Estimating component distributions of series systems from masked failure data

Alexander Towell

atowell@siue.edu

Contents

List of Figures

List of Algorithms

1 Introduction

> All models are wrong but some are useful.
>
> *George Box*

George Box, an influential statistician, popularly coined the phrase "all models are wrong, but some are useful." A model's *raison d'etre* is not necessarily *truth* but *utility*.

Models are simplified abstractions of a more complex underlying reality. For example, to facilitate navigation from point A to point B, a map maker may decide to only represent roads, highways, and named places on a map, with the assumption that representing other features detracts from the navigation objective. For a more mathematical example, consider Hook's law, a mathematical relationship in which the force required to extend or compress a spring by some distance is proportional to that distance. This model, the ideal spring model, does not exactly represent the behavior of any physical spring, but in many cases it is still a useful approximation. Moreover, its simplicity makes it easier to reason about, which can contribute to a deeper understanding than would be possible with a more accurate, but also more complex, model.

Hooke's law is a deterministic model, but sometimes we need to model the uncertainty (often due to incomplete information) exhibited by a system. In reliability analysis, a primary objective is to model a system's lifetime distribution. Once this model is attained, it may be used to answer questions about the system. For example, one may be interested in answering the question, "What is the probability a given television will continue to function after one year of operation?" There are conceivably an uncountable number of ways for the television to fail, but since we do not know how to model all such events we represent the television's lifetime using a much simpler statistical model. The statistical model is *useful* if it is sufficiently able to predict the probability that the television will continue to operate after one year of operation.

Research objectives The lifetime distribution of complex systems can often be usefully approximated as series systems consisting of m components with independently distributed lifetimes. First, we consider generalized series systems in which the lifetime of component j for $j = 1, \ldots, m$ is distributed according to some parametrized distribution family, $T_j \sim f_j(t \mid \boldsymbol{\theta}_j^\star)$ where $\boldsymbol{\theta}_j^\star$ is the true parameter value. Second, we show how maximum likelihood estimation (MLE) on a sample of masked system failures can be used to estimate parameter characteristics of interest that depend on $\boldsymbol{\theta}_j^\star$ for $j = 1, \ldots, m$. Moreover, we demonstrate ways to estimate the accuracy of such estimators. After developing the general series system, we focus our

attention on specific kinds of series systems in which components have lifetimes distributed from either the Exponential distribution, the Weibull distribution, or the Pareto distribution.

2 General series systems

2.1 Component lifetimes

Consider a series system composed of m components. Each component j has an uncertain lifetime denoted by the continuous random variable T_j for $j = 1, \ldots, m$. We make the following assumptions about the components. (i) The random variables $T_1, \ldots, T_m$ are independent of each other, i.e., the probability that component j has failed is not affected by any other set of components failing. (ii) Components are either in a failed state or a non-failed state, i.e., there is no degree-of-failure.

The random variable T_j can be uniquely specified in many different ways, but given any unique specification it must be possible to derive any other specification through mathematical manipulation. Consequently, to represent random variable T_j, only one unique specification is necessary. However, for convenience we will specify T_j with respect to the three unique specifications: $R_j(t)$, $F_j(t)$, and $f_j(t)$.

The function $R_j(t)$ is the reliability (or survival) function for the lifetime of component j; it maps time t to the probability component j will not fail before the given time t. The function $F_j(t)$ is the cumulative density function (CDF) for the lifetime of component j; it is defined to be the complement of $R_j(t)$) and therefore it maps time t to the probability component j will fail before time t. For a continuous random variable T_j, the function $f_j(t)$ is the probability density function (PDF) of the lifetime of component j; it maps time t to the *relative likelihood* component j will fail at time t. More formally, its integral, $\int_a^b f_j(s)\mathrm{d}s$, calculates the probability that component j will fail at some time $t \in (a, b)$.

Definition 1. *The three functions, $R_j(t)$, $F_j(t)$, and $f_j(t)$ are related in the following ways:*

$$R_j(t) = \Pr[T_j > t] = 1 - F_j(t) = \int_t^\infty f_j(s)\mathrm{d}s \tag{1}$$

$$F_j(t) = \Pr[T_j <= t] = 1 - R_j(t) = \int_0^t f_j(s)\mathrm{d}s \tag{2}$$

$$f_j(t) = \frac{\mathrm{d}}{\mathrm{d}t}F_j(t) = -\frac{\mathrm{d}}{\mathrm{d}t}R_j(t) \tag{3}$$

where $f_j(t) = 0$, $F_j(t) = 0$, and $R_j(t) = 1$ if $t < 0$.

Parametric distributions We make one additional assumption about the components. Recall that component j has an uncertain lifetime $T_j \sim f_j(t)$. Ideally, the lifetime distribution of component j could be completely specified, but in practice this is usually not possible. Instead, we assume that the distribution of T_j is specified by some parametrized distribution such that $T_j \sim f_j(t \mid \boldsymbol{\theta}_j)$.

These distribution functions are used to examine characteristics of interest about the series system. For instance, the instantaneous failure rate of component j is a popular characteristic of interest in reliability analysis.

Definition. *[Failure rate] The instantaneous failure rate $\lambda_j(t)$ denotes the relative frequency of failures at time t for component j and is defined as*

$$\lambda_j(t) = \frac{f_j(t)}{R_j(t)}. \tag{4}$$

Intuitively, instantaneous failure rate is to failures as instantaneous velocity is to distance. To formalize this intuition, consider the conditional probability $\Pr[t < T_j < t + h \mid T_j > t]$, which denotes the probability that component j will fail during the time interval $(t, t + h)$ given that it has not failed before time t. By Bayes' law, this is equivalent to

$$\frac{\Pr[t < T_j < t + h]}{\Pr[T_j > t]} = \frac{F_j(t + h) - F_j(t)}{R_j(t)}.$$

If we divide this conditional probability by the length of the time interval, h, we arrive at the average failure rate over the time interval $(t, t + h)$. Letting $h \to 0$ we arrive at the instantaneous failure rate

$$\lambda_j(t) = \lim_{h \to 0} \frac{F_j(t + h) - F_j(t)}{h} \cdot \frac{1}{R_j(t)} = \frac{f_j(t)}{R_j(t)}.$$

2.2 System lifetimes

A system composed of m components can have many different kinds of structures. For instance, the system could be parallel, series, or any intermediate structure between these two extremes. Each structure can be represented by a structure function, $\phi(x_1, \ldots, x_m) : \{0, 1\}^m \mapsto \{0, 1\}$. The variable x_j is a Boolean variable indicating whether the j^{th} component has failed, denoted by $x_j = 0$, or not failed, denoted by $x_j = 1$. Analogously, for a given $x_1, \ldots, x_m$, if $\phi(\cdot) = 0$, the system has failed, and if $\phi(\cdot) = 1$, the system has not failed.

We restrict our attention to series systems composed of m components. A series system has a structure

function

$$\phi(x_1, \ldots, x_m) = x_1 \cdot x_2 \cdot \ldots \cdot x_m = \min(x_1, x_2, \ldots, x_m).$$

Consequently, in a series system, the system is in a failed state if $x_j = 0$ for any $j \in \{1, \ldots, m\}$. That is, the series system fails if any component fails. Due to the necessity that m out of the m components must be in a non-failed state for the system to be in a non-failed state, series systems composed of m components are sometimes denoted m-out-of-m systems.

Remark. Restricting attention to m-out-of-m systems is not as limiting as it may seem since any coherent system[1] can be decomposed into a set of k series structures with structure functions ϕ_1, ..., ϕ_k in which the system fails if and only if $\phi_j(x_1, \ldots, x_m) = 0$ for $j = 1, \ldots, k$.

Consider a series system composed of m components.

