) have the following desirable properties:

1. $\hat{\boldsymbol{\theta}}$ exists and is unique.

2. $\hat{\boldsymbol{\theta}}$ is an asymptotically unbiased estimator.

3. $\hat{\boldsymbol{\theta}}$ is asymptotically the UMVUE, the uniform minimum variance unbiased estimator.

---

[2]For example, one approach may be more computationally efficient.

4. $\hat{\boldsymbol{\theta}}$ is asymptotically normal with a mean $\boldsymbol{\theta}$ and a variance-covariance that is the inverse of the Fisher information matrix, whose (i, j)-th component is given by

$$I(\boldsymbol{\theta})_{ij} = nE_{\boldsymbol{\theta}}\left(-\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(S_i, \delta_i, C_i; \boldsymbol{\theta})\right).$$

Since it is frequently problematic taking the expectation to derive $I(\theta)$, we instead use the *observed* information matrix, which is conditioned on the observed data,

$$J(\boldsymbol{\theta})_{ij} = -\frac{\partial}{\partial\theta_i\partial\theta_j}\ell(\boldsymbol{\theta}).$$

Since we do not know $\boldsymbol{\theta}$, we estimate $J(\boldsymbol{\theta})$ with $J(\hat{\boldsymbol{\theta}})$. Thus, assuming the regularity conditions are satisfied, then approximately,

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\hat{\boldsymbol{\theta}}, J^{-1}(\hat{\boldsymbol{\theta}}))$$

and as the sample size goes to infinity, $\hat{\boldsymbol{\theta}}$ converges in distribution to $\mathcal{N}(\hat{\boldsymbol{\theta}}, J^{-1}(\hat{\boldsymbol{\theta}}))$.

7 Analysis of the MLE

7.1 Bootstrapping

Suppose we have some random sample $\{\boldsymbol{X_i}\}_1^n$ and we assume some parametric model

$$\boldsymbol{X_i} \sim f(\,\cdot\,; \boldsymbol{\beta}).$$

In the parametric Bootstrap, we estimate $\boldsymbol{\beta}$ with $\hat{\boldsymbol{\beta}}$ and generate B random samples of size n from the estimate of the distribution,

$$\boldsymbol{X_{ij}} \sim f(\,\cdot\,; \hat{\boldsymbol{\beta}}).$$

for $j = 1, \ldots, B$ and $i = 1, \ldots, n$, resulting in B bootstrapped samples each of size n,

$$\mathcal{D} = \big\{\{\boldsymbol{X_{11}}, \boldsymbol{X_{12}}, \ldots, \boldsymbol{X_{1B}}\}, \ldots, \{\boldsymbol{X_{n1}}, \boldsymbol{X_{n2}}, \ldots, \boldsymbol{X_{nB}}\}\big\}.$$

We may then apply a statistic to each sample in $\mathcal{D}$, such as the MLE, to generate a empirical sampling distribution of the statistic.

The most common form of the bootstrap is the non-parametric bootstrap. In the nonparametric bootstrap, the random data is created by resampling with replacement from the original data.

Since we do not know (nor attempt to model) the distribution of candidate sets, this non-parametric form is ideal, since we can simply resample from the original data to approximate its empirical distribution.

7.2 Estimates of accuracy and precision

We would like to measure the accuracy and precision of $\hat{\boldsymbol{\theta}}$. The bias

$$\mathrm{b}(\hat{\boldsymbol{\theta}}) = E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}$$

is a measure of accuracy and variance is a measure of precision.

Assuming the regularity conditions are met, asymptotically the bias of the MLE is 0. However, as claimed previously, these regularity conditions may not be met depending upon how candidate sets are generated. Since we know $\boldsymbol{\theta}$, we can estimate the bias quite easily.