Theorem 2.1. *An m-out-of-m system has a random lifetime T that is a function of the random lifetimes $T_1, \ldots, T_m$ of the m components,*

$$T = \min(T_1, \ldots, T_m).$$

where $T_1, \ldots, T_m$ are the random variables denoting the lifetime distributions of the m components in the system.

Proof. A series system fails whenever any component fails. Therefore, its lifetime is equal to the lifetime of the component with the minimum[2] lifetime. □

The system reliability function of the system's lifetime, $R_T(t \mid \boldsymbol{\theta})$, denotes the probability the system will not fail before time t and is given in the following theorem.

Theorem 2.2 (Reliability function). *The random variable T for the lifetime of the m-out-of-m system has a reliability function*

$$R_T(t \mid \boldsymbol{\theta}) = \prod_{j=1}^{m} R_j(t \mid \boldsymbol{\theta}_j), \tag{5}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is partitioned into a disjoint set of vectors $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m\}$ such that the lifetime of component j has a distribution parametrized by $\boldsymbol{\theta}_j$ for $j = 1, \ldots, m$.

Proof. In order for the series system to not fail before time t, it must be the case that $T > t$, in which case

$$R_T(t) = \Pr[T > t] = \Pr[\min(T_1, \ldots, T_m) > t] = \Pr[\text{no component fails before time } t].$$

[1] In a coherent system, all components in the system are relevant. See [1] for more on coherent systems.
[2] A chain is only as strong as its weakest link.

In order for no component to fail before time t, it must be the case that

$$R_T(t) = \Pr[T_1 > t, \ldots, T_m > t].$$

By the assumption that the component lifetimes are independent, this is equivalent to

$$R_T(t) = \Pr[T_1 > t] \cdots \Pr[T_m > t] = \prod_{j=1}^{m} R_j(t).$$

□

The CDF of the system's lifetime, $F_T(t \mid \boldsymbol{\theta})$, denotes the probability the system will fail before time t and is given in the following theorem.

Theorem 2.3 (CDF). *The random variable T for the lifetime of the m-out-of-m system has a reliability function CDF*

$$F_T(t \mid \boldsymbol{\theta}) = 1 - \prod_{j=1}^{m} R_j(t \mid \boldsymbol{\theta}_j), \tag{6}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is partitioned into a disjoint set of vectors $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m\}$ such that the lifetime of component j has a distribution parametrized by $\boldsymbol{\theta}_j$ for $j = 1, \ldots, m$.

Proof. By definition, $F_T(t)$ is the complement of the reliability function eq. (5). Thus,

$$F_T(t) = 1 - R_T(t) = 1 - \prod_{j=1}^{m} R_j(t).$$

□

The PDF of the system's lifetime, $f_T(t \mid \boldsymbol{\theta})$, denotes the *relative likelihood* the system will fail at time t and is given in the following theorem.

Theorem 2.4 (PDF). *The random variable T for the lifetime of the m-out-of-m system has a PDF*

$$f_T(t \mid \boldsymbol{\theta}) = \sum_{j=1}^{m} \left(f_j(t \mid \boldsymbol{\theta}_j) \prod_{\substack{k=1 \\ k \neq j}}^{m} R_k(t \mid \boldsymbol{\theta}_k) \right). \tag{7}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is partitioned into a disjoint set of vectors $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m\}$ such that the lifetime of component j has a distribution parametrized by $\boldsymbol{\theta}_j$ for $j = 1, \ldots, m$.

Proof. The PDF $f_T(t)$ is just the derivative (with respect to time t) of the CDF $F_T(t)$,

$$f_T(t) = \frac{\mathrm{d}}{\mathrm{d}t} FT(t) = -\frac{\mathrm{d}}{\mathrm{d}t} RT(t).$$

By theorem 2.2, this is equivalent to

$$f_T(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \prod_{j=1}^{m} R_j(t).$$

By the product rule, this is equivalent to

$$\begin{aligned} f_T(t) &= -\left(R_1(t) \frac{\mathrm{d}}{\mathrm{d}t} \prod_{j=2}^{m} R_j(t) - f_1(t) \prod_{j=2}^{m} R_j(t) \right) \\ &= f_1(t) \prod_{j=2}^{m} R_j(t) - R_1(t) \frac{\mathrm{d}}{\mathrm{d}t} \prod_{j=2}^{m} R_j(t) \end{aligned}$$

By applying the product rule again, we arrive at

$$\begin{aligned} f_T(t) &= f_1(t) \prod_{j=2}^{m} R_j(t) + R_1(t) \left(f_2(t) \prod_{j=3}^{m} R_j(t) - R_2(t) \frac{\mathrm{d}}{\mathrm{d}t} \prod_{j=2}^{m} R_j(t) \right) \\ &= f_1(t) \prod_{j=2}^{m} R_j(t) + f_2(t) \prod_{\substack{j=1 \\ j \neq 2}}^{m} R_j(t) - R_1(t) R_2(t) \frac{\mathrm{d}}{\mathrm{d}t} \prod_{j=3}^{m} R_j(t) \end{aligned}$$

We begin to see a pattern emerge. After applying the product rule r times, we arrive at

$$f_T(t) = \sum_{j=1}^{r} f_j(t) \prod_{\substack{k=1 \\ k \neq j}}^{m} R_k(t) - \prod_{j=1}^{r} R_k(t) \frac{\mathrm{d}}{\mathrm{d}t} \prod_{j=r+1}^{m} R_r(t)$$

Continuing in this fashion, after applying the product rule m times, we finally arrive at

$$f_T(t) = \sum_{j=1}^{m} \left(f_j(t) \prod_{\substack{k=1 \\ k \neq j}}^{m} R_k(t) \right)$$

□

The failure rate of the system, $\lambda(t)$, is given in the following theorem.

Theorem 2.5 (Failure rate). *The random variable T for the lifetime of the m-out-of-m system has a failure*

rate function

$$\lambda(t) = \sum_{j=1}^{m} \lambda_j(t), \tag{8}$$

where $\lambda_j(t)$ is the failure rate of component j.

Proof. The failure rate is defined to be

$$\lambda(t) = \frac{f_T(t)}{R_T(t)}.$$

Plugging in these distribution functions, we get

$$\lambda(t) = \frac{\sum_{j=1}^{m} \left(f_j(t) \prod_{\substack{k=1 \\ k \neq j}}^{m} R_k(t) \right)}{\prod_{j=1}^{m} R_j(t)}.$$

This simplifies to

$$\lambda(t) = \sum_{j=1}^{m} \left(f_j(t)/R_j(t) \right) = \sum_{j=1}^{m} \lambda_j(t).$$

□

2.3 Conditional system lifetimes

Suppose it is given that one of the components in a candidate set of k components, $\mathrm{c} = \{j_1, \ldots, j_k\} \subset \{1, \ldots, m\}$, is the cause of the system failure.

Theorem 2.6. *An m-out-of-m system has a conditional random lifetime $T \mid \mathrm{c}$ that is a function of the lifetimes of the candidate components,*

$$T \mid \mathrm{c} = \min(T_{j_1}, \ldots, T_{j_k}),$$

where $T_{j_1}, \ldots, T_{j_k}$ are the random variables denoting the lifetime distributions of the k components in the candidate set c.

Proof. A series system fails whenever any of the components in the candidate set c fails. Therefore, its lifetime is equal to the lifetime of the candidate component with the minimum lifetime. □

The CDF of the system's conditional lifetime, $F_{T|\mathrm{c}}(t \mid \mathrm{c}, \boldsymbol{\theta})$, denotes the probability the system will fail before time t given that one of the components in the candidate set c causes the system failure. The CDF is given in the following theorem.

Theorem 2.7 (Conditional CDF). *The conditional random variable* $T \mid \mathrm{c}$ *for the lifetime of the* m*-out-of-*m *system has a conditional CDF*

$$F_{T|\mathrm{c}}(t \mid \mathrm{c}, \boldsymbol{\theta}) = \int_0^t \left[\sum_{j \in \mathrm{c}} f_j(s \mid \boldsymbol{\theta}_j) \prod_{\substack{k=1 \\ k \neq j}}^{m} R_k(s \mid \boldsymbol{\theta}_k) \right] \mathrm{d}s \tag{9}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is partitioned into a disjoint set of vectors $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m\}$ such that the lifetime of component j has a distribution parametrized by $\boldsymbol{\theta}_j$ for $j = 1, \ldots, m$.

Proof. Let $\mathrm{c} = \{j_1, \ldots, j_k\}$ be the set of k components that are candidates for causing the system failure. The $F_{T|\mathrm{c}}(t \mid \mathrm{c})$ corresponds to probability that one of the components in the candidate set c causes a system failure to occur before time t, which corresponds to the probability of the event $E_{j_1} \cup \cdots \cup E_{j_k}$, where E_j is the event

$$[T_j \leq t \bigcap_{\substack{p=1 \\ p \neq j}}^{m} T_p > T_j].$$

However, since the probability that more than one component fails at the same time is zero, only one of the candidate components in the c can cause of the system failure. Thus, the events $E_{j_1}, \ldots, E_{j_k}$ are mutually exclusive which means the probability of the union of these events is just the sum of their probabilities. That is,

$$\Pr[E_{j_1} \cup \cdots \cup E_{j_k}] = \sum_{j \in \mathrm{c}_i} \Pr[T_j \leq t \bigcap_{\substack{p=1 \\ p \neq j}}^{m} T_p > T_j].$$

This is equivalent to

$$\Pr[E_{j_1} \cup \cdots \cup E_{j_k}] = \sum_{j \in \mathrm{c}_i} \int_0^t f_j(s) \prod_{\substack{p=1 \\ p \neq j}}^{m} R_p(s) \mathrm{d}s = F_{\mathrm{c}}(t).$$

□

The reliability function of the system's conditional lifetime, $R_{T|\mathrm{c}}(t \mid \mathrm{c}, \boldsymbol{\theta})$, denotes the probability the system does not fail before time t given that one of the components in the candidate set c causes the system failure. It is given in the following theorem.

Theorem 2.8 (Conditional reliability function). *The conditional random variable* $T \mid \mathrm{c}$ *for the lifetime of the*

m-out-of-m system has a conditional reliability function

$$R_{T|\mathrm{c}}(t \mid \mathrm{c}, \boldsymbol{\theta}) = 1 - \int_0^t \left[\sum_{j \in \mathrm{c}} f_j(s \mid \boldsymbol{\theta}_j) \prod_{\substack{k=1 \\ k \neq j}}^{m} R_k(s \mid \boldsymbol{\theta}_k) \right] \mathrm{d}s. \tag{10}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is partitioned into a disjoint set of vectors $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m\}$ such that the lifetime of component j has a distribution parametrized by $\boldsymbol{\theta}_j$ for $j = 1, \ldots, m$.

Proof. By definition, $R_{T|\mathrm{c}}(t \mid \mathrm{c})$ is the complement of the CDF in eq. (9). Thus,

$$R_{T|\mathrm{c}}(t \mid \mathrm{c}) = 1 - F_{T|\mathrm{c}}(t \mid \mathrm{c}) = 1 - \int_0^t \left[\sum_{j \in \mathrm{c}} f_j(s) \prod_{\substack{p=1 \\ p \neq j}}^{m} R_p(s) \right] \mathrm{d}s.$$

□

The PDF of the system's conditional lifetime, $f_{T|\mathrm{c}}(t \mid \mathrm{c}, \boldsymbol{\theta})$, denotes the *relative likelihood* the system fails at time t. It is given in the following theorem.

Theorem 2.9 (Conditional PDF). *The conditional random variable $T \mid \mathrm{c}$ for the lifetime of the m-out-of-m system has a conditional PDF*

$$f_{T|\mathrm{c}}(t \mid \mathrm{c}, \boldsymbol{\theta}) = \sum_{j \in \mathrm{c}} f_j(t \mid \boldsymbol{\theta}_j) \prod_{\substack{k=1 \\ k \neq j}}^{m} R_k(t \mid \boldsymbol{\theta}_k). \tag{11}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is partitioned into a disjoint set of vectors $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m\}$ such that the lifetime of component j has a distribution parametrized by $\boldsymbol{\theta}_j$ for $j = 1, \ldots, m$.

Proof. The PDF $f_{T|\mathrm{c}}(t \mid \mathrm{c})$ is the derivative of the CDF $F_{T|\mathrm{c}}(t \mid \mathrm{c})$,

$$f_{T|\mathrm{c}}(t \mid \mathrm{c}) = \frac{\mathrm{d}}{\mathrm{d}t} F_{T|\mathrm{c}}(t \mid \mathrm{c}) = \frac{\mathrm{d}}{\mathrm{d}t} \int_0^t \left[\sum_{j \in \mathrm{c}} f_j(s) \prod_{\substack{p=1 \\ p \neq j}}^{m} R_p(s) \right] \mathrm{d}s.$$

By first fundamental theorem of calculus, this is

$$f_{T|\mathrm{c}}(t \mid \mathrm{c}) = \sum_{j \in \mathrm{c}} f_j(t) \prod_{\substack{p=1 \\ p \neq j}}^{m} R_p(t).$$

□

2.4 Random samples and likelihood functions

Random sample of system failure times The random variable T has a lifetime distribution $f_T(t \mid \boldsymbol{\theta})$. Thus, a sample of n i.i.d. random variables, $T_1, \ldots, T_n$, has a joint density $f_{T_1,\ldots,T_n}(t_1, \ldots, t_n \mid \boldsymbol{\theta})$. Since the random variables are independent, the joint density is the product of its density evaluated at each point in the sample. That is,

$$f_{T_1,\ldots,T_n}(t_1, \ldots, t_n \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f_T(t_i \mid \boldsymbol{\theta}).$$

Fixing the values of the random variables, $T_1 = t_1, \ldots, T_n = t_n$, and allowing the parameter $\boldsymbol{\theta}$ of the joint density to change, we have what is called the likelihood function.

Definition 2 (Likelihood function)**.** *The joint density of an i.i.d. random sample $T_1, \ldots, T_n$ evaluated at $t_1, \ldots, t_n$ with respect to parameter $\boldsymbol{\theta}$ is the* likelihood function *and is defined as*

$$L(\boldsymbol{\theta} \mid t_1, \ldots, t_n) = \prod_{i=1}^{n} f_T(t_i \mid \boldsymbol{\theta}), \tag{12}$$

where $f_T(\cdot)$ is the lifetime PDF of the system given in eq. (7)*.*

The likelihood function represents the *relative likelihood* of observing the given i.i.d. sample $(T_1 = t_1, \ldots, T_n = t_n)$ with respect to parameter $\boldsymbol{\theta}$.

Random sample of conditional system failure times The conditional random variable $T \mid \mathrm{c}$ has a conditional lifetime distribution $f_{T|\mathrm{c}}(t \mid \mathrm{c}, \boldsymbol{\theta})$. Thus, a sample of n i.i.d. random variables, $T_1 \mid \mathrm{c}_1, \ldots, T_n \mid \mathrm{c}_n$ has a joint conditional density $f_{T_1,\ldots,T_n|\mathrm{c}_1,\ldots,\mathrm{c}_n}(t_1, \ldots, t_n \mid \mathrm{c}_1, \ldots, \mathrm{c}_n, \boldsymbol{\theta})$. Since the conditional random variables are independent, the joint conditional density is the product of its conditional density evaluated at each point in the sample. That is,

$$f_{T_1,\ldots,T_n|\mathrm{c}_1,\ldots,\mathrm{c}_n}(t_1, \ldots, t_n \mid \mathrm{c}_1, \ldots, \mathrm{c}_n, \boldsymbol{\theta}) = \prod_{i=1}^{n} f_{T_i|\mathrm{c}_i}(t_i \mid \mathrm{c}_i, \boldsymbol{\theta}).$$

We denote the sample of conditional random variables, $T_1 \mid \mathrm{c}_1, \ldots, T_n \mid \mathrm{c}_n$, the masked system failure sample.