Since the inverse of observed information matrix $J^{-1}(\hat{\boldsymbol{\theta}})$ is an estimate of the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, we can estimate the variance of the jth component of $\hat{\boldsymbol{\theta}}$ with

$$\hat{\sigma}_j^2 = J^{-1}(\hat{\boldsymbol{\theta}})_{jj}.$$

The mean squared error, MSE, is a measure of estimator error that incorporates both the bias and the variance. If the regularity conditions are met,

$$\text{MSE}(\hat{\boldsymbol{\theta}}) = \text{trace}\big(J^{-1}(\hat{\boldsymbol{\theta}})\big) + \text{b}^2(\hat{\boldsymbol{\theta}})$$

and as $n \to \infty$, the asymptotic MSE goes to

$$\text{trace}\big(J^{-1}(\hat{\boldsymbol{\theta}})\big),$$

in which case for sufficiently large samples, $\text{MSE}(\hat{\boldsymbol{\theta}})$ is approximately equal to the trace of inverse of the observed information matrix.

Assuming the regularity conditions, an independent $(1-\alpha) \times 100\%$ confidence interval for θ_j is given by

$$\hat{\theta}_j \pm z_{1-\alpha/2}\sqrt{J^{-1}(\hat{\boldsymbol{\theta}})_{jj}}.$$

7.3   Goodness of fit for real-world data

In the case of real-world data sets, we do not know the true distribution and thus cannot use a performance measure that is a function of the unknowns, like $\boldsymbol{\theta}$. Moreover, the assumptions baked into our statistical model may not be appropriate.

We seek to assess the goodness of fit of our model with respect to the observed system lifetimes $T_1, \ldots, T_n$ using the Cramér-von Mises test, in which we judge the goodness of fit of the estimated theoretical parametric cdf $F_{T_i}(t; \hat{\boldsymbol{\theta}})$ to the empirical cdf F_n.

8   Components with exponentially distributed lifetimes

Consider a series system in which the components have exponentially distributed lifetimes. The jth component of the ith has a lifetime distribution given by

$$T_{ij} \sim \text{EXP}(\lambda_j)$$

for $j = 1, \ldots, m$. Thus, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$. The random variable T_{ij} has a reliability, pdf, and hazard function given respectively by

$$R_j(t; \lambda_j) = \exp(-\lambda_j t), \tag{8.1}$$
$$f_j(t; \lambda_j) = \lambda_j \exp(-\lambda_j t), \tag{8.2}$$
$$h_j(\cdot; \lambda_j) = \lambda_j \tag{8.3}$$

where $t > 0$ and $\lambda_j > 0$ is the failure rate (or inverse scale) parameter.

The lifetime of the series system composed of m components with exponentially distributed lifetimes has a reliability function given by

$$R(t; \boldsymbol{\theta}) = \exp\left(-\sum_{j=1}^{m} \lambda_j t\right) \tag{8.4}$$

where $t > 0$.

Proof. By Theorem 3.1,

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

Plugging in the component reliability functions given by Equation (8.1) obtains the result

$$R(t; \boldsymbol{\lambda}) = \prod_{j=1}^{m} = \exp(-\lambda_j t)$$

$$= \exp\left(-\sum_{j=1}^{m} \lambda_j t\right).$$

$\square$

A series system with exponentially distributed lifetimes is also exponentially distributed.

Theorem 8.1. *The random lifetime T_{ij} of a series system composed of m components with exponentially distributed lifetimes is exponentially distributed with a failure rate that is the sum of the component failure rates,*

$$T_i \sim \mathrm{EXP}\Big(\sum_{j=1}^{m} \lambda_j\Big).$$

Proof. By Equation (8.4), the series system has a reliability function in the family of the exponential distribution with a failure rate that is the sum of the component failure rates. □

The series system's failure rate function is given by

$$h(\cdot; \boldsymbol{\lambda}) = \sum_{j=1}^{m} \lambda_j \tag{8.5}$$

whose proof follows from Theorem 3.3.

We see that the failure rate $\lambda = \sum_{j=1}^{n} \lambda$ is *constant*, consistent with the the exponential distribution being the only continuous distribution that has a constant failure rate.