Definition 3 (Masked system failure sample)**.** *A masked system failure sample of size n, denoted by $\mathbb{F}_n$, is a sample of n system lifetimes in which it is given that the i^{th} system failure with lifetime t_i was caused by*

a component in a corresponding set of candidate components, denoted by c_i*. That is,*

$$\mathbb{F}_n = \{T_1 = t_1, \ldots, T_n = t_n \mid \mathrm{c}_1, \ldots, \mathrm{c}_n\}.$$

Fixing the values of the random variables, $T_1 = t_1 \mid \mathrm{c}_1, \ldots, T_n = t_n \mid \mathit{cand}_n$, and allowing the parameter $\boldsymbol{\theta}$ of the joint conditional density to change, we have what is called the conditional likelihood function.

Definition 4. *[Conditional likelihood function] The joint conditional density function of an i.i.d.* $\mathbb{F}_n$ *sample evaluated at* $t_1, \ldots, t_n \mid \mathrm{c}_1, \ldots, \mathrm{c}_n$ *is*

$$\mathcal{L}(\boldsymbol{\theta} \mid \mathbb{F}_n) = \prod_{i=1}^{n} f_{T|\mathrm{c}_i}(t_i \mid \mathrm{c}_i, \boldsymbol{\theta}), \tag{13}$$

where $f_{T|\mathrm{c}}(\cdot)$ *is the conditional lifetime PDF of the system given in* **??**.

2.5 Point estimation

Suppose component j has a lifetime distributed $T_j \sim f_j(t \mid \boldsymbol{\theta}_j{}^\star)$, where $\boldsymbol{\theta}_j{}^\star$ is the unknown true parameter value the lifetime of component j for $j = 1, \ldots, m$.

We are interested in characteristics about the population that are functions of the lifetime distribution of one or more of the components in the m-out-of-m system. For example, we may be interested in the expected failure time of component j. This characteristic has a true parameter value $\mathbb{E}[T_j] = \int_0^\infty t \cdot f_j(s \mid \boldsymbol{\theta}_j{}^\star)\mathrm{d}s$.

Remark. If the characteristic of interest is unrelated to individual component lifetimes, then all else being equal we prefer simpler models that do not model component lifetimes.

Since the characteristics of interest are functions of the lifetime distributions of the components, we are interested in estimating the lifetime distributions of the components. Given an assumed distribution family, $T \sim f_T(t \mid \boldsymbol{\theta})$, we may use the information in an $\mathbb{F}_n$ sample to select *plausible* estimates of the unknown true parameter value $\boldsymbol{\theta}^\star$.[3]

Properties of estimators Let our estimator of $\boldsymbol{\theta}^\star$ be $\hat{\boldsymbol{\theta}}_n = \psi\left(T_1, \ldots, T_n, \mathbb{C}_1, \ldots, \mathbb{C}_n\right)$, where ψ is a statistic since it does not depend on $\boldsymbol{\theta}^\star$. If $\mathbb{F}_n = (t_1, \ldots, t_n, \mathrm{c}_1, \ldots, \mathrm{c}_n)$ is an observed sample, then $\hat{\boldsymbol{\theta}}_n = \psi(\mathbb{F}_n)$ is our statistical estimate of $\boldsymbol{\theta}^\star$.

[3]Note that if we only have a sample of system lifetimes and the m components have lifetime distributions drawn from the same parametrized distribution family, then there is not enough information in the sample to statistically model the component lifetimes.

The estimator $\hat{\boldsymbol{\theta}}_n$ is a function of random variables $T_1, \ldots, T_n, \mathbb{C}_1, \ldots, \mathbb{C}_n$, therefore $\hat{\boldsymbol{\theta}}_n$ is a random vector. As a random vector, $\hat{\boldsymbol{\theta}}_n$ will have a distribution. All else being equal, we prefer estimators which vary only *slightly* from sample to sample with a *central tendency* around the true parameter value $\boldsymbol{\theta}^\star$. That is, we prefer unbiased estimators ($\mathbb{E}[\hat{\boldsymbol{\theta}}_n] = \boldsymbol{\theta}^\star$) with small variance.

We are often faced with a trade-off between bias and variance, e.g., the minimum-variance unbiased estimator may have a much larger variance than a biased estimator that has an expected value sufficiently close to the true parameter value. A measure of estimator accuracy that considers both the bias and the variance of the estimator is the mean squared error function (MSE).

$$\mathrm{MSE}(\hat{\theta_j}) = \mathbb{E}[\hat{\theta_j} - \theta_j]^2 = \mathrm{Var}[\hat{\theta_j}] + b[\hat{\theta_j}]^2,$$

where $\hat{\theta_j}_n$ denotes the estimator of the j^{th} parameter $\theta_j{}^\star$ of the parameter vector $\boldsymbol{\theta}^\star$ and $b[\hat{\theta_j}_n] = \mathbb{E}[\hat{\theta_j}_n] - \theta_j{}^\star$ denotes the estimator's bias.

Maximum likelihood estimation Maximum likelihood estimation (MLE) is a general strategy for finding parameter estimators. Under certain regularity conditions, it has several desirable large-sample properties, three of which given in the following.

1. $\hat{\boldsymbol{\theta}}_n$ exists and is unique.
2. $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}^\star$. The distribution of $\hat{\boldsymbol{\theta}}_n$ converges in probability to the true parameter $\boldsymbol{\theta}^\star$, written $\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}^\star$, since $\lim_{n\to\infty} \Pr[|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^\star| < \boldsymbol{\epsilon}] = 1$ for every $\boldsymbol{\epsilon} > \mathbf{0}$.
3. $\hat{\boldsymbol{\theta}}_n$ is asymptotically efficient. No consistent estimator has lower asymptotic mean squared error.

These large-sample properties of MLE $\hat{\boldsymbol{\theta}}_n$ suggests that we can use maximum likelihood estimation to find an asymptotically unbiased estimator of $\boldsymbol{\theta}^\star$ which asymptotically achieves the minimum variance among all unbiased estimators when the true parameter $\boldsymbol{\theta}^\star$ can uniformly take on any value in the feasible parameter space of $f(\cdot \mid \boldsymbol{\theta} \in \boldsymbol{\Theta})$. Such an estimator is known as a uniformly minimum-variance unbiased estimator (UMVUE).

Thus, for sufficiently large sample sizes, $\hat{\boldsymbol{\theta}}_n$ varies only *slightly* from sample to sample with a *central tendency* around the true parameter value $\boldsymbol{\theta}^\star$. Note that if the large-sample assumption does not hold, this approach may produce inefficient, biased estimators.

Intuitively, maximum likelihood estimation finds a value $\hat{\boldsymbol{\theta}}_n$ in the feasible parameter space of the assumed distribution family $f(t \mid \boldsymbol{\theta} \in \boldsymbol{\Theta})$ that maximizes the *likelihood* of observing the given sample.

The objective in maximum likelihood estimation is to choose a parameter value that has the greatest *likelihood* of generating the observed sample. Specifically, we find a value $\hat{\boldsymbol{\theta}}$ that maximizes the $\mathbb{F}$-likelihood function $\mathcal{L}$.

Definition 5 (Maximum likelihood estimator)**.** *For a given $\mathbb{F}_n$ sample, the parameter value $\hat{\boldsymbol{\theta}}_n$ that maximizes the likelihood function $\mathcal{L}$ is the maximum likelihood estimator of true parameter value $\boldsymbol{\theta}^\star$. That is,*

$$\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\max}\, \mathcal{L}(\boldsymbol{\theta} \mid \mathbb{F}_n) \tag{14}$$

where $\boldsymbol{\Theta}$ is the feasible parameter space of $f(t \mid \boldsymbol{\theta} \in \boldsymbol{\Theta})$.

The logarithm function is a monotonically increasing function which means the parameter value $\hat{\boldsymbol{\theta}}_n$ that maximizes $\mathcal{L}$ also maximizes the logarithm of $\mathcal{L}$. Since the logarithm transforms products into sums, it is often easier to find the MLE by maximizing the log-likelihood instead.

Theorem 2.10. *The $\mathbb{F}_n$ log-likelihood function of an i.i.d. random sample of size n is*

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f_{T|\mathrm{c}_i}(t_i \mid \mathrm{c}_i, \boldsymbol{\theta}) \tag{15}$$

The problem of finding the MLE may be simplified if eq. (15) is differentiable and assumes a global maximum $\hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$.

Definition 6. *If eq.* (15) *is differentiable and assumes a maximum $\hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$, then the MLE $\hat{\boldsymbol{\theta}}$ is the stationary point that takes on the maximum value with respect to $\ell(\boldsymbol{\theta})$. That is,*

$$\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}\in\mathbb{S}}{\arg\max}\, \ell(\boldsymbol{\theta} \mid \mathbb{F}_n) \tag{16}$$

where $\mathbb{S} \subset \boldsymbol{\Theta}$ is the set of stationary points satisfying

$$\frac{\partial}{\partial\theta_r}\ell(\boldsymbol{\theta}) = 0, r = 1, \ldots, q$$

where q is the dimension of $\boldsymbol{\theta}$ and

$$\frac{\partial}{\partial\theta_r}\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[\frac{\partial}{\partial\theta_r} f_{T|\mathrm{c}_i}(t_i \mid \mathrm{c}_i, \boldsymbol{\theta})\right] \left[f_{T|\mathrm{c}_i}(t_i \mid \mathrm{c}_i, \boldsymbol{\theta})\right]^{-1}$$

This is one of the more popular ways to solve the MLE since it restricts the optimization to searching over a proper subset $\mathbb{S} \subset \boldsymbol{\Theta}$, which is frequently a finite set of points.

A significant aspect of statistical parameter estimation involves solving optimization problems such as eq. (16). Since there is in general no closed-form solution for solving the MLE $\hat{\boldsymbol{\theta}}$, numerical methods are typically employed. See appendix A.1 for a detailed discussion on this.

2.6 Confidence intervals

Statistics studies ways of drawing inferences about a population from incomplete and imperfect data. Experience has shown that each time an experiment based on such data is repeated, a different inference may be drawn. Therefore, it is a mistake to treat any particular inference as certain.

In our case, we are interested in estimating some characteristic, $\psi(\boldsymbol{\theta}^\star)$, of a series system with a lifetime $T \sim f_{T|\mathrm{c}}(t \mid \mathrm{c}, \boldsymbol{\theta}^\star)$ from a sample of masked system failures. However, because we are performing this estimation from an incomplete, imperfect sample, we expect that we will arrive at a different estimate of the characteristic each time the experiment is repeated. Consequently, we should treat any particular estimate as tentative and uncertain.

That is, by the invariance property of the ML estimator $\hat{\boldsymbol{\theta}}_n$, $\psi(\hat{\boldsymbol{\theta}}_n)$ is the MLE point estimator of the true value of the characteristic $\psi(\boldsymbol{\theta}^\star)$. However, since $\hat{\boldsymbol{\theta}}_n$ is a random vector, $\hat{\psi}_n$ will have a distribution, sometimes called a *sampling distribution* due to its dependence on a random sample. We are interested in the *sampling distribution* of $\hat{\psi}_n$ to facilitate the ability to calculate the probability that, given a true value ψ, the estimated value $\hat{\psi}_n$ will take on some given set of values. Thus, if we can derive its sampling distribution, then we can estimate how "close" $\hat{\psi}_n$ is to true value ψ.

As discussed in section 2.5, the MSE is one way to evaluate the accuracy of an estimator. The confidence interval is another common approach, where a $(1-\alpha)\cdot 100\%$-confidence interval estimates a minimum-width interval (L, U) that contains the true value of the characteristic, ψ, with probability $(1-\alpha)$, i.e., $\Pr[\mathrm{L} < \psi < \mathrm{U}] = 1-\alpha$. Essentially, after multiple experiments, it is expected that $(1-\alpha)\cdot 100\%$ of the confidence intervals generated will contain the true value ψ.

Frequently, the derivation of a sampling distribution is intractable. In such cases, instead of deriving the exact sampling distribution of the statistic (the estimator of the characteristic of interest), approximate techniques may be used.

2.6.1 Asymptotic confidence intervals

An obvious characteristic of interest about the population, $\psi(\boldsymbol{\theta}^\star)$, is the identity function $\psi(\boldsymbol{\theta}^\star) = \boldsymbol{\theta}^\star$. That is, we are interested in the true parametrization of the series system's lifetime distribution. A statistical estimator of this characteristic is the ML estimator, $\hat{\boldsymbol{\theta}}$, as described in section 2.5.

We are interested in finding the limiting sampling distribution of $\hat{\boldsymbol{\theta}}_n$ to facilitate the calculation of approximate confidence intervals for true parameter value $\boldsymbol{\theta}^\star$. Recall that maximum likelihood estimators have several desirable large-sample properties such as consistency and efficiency. In addition, for the purpose of approximating the *sampling distribution* of $\hat{\boldsymbol{\theta}}_n$, under certain regularity conditions another desirable large-sample property of ML estimators is their asymptotic normality. That is, as the sample size $n \to \infty$, $\hat{\boldsymbol{\theta}}_n$ converges in distribution to a multivariate normal distribution with asymptotic mean $\boldsymbol{\theta}^\star$ and asymptotic covariance matrix $\Sigma_n = \boldsymbol{\mathcal{I}}_n^{-1}(\boldsymbol{\theta}^\star)$, written $\hat{\boldsymbol{\theta}}_n \xrightarrow{d} N(\boldsymbol{\theta}^\star, \boldsymbol{\mathcal{I}}_n^{-1}(\boldsymbol{\theta}^\star))$.

The covariance matrix Σ_n of the limiting distribution is the inverse of the Fisher information matrix, $\boldsymbol{\mathcal{I}}_n(\boldsymbol{\theta}^\star)$.

Definition 7 (Information matrix)**.** *Consider an m-out-of-m system $T \sim f_{T|\mathrm{c}}(\cdot \mid \boldsymbol{\theta}^\star \in \mathbb{R}^q, \mathrm{c})$. The Fisher information matrix $\boldsymbol{\mathcal{I}}_n(\boldsymbol{\theta}^\star)$ for an $\mathbb{F}_n$ sample is a q-by-q symmetric matrix whose (i,j)-th element is the negative of the expectation of the second partial derivative of the conditional log-likelihood ℓ with respect to scalar parameters θ_i and θ_j.*

$$[\boldsymbol{\mathcal{I}}_n(\boldsymbol{\theta}^\star)]_{i,j} = -\,\mathbb{E}\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j}\ell(\boldsymbol{\theta}^\star \mid \mathbb{F}_n)\right] \tag{17}$$

Since we are estimating parameter $\boldsymbol{\theta}^\star$ its Fisher information matrix $\boldsymbol{\mathcal{I}}_n(\boldsymbol{\theta}^\star)$ is not known. However, since MLE $\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}^\star$ and $\hat{\boldsymbol{\theta}}_n \xrightarrow{d} N(\boldsymbol{\theta}^\star, \boldsymbol{\mathcal{I}}_n(\boldsymbol{\theta}^\star))$, $\hat{\boldsymbol{\theta}}_n \xrightarrow{d} N(\boldsymbol{\theta}^\star, \boldsymbol{\mathcal{I}}_n(\hat{\boldsymbol{\theta}}))$.