The pdf of the series system is given by

$$f(t; \boldsymbol{\lambda}) = \Big(\sum_{j=1}^{m} \lambda_j\Big) \exp\Big(-\sum_{j=1}^{m} \lambda_j t\Big) \tag{8.6}$$

where $t > 0$.

Proof. By definition,

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

Plugging in the failure rate and reliability functions given respectively by Equations (8.5) and (8.1) completes the proof. □

The conditional probability that component k is the cause of a system failure at time t is given by

$$f_{K_i|T_i}(k|t, \boldsymbol{\lambda}) = f_{K_i}(k|\boldsymbol{\lambda}) = \frac{\lambda_k}{\sum_{p=1}^{m} \lambda_p} \tag{8.7}$$

where $k \in \{1, \ldots, m\}$ and $t > 0$.

Proof. By Theorem **??**,

$$f_{K_i|T_i}(k; \boldsymbol{\lambda}|t) = \frac{h_k(t; \lambda_k)}{h(t; \boldsymbol{\lambda})}.$$

Plugging in the failure rate of the component indexed by k and the failure rate of the system given respectively by Equations (8.5) and (8.3) completes the proof. □

Due to the constant failure rates of the components, K_i and T_i are mutually independent. The joint pdf of K_i and T_i is given by

$$f_{K_i, T_i}(k, t; \boldsymbol{\lambda}) = \lambda_k \exp\Big(-\sum_{j=1}^{m} \lambda_j t\Big) \tag{8.8}$$

where $k \in \{1, \ldots, m\}$ and $t > 0$.

Proof. By definition,

$$f_{K_i, T_i}(k, t; \boldsymbol{\lambda}) = f_{K_i|T_i}(k; \boldsymbol{\lambda}|t) f_{T_i}(t; \boldsymbol{\lambda}).$$

Plugging in the conditional probability and the marginal probability given respectively by Equations (8.7) and (8.6) completes the proof. □

8.1 Maximum likelihood estimation

The log-likelihood for our masked data model when we assume Conditions 5.1, 5.2, and 5.3 is given by

$$\ell(\boldsymbol{\lambda}) = \sum_{i=1}^{n}(1-\delta_i)\log\left(\sum_{j\in c_i}\lambda_j\right) - \left(\sum_{i=1}^{n}s_i\right)\left(\sum_{j=1}^{m}\lambda_j\right). \tag{8.9}$$

Proof. By Equation (6.3),

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\log R(s_i;\boldsymbol{\theta}) + \sum_{i=1}^{n}(1-\delta_i)\log\left\{\sum_{k\in c_i}h_k(s_i;\boldsymbol{\theta_k})\right\}.$$

Plugging in the component failure rate and system reliability functions given respectively by Equations (8.3) and (8.4) and simplifying completes the proof. $\qquad\square$

The set of solutions to the MLE equations must be stationary points, i.e., a point at which the score function of type $\mathbb{R}^m \mapsto \mathbb{R}^m$ is zero. The j-th component of the output of the score function is given by

$$\frac{\partial\ell}{\partial\lambda_p} = \sum_{i=1}^{n}\left(\sum_{j\in c_i}\lambda_j\right)^{-1}1_{\{p\in c_i \text{ and } \delta_i=0\}} - \sum_{i=1}^{n}s_i. \tag{8.10}$$

We may find an MLE by solving the maximum likelihood equation (6.4), i.e., a set of (stationary) points satisfying

$$\left.\frac{\partial\ell}{\partial\lambda_j}\right|_{\hat{\lambda}_j} = 0$$

for $j = 1, \ldots, m$. We approximate a solution to this problem by using the iterative Newton-Raphson method as described in Section 6.2.