Definition 8. *For sufficiently large sample sizes,*

$$\hat{\boldsymbol{\theta}}_n \approx N(\boldsymbol{\theta}^\star, \boldsymbol{\mathcal{I}}_n^{-1}(\hat{\boldsymbol{\theta}})) \tag{18}$$

We can use this asymptotic normal approximation to calculate approximate confidence intervals. To estimate the confidence intervals for the parameters $\boldsymbol{\theta}^\star = (\theta_1{}^\star, \ldots, \theta_m{}^\star)$, we are only interested in the variance of each scalar parameter $\hat{\theta_j}$, which is just the diagonal of the covariance matrix.

Definition 9. *The $(1-\alpha)\cdot 100\%$ confidence interval for parameter value $\theta_j{}^\star$ is*

$$\hat{\theta}_j \pm z_{\alpha/2}\sqrt{\left[\boldsymbol{\mathcal{I}}_n^{-1}(\hat{\boldsymbol{\theta}})\right]_{j,j}} \tag{19}$$

where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal.

Note that the inverse of the information matrix is equivalent to the Cramér-Rao lower bound, the minimum possible variance for unbiased estimators of $\theta^\star$. Mathematically,

$$\mathrm{Var}(\hat{\theta}_j) \geq \left[\boldsymbol{\mathcal{I}}_n^{-1}(\boldsymbol{\theta}^\star)\right]_{j,j}$$

thus demonstrating the asymptotic efficiency of $\hat{\boldsymbol{\theta}}_n$.

2.6.2 Parametric bootstrap confidence intervals

The characteristic of interest, $\hat{\psi}$, may have a distribution that cannot readily be derived or approximated using analytical methods. However, if we could repeat an experiment many times, we could simply observe the resulting *empirical sampling distribution* of $\hat{\psi}$ and use that as an estimate of the true *sampling distribution*. This is the motivation behind the bootstrap method.

The bootstrap depends on being able to simulate the construction of random $\mathbb{F}_n$ samples. To generate $\mathbb{F}_n$ samples, we must estimate the joint density $f_{T,\mathbb{C}}(t, \mathrm{c} \mid \boldsymbol{\theta})$.

By Bayes law, $f_{T,\mathbb{C}}(t, \mathrm{c} \mid \boldsymbol{\theta}) = f_{T|\mathrm{c}}(t \mid \mathrm{c}) \cdot \mathbb{P}_{\mathbb{C}}(c)$, where $\mathbb{P}_{\mathbb{C}}(c)$ is the distribution of candidate sets and $f_{T|\mathrm{c}}(t \mid \mathrm{c})$ is the conditional lifetime PDF of the system as given in eq. (11).

If we do not know $T \sim f_T(t \mid \boldsymbol{\theta}^\star)$, we can estimate it with the asymptotically consistent MLE $f_T(t \mid \hat{\boldsymbol{\theta}}_n)$. If we do not know $\mathbb{P}_{\mathbb{C}}(c)$, we can estimate it with the empirical distribution. That is, the relative frequencies in the original $\mathbb{F}_n$ sample.

To perform the bootstrap method, first we generate r $\mathbb{F}_n$ samples each with n elements of the form $(T_i, \mathbb{C}_i)$ for $i = 1, \ldots, n$. Second, for each of the r samples, we apply the statistic $\hat{\psi}$ and add the result to the simulated sampling distribution $\mathbf{s}$. Finally, we use distribution $\mathbf{s}$ to estimate the quantiles of the true sampling distribution, e.g., the $\alpha/2$ quantile and $(1-\alpha/2)$ quantile for the $(1-\alpha)\cdot 100\%$-confidence interval.

See line 1 for a precise algorithmic description of the bootstrap method.

Data: $\mathbf{s}$ is the bootstrap sampling distribution for $\hat{\psi}$ of size r, $\mathbb{F}_n$ is original sample, $\mathbb{F}_{n,k}$ is the k^{th} bootstrap resample

$\hat{\boldsymbol{\theta}}_n \leftarrow$ ML estimator applied to $\mathbb{F}_n$;

for $k \leftarrow 1$ **to** r **do**

  for $i \leftarrow 1$ **to** n **do**

    draw a candidate set $\mathrm{c}_i = \{t_{j_1}, \ldots, t_{j_k}\}$ from $\mathbb{C} \sim f_{\mathbb{C}}(\cdot)$;

    $t_i \leftarrow \min(t_{j_1}, \ldots, t_{j_k})$;

    $\mathbb{F}_{n,k} \leftarrow \mathbb{F}_{n,k} \bigcup \{t_i, \mathrm{c}_i\}$;

  end

  $\hat{\boldsymbol{\theta}_{n,k}} \leftarrow$ ML estimator applied to resample $\mathbb{F}_{n,k}$;

  $\mathbf{s} \leftarrow \mathbf{s} \bigcup \hat{\psi}(\hat{\boldsymbol{\theta}_{n,k}})$;

end

use $\mathbf{s}$ to estimate quantiles of sampling distribution $\hat{\psi}$;

Algorithm 1: Parametric bootstrap algorithm

Before the age of computers, the bootstrap method was somewhat impractical since it uses brute-force computation to generate a simulated empirical sampling distribution. Thus, analytical or asymptotic approaches were necessary. However, with modern computer hardware, the bootstrap method's brute-force approach is quite practical and it has the advantage that it can be used to estimate the accuracy of many statistics of interest with otherwise analytically intractable distributions.

Note that the bootstrap method is not foolproof. It can perform poorly, especially if the statistic of interest is sensitive to small differences between $f_T(\cdot \mid \hat{\boldsymbol{\theta}}_n)$ and $f_T(\cdot \mid \boldsymbol{\theta}^\star)$.

2.7 Applications

Modeling the lifetime distributions of a system's components has practical applications. We have already discussed some of these applications. For example, in an earlier section we discussed estimating the failure rates of components of an m-out-of-m series system. In this section, we discuss additional applications.

2.7.1 Probability of component failure

If a system failure occurs, each component has a particular probability of being the cause. Let K be the discrete random variable[4] representing the distribution of component indexes responsible for m-out-of-m

[4]The component indexes are not numerically meaningful, thus random variable K is specifically a nominal random variable.

system failures. The support set of K is $\{1, \ldots, m\}$, where each index value uniquely picks out one of the m components.

Recall that we assume component k in the m-out-of-m system has a lifetime distribution $f_k(t \mid \boldsymbol{\theta}^\star)$ for $k = 1, \ldots, m$. Thus, the distribution of K is a function of this assumed model.

Theorem 2.11. *The probability mass function $\mathbb{P}_{\boldsymbol{\theta}^\star}(k)$ maps each component $k \in \{1, \ldots, m\}$ to the probability that component k is the cause of an m-out-of-m system failure, i.e., $\mathbb{P}_{\boldsymbol{\theta}^\star}(k) = \Pr[K = k]$.*

$$\mathbb{P}_{\boldsymbol{\theta}^\star}(k) = \int_0^\infty f_k(s \mid \boldsymbol{\theta}_k^\star) \prod_{\substack{j=1 \\ j \neq k}}^{m} R_j(s \mid \boldsymbol{\theta}_j^\star) \mathrm{d}s \tag{20}$$

where $\boldsymbol{\theta}^\star$ is the true parametrization of the lifetime distribution for the m-out-of-m system.

Proof. Note that component k causes a system failure whenever the other components survive longer than it. That is,

$$\Pr\left[0 < T_k < \infty, T_k < T_1, \ldots, T_k < T_{k-1}, T_k < T_{k+1}, \ldots, T_k < T_m\right],$$

where T_j is a random variable denoting the lifetime of component j for $j = 1, \ldots, m$. This is equal to eq. (20).

To verify that $\mathbb{P}(k)$ is a probability mass function (PMF), it must be the case that $\mathbb{P}(k) \geq 0$ for $k = 1, \ldots, m$ and $\sum_{j=1}^{m} \mathbb{P}(j) = 1$. To show that $\mathbb{P}(k) \geq 0$, by definition

$$\mathbb{P}(k) = \int_0^\infty f_k(s) \prod_{j \neq k} R_j(s) \mathrm{d}s.$$

Since $f_k(t)$ and $R_j(t)$ are non-negative, any summation or product of them must also be non-negative, and since we are integrating non-negative values over a positive interval, its definite integral must be non-negative.

To show that $\mathbb{P}(1) + \ldots + \mathbb{P}(m) = 1$, note that

$$\sum_{j=0}^{m} \mathbb{P}(j) = \sum_{j=1}^{m} \int_0^\infty f_j(s) \prod_{p \neq j} R_p(s) \mathrm{d}s.$$

Since the limits of the integral do not depend on the summation over j, we can move the summation inside the integral.

$$\sum_{j=0}^{m} \mathbb{P}(j) = \int_0^\infty \sum_{j=1}^{m} f_j(s) \prod_{p \neq j} R_p(s) \mathrm{d}s.$$

Notice that this is the integral of the PDF of the m-out-of-m system eq. (7) integrated over its entire support

set. This must be equal to 1. □

Conditional probability of component failure If a system failure is known to have occurred at a given time t, we have additional information about which component may have caused the system failure.

Given a system failure at time t, each component has a conditional probability of being the cause of the system failure. Let K_t be the conditional discrete random variable representing the distribution of component indexes responsible for m-out-of-m system failures given system failures at time t. The support set of K_t is $\{1, \ldots, m\}$, where each index value uniquely picks out one of the m components.

Theorem 2.12. *The conditional probability mass function $\mathbb{P}_{\boldsymbol{\theta}^\star}(k \mid t)$ maps each component $k \in \{1, \ldots, m\}$ to the probability that component k is the cause of the m-out-of-m system failure given a system failure at time t. That is, $\mathbb{P}_{\boldsymbol{\theta}^\star}(k \mid t) = \Pr[K = k \mid T = t]$.*

$$\mathbb{P}_{\boldsymbol{\theta}^\star}(k \mid t) = \frac{f_k(t \mid \boldsymbol{\theta}_k^\star)}{f_T(t \mid \boldsymbol{\theta}^\star)} \prod_{\substack{j=1 \\ j \neq k}}^{m} R_j(t \mid \boldsymbol{\theta}_j^\star) \tag{21}$$

where $\boldsymbol{\theta}^\star$ is the true parametrization of the lifetime distribution for the m-out-of-m system.

Proof. Let it be given that the system failure occurred at some time t' in the interval $(t, t + \Delta t)$. Observe that component k caused the system failure if it failed at time t' and the other components survived longer than component k. Thus, the conditional probability that component k caused the system failure given a system failure at time $t' \in (t, t + \Delta t)$ is

$$\Pr\left[t < T_k < t + \Delta t \cap_{j \neq k} T_j > T_k \mid t < T < t + \Delta t\right],$$

where T_j is a random variable denoting the lifetime of component j for $j = 1, \ldots, m$ and T is a random variable denoting the lifetime of a series system composed of those m components.

By Bayes law, the above conditional probability is equivalent to

$$\frac{\Pr\left[t < T_k < t + \Delta t \cap_{j \neq k} T_j > T_k\right]}{\Pr\left[t < T < t + \Delta t\right]} = \frac{\int_t^{t+\Delta t} f_k(s) \prod_{j \neq k} R_j(s) \mathrm{d}s}{\int_t^{t+\Delta t} f_T(s) \mathrm{d}s}.$$

If we let Δt be some infinitesimally small value, then $f_k(t)$, $f_T(t)$, and $R_j(t)$ for all j are approximately

constant over the interval $(t, t + \Delta t)$, therefore an approximation of the above equation is

$$\frac{\Delta t \cdot f_k(t) \prod_{j \neq k} R_j(t)}{\Delta t \cdot f_T(t)}.$$

After canceling Δt in the numerator and denominator, we arrive at

$$\frac{f_k(t) \prod_{j \neq k} R_j(t)}{f_T(t)},$$

which is equivalent to eq. (21). Also, note that since Δt is infinitesimally small, it is given that the system failure occurred at exact time t.

To verify that $\mathbb{P}(k \mid t)$ is a conditional probability, it must be the case that $\mathbb{P}(j \mid t) \geq 0$ for $j = 1 \ldots, m$ and $\sum_{j=1}^{m} \mathbb{P}(j \mid t) = 1$. By definition,

$$\sum_{k=1}^{m} \mathbb{P}(k \mid t) = \sum_{k=1}^{m} \frac{f_k(t)}{f_T(t)} \prod_{j \neq k} R_j(t).$$

Notice that since $f_k(t)$, $f_T(t)$, and $R_j(t)$ are non-negative, any summation or product of them must also be non-negative. Substituting $f_T(t)$ with its definition, we arrive at

$$\sum_{k=1}^{m} \mathbb{P}(k \mid t) = \sum_{k=1}^{m} \frac{f_k(t) \prod_{j \neq k}^{m} R_j(t)}{\sum_{r=1}^{m} f_r(t) \prod_{s \neq r} R_s(t)} = \frac{\sum_{k=1}^{m} f_k(t) \prod_{j \neq k} R_j(t)}{\sum_{k=1}^{m} f_k(t) \prod_{j \neq k} R_j(t)} = 1.$$

□

If the component failure rates, $\lambda_j(t)$ for $j = 1, \ldots, m$ are available, there is another way to calculate $\mathbb{P}_{\boldsymbol{\theta}^\star}(k \mid t)$ given in the following theorem.

Theorem 2.13 (Conditional probability of failure)**.** *The conditional probability mass function $\mathbb{P}_{\boldsymbol{\theta}^\star}(k \mid t)$ maps each component $k \in \{1, \ldots, m\}$ to the probability that component k is the cause of the m-out-of-m system failure given a system failure at time t.*

$$\mathbb{P}_{\boldsymbol{\theta}^\star}(k \mid t) = \frac{\lambda_k(t)}{\sum_{j=1}^{m} \lambda_j(t)} = \frac{\lambda_k(t)}{\lambda(t)} \tag{22}$$

where $\boldsymbol{\theta}^\star$ is the true parametrization of the lifetime distribution for the system and $\lambda_j(t)$ is the instantaneous failure rate of component j for $j = 1, \ldots, m$, and $\lambda(t)$ is the failure rate of the system.