The Newton-Raphson method needs the observed information matrix, which is a function of $\boldsymbol{\lambda}$ of type $\mathbb{R}^m \mapsto \mathbb{R}^{m\times m}$. The (j,k)-th element of $J(\boldsymbol{\lambda})$ is given by

$$\frac{\partial^2\ell}{\partial\lambda_j\partial\lambda_k} = \sum_{i=1}^{n}\left(\sum_{j\in c_i}\lambda_j\right)^{-2}1_{\{j,k\in c_i \text{ and } \delta_i=0\}}. \tag{8.11}$$

8.2 Simulation

We wish to simulate a masked data set consisting with Conditions 5.1, 5.2, and 5.3. We setup the parameters of this simulation with:

```
# sample size
n <- 500
# m components in series system
m <- 3
# with probability q an observation will be right-censored
q <- 0.25
# true parameter value of series system
lambda <- c(2,3,2.5)
# right censoring time
tau <- rep(-(1/sum(lambda))*log(q),n)
```

We generate a masked data sample of size $n = 500$, $m = 3$, $\tau_i = 0.1848$, $\boldsymbol{\lambda} = (2, 3, 2.5)'$, and using the simple Bernoulli candidate model that is consistent with conditions C_1, C_2, and C_3 with:

```r
md <- tibble(t1=stats::rexp(n,rate=lambda[1]),
             t2=stats::rexp(n,rate=lambda[2]),
             t3=stats::rexp(n,rate=lambda[3])) %>%
    md_series_lifetime() %>%
    md_series_lifetime_right_censoring(tau) %>%
    md_bernoulli_candidate_C1_C2_C3(m)
```

See Appendix A for more information about the `md_*` functions and see The above code produces a data frame `md` whose first 10 rows are given by Table **??**.

Now, we find the MLE with the following R code:

```r
loglike.exp <- md_loglike_exp_series_C1_C2_C3(md)
lambda.hat <- algebraic.mle::mle_gradient_ascent(
    l=loglike.exp,
    theta0=lambda)
points <- cbind(point(lambda.hat),as.matrix(lambda))
colnames(points) <- c("MLE","Parameter")
cbind(points,confint(lambda.hat))
```

In Section 7.2, we consider various ways of analyzing the accuracy and precision of the MLE $\hat{\boldsymbol{\lambda}}$. To do these estimations, we perform Bootstrapping to generate $R = 1000$ MLEs.

TODO: I show a bunch of different estimates, like bias and confidence intervals, using the Bootstrap method. I need to work on presenting only interesting estimates in a way that helps show what the MLE is like for small samples, then show how as the sample gets larger, we become consistent with the asymptotic theory of the MLE. So, basically a graph whose x-axis is sample size and whose y-axis is MSE or bias.

```r
N <- c(100,200,300,400,500)
for (n in N)
{
    md <- tibble(t1=stats::rexp(n,rate=lambda[1]),
                 t2=stats::rexp(n,rate=lambda[2]),
                 t3=stats::rexp(n,rate=lambda[3])) %>%
        md_series_lifetime() %>%
        md_bernoulli_candidate_C1_C2_C3(m)

    loglike.exp <- md_loglike_exp_series_C1_C2_C3(md)
    loglike.scr <- md_score_exp_series_C1_C2_C3(md)
    mle.exp <- mle_gradient_ascent(l=loglike.exp,theta0=lambda,score=loglike.scr)
    mle.boot <- mle_boot_loglike(mle=mle.exp,
                                 loglike.gen=md_loglike_exp_series_C1_C2_C3,
                                 data=md)

    cat("Sample size: ", n, "\n")
    cat("--------------\n")

    print("Estimate of bias using Bootstrap:")
    print(bias(mle.boot,lambda))

    print("Estimate of confidence intervals:")
    print("Asymptotic")
    print(confint(mle.exp))
    print("Bootstrap")
    print(confint(mle.boot))
```

```
    print("Estimate of variance-covariance:")
    print("Asymptotic")
    print(vcov(mle.exp))
    print("Bootstrap")
    print(vcov(mle.boot))

    print("Estimate of MSE:")
    print("Asymptotic")
    print(mse(mle.exp))
    print("Bootstrap")
    print(mse(mle.boot,lambda))
}
```

The asymptotic mean squared error of $\hat{\boldsymbol{\lambda}}^{(1)}, \ldots, \hat{\boldsymbol{\lambda}}^{(N)}$ is approximately equal to the trace of the inverse of the Fisher information matrix evaluated at $\boldsymbol{\lambda}$, which is given by

8.3 Candidate sets that generate non-unique MLEs

In some cases, the MLE may be non-unique due to a small sample size. However, there are cases where as the sample size $n \to \infty$, the MLE remains non-unique, i.e., it is not guaranteed to converge to any value. One way in which this can occur for a series system with masked component causes of failure is given by the following setup.

Let the exponential series system consist of $m = 3$ components. Components 1 and 2 are on the same circuit board, and component 3 is on another circuit board. Whenever the series system fails, the potential components that may have caused the failure is determined by replacing the entire circuit board. Thus, when compoent 1 or 2 fails, the series system is fixed by replacing the circuit board they both reside on, and thus the candidate set is $\{1, 2\}$. If component 3 fails, the candidate set is $\{3\}$. In this case, the log-likelihood function is given by

$$\ell(\lambda_1, \lambda_2, \lambda_3) = \eta_3 \log(\lambda_3) + \eta_{12} \log(\lambda_1 + \lambda_2) - (\eta_3 + \eta_{12})\bar{s}(\lambda_1 + \lambda_2 + \lambda_3),$$

where η_3 denotes the number of observations in which the candidate set is $\{3\}$ and η_{12} denotes the number of observations in which the candidate set is $\{1, 2\}$.

Let $\hat{\lambda}_1 = 1$, $\hat{\lambda}_2 = 2$, and $\hat{\lambda}_3 = 3$. Then,

$$\ell(1, 2, 3) = \eta_3 \log(3) + \eta_{12} \log(1 + 2) - (\eta_3 + \eta_{12})\bar{s}(1 + 2 + 3).$$

However, if we interchange the values for $\hat{\lambda}_1$ and $\hat{\lambda}_2$, we get the same value for the log-likehood function. Thus, any time $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3)$ is an MLE, so is $(\hat{\lambda}_2, \hat{\lambda}_1, \hat{\lambda}_3)$. This is known as *non-identifiability*, which means that two or more values of the parameters result in the same likelihood of the observed data.

To find an MLE, we take the derivative of ℓ with respect to the parameters, obtaining the simultaneous equations

$$\hat{\lambda}_1 + \hat{\lambda}_2 = \frac{\eta_{12}}{(\eta_{12} + \eta_3)\bar{s}} \tag{8.12}$$

$$\hat{\lambda}_3 = \frac{\eta_3}{(\eta_{12} + \eta_3)\bar{s}} \tag{8.13}$$

We see that

$$\hat{\lambda}_1 + \hat{\lambda}_2 = \frac{\eta_{12}}{(\eta_{12} + \eta_3)\bar{s}}$$

defines a *line*, and thus any point on this line is an MLE of $\hat{\lambda}_1$ and $\hat{\lambda}_2$.

As a demonstration of this occurrence, we run $N = 500000$ simulations for an exponential series system with a true parameter $\boldsymbol{\lambda} = (2, 3, 2.5)'$, starting at a random location with the parameter space Ω. In Table **??**, we show a small sample of the generated masked data.

We find an MLE for each, and do density plots for $\hat{\lambda}_1$ and $\hat{\lambda}_2$ on the left side and $\hat{\lambda}_3$ on the right side in Figure **??**.

9 References

[1] 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.

[2] GUO, H., NIU, P. and SZIDAROVSZKY, F. (2013). Estimating component reliabilities from incomplete system failure data. *Annual Reliability and Maintainability Symposium (RAMS)* 1–6.

[3] ENGELHARDT, B. (1992). Introduction to probability and mathematical statistics. 294.

10 Appendix

.1 Data

.2 R code