Proof. Consider $\frac{\lambda_k(t)}{\sum_{j=1}^{m} \lambda_j(t)}$. When we substitute the definition of instantaneous failure rate, $\lambda_j(t) = f_j(t)/R_j(t)$,

into this expression we get

$$\frac{f_k(t)}{R_k(t)} \cdot \frac{1}{\frac{f_1(t)}{R_1(t)} + \ldots + \frac{f_m(t)}{R_m(t)}}.$$

To add the fractions in the right-hand-side, $f_1(t)/R_1(t)+\ldots+f_m(t)/R_m(t)$, we must make their denominators the same. That is,

$$\frac{f_k(t)}{R_k(t)} \cdot \frac{1}{\frac{[\prod_{p\neq 1} R_p(t)]f_1(t)}{[\prod_{p\neq 1} R_p]R_1(t)} + \ldots + \frac{[\prod_{p\neq m} R_p(t)]f_m(t)}{[\prod_{p\neq m} R_p]R_m(t)}} = \frac{f_k(t)}{R_k(t)} \cdot \frac{\prod_{j=1}^m R_j(t)}{\sum_{j=1}^m \prod_{p\neq j} R_p(t) f_j(t)}.$$

Dividing $\prod_{j=1}^m R_j(t)$ in the numerator by $R_k(t)$ in the denominator, we get

$$\frac{f_k(t) \prod_{j\neq k} R_j(t)}{\sum_{j=1}^m \prod_{p\neq j} R_p(t) f_j(t)}.$$

Recall that $\sum_{j=1}^m \prod_{p\neq 1} R_p(t) f_j(t)$ is just the PDF of the system's lifetime, $f_T(t)$. Performing this substitution, we get

$$\frac{f_k(t) \prod_{j\neq k} R_j(t)}{f_T(t)},$$

which is equivalent to eq. (21). □

2.7.2 Estimating probability of component failure

By the invariance property of the maximum likelihood estimator, if $\hat{\boldsymbol{\theta}}_n$ is the ML estimator of $\boldsymbol{\theta}^\star$ (as described in section 2.5), then for any population characteristic $\psi(\boldsymbol{\theta}^\star)$ its corresponding ML estimator is $\psi(\hat{\boldsymbol{\theta}}_n)$.

Consequently, if the probability that component k caused a system failure, $\mathbb{P}_{\boldsymbol{\theta}^\star}(k)$, is not known, then instead we may derive its MLE estimator, $\mathbb{P}_{\hat{\boldsymbol{\theta}}_n}(k)$. Note that, assuming the $\mathbb{F}$ sample is sufficiently large, a more straightforward estimator of $\mathbb{P}_{\boldsymbol{\theta}^\star}(k)$ is the proportion[5] of component indexes in the $\mathbb{F}$ sample equal to k. This nonparametric estimator has the additional advantage of not assuming any particular lifetime model for the system and its components, unlike the parametrized model $\mathbb{P}_{\boldsymbol{\theta}}(k)$.

If it is given that a system failure occurred at time t, then we are interested in estimating the conditional probability, $\mathbb{P}_{\boldsymbol{\theta}^\star}(k \mid t)$. In this case, the nonparametric estimator is relatively inefficient compared to the MLE estimator $\mathbb{P}_{\hat{\boldsymbol{\theta}}_n}(k \mid t)$ since the nonparametric estimator is a function of a subset of the observations in the $\mathbb{F}$ sample (only those observations which have system failure times *approximately* equal to the given time t) while the parametric estimator $\mathbb{P}_{\hat{\boldsymbol{\theta}}_n}(k \mid t)$ is a function of the entire $\mathbb{F}$ sample.

[5] The nonparametric bootstrap method may be used to estimate the confidence interval of this proportion.

As an estimator, $\mathbb{P}_{\hat{\boldsymbol{\theta}}_n}(k \mid t)$ is a random variable for each $k \in \{1, \ldots, m\}$. To measure the estimator's "accuracy," we derive a $(1-\alpha) \cdot 100\%$-confidence interval for true parameter value $\mathbb{P}_{\boldsymbol{\theta}^\star}(k \mid t)$. The asymptotically normal $(1-\alpha) \cdot 100\%$-confidence interval for each true parameter value ${\theta_j}^\star$ (see section 2.6) can be used for this purpose.

Theorem 2.14. *The $(1-\alpha) \cdot 100\%$-confidence interval for parameter $\mathbb{P}_{\boldsymbol{\theta}^\star}(k \mid t)$ is $(L_t(k), U_t(k))$*

$$L_t(k) = \min_{\boldsymbol{\theta} \in D} \mathbb{P}_{\boldsymbol{\theta}}(k \mid t)$$

$$U_t(k) = \max_{\boldsymbol{\theta} \in D} \mathbb{P}_{\boldsymbol{\theta}}(k \mid t)$$

where D is a Cartesian product of q $(1-\alpha) \cdot 100\%$-confidence intervals for the q parameters in $\boldsymbol{\theta}^\star = \left(\theta_1^\star, \ldots, \theta_q^\star\right)$.

In general, theorem 2.14 involves solving two nonlinear programming problems to determine $L_t(k)$ and $U_t(k)$. Thus, it may be preferable to use the parametric Bootstrap (see section 2.6.2) to estimate the $(1-\alpha) \cdot 100\%$-confidence interval $(L_t(k), U_t(k))$.

2.7.3 Distribution of inspection counts

If an m-out-of-m system failure occurs, then one its m components failed. To repair the system, the failed component must be repaired or replaced. If it is not known which component failed, components must be inspected until the failed component is identified.

In general, the number of component inspections required to identify the failed component is a function of the order in which components are inspected and the distribution of random variable K_t. Let $N_{t,\pi} \in \{1, \ldots, m\}$ be a discrete conditional random variable denoting the distribution of the number of inspections required to identify the component responsible for a system failure occurring at a given time t using order policy $\pi \colon \{1, \ldots, m\} \mapsto \{1, \ldots, m\}$, which is a function defining a one-to-one correspondence that maps each inspection order $n \in \{1, \ldots, m\}$ to a unique component index $j \in \{1, \ldots, m\}$. For instance, $\pi(n) = k$ denotes that component k should be the n^{th} inspection.

Suppose we have complete information about the distribution of random variable K_t. Thus, intuitively, it would seem most appropriate to inspect the components in order of decreasing probability of having caused the system failure at time t.

Theorem 2.15. *Assuming the distribution of the conditional random variable K_t is known, $\pi = \sigma$ is the*

order policy in which the components of the m-out-of-m system that failed at given time t are inspected in the order of decreasing probability $\mathbb{P}_{\boldsymbol{\theta}^\star}(\cdot \mid t)$ *of being the cause of the system failure. The discrete conditional random variable* $N_{t,\sigma}$ *has a conditional probability mass function*

$$\mathbb{P}_\sigma(n \mid t) = \begin{cases} \mathbb{P}_{\boldsymbol{\theta}^\star}(\sigma(n) \mid t) & n \in \{1, \ldots, m\} \\ 0 & n \notin \{1, \ldots, m\} \end{cases} \tag{23}$$

where n *denotes the number of inspections required to identify the failed component and* $\mathbb{P}_{\boldsymbol{\theta}^\star}(\sigma(1) \mid t) \geq \mathbb{P}_{\boldsymbol{\theta}^\star}(\sigma(2) \mid t) \geq \ldots \geq \mathbb{P}_{\boldsymbol{\theta}^\star}(\sigma(m-1) \mid t) \geq \mathbb{P}_{\boldsymbol{\theta}^\star}(\sigma(m) \mid t)$.

Note that if the system failure time t *is not given, the distribution of the random variable* $K \sim \mathbb{P}_{\boldsymbol{\theta}^\star}(k)$ *may be used to derive the distribution of random variable* $N_\sigma \sim \mathbb{P}_\sigma(n)$*, which is independent of time* t*.*

Proof. One inspection is required to identify the failed component if the first component inspected, component $\sigma(1)$, has failed. The probability that the component $\sigma(1)$ has failed given a system failure at time t is $\Pr[K_t = 1] = \mathbb{P}_{\boldsymbol{\theta}^\star}(\sigma(1) \mid t)$. Therefore, $\Pr[N_{t,\sigma} = 1] = \mathbb{P}_{\boldsymbol{\theta}^\star}(\sigma(1) \mid t)$ computes probability that one inspection is required.

Two inspections are required to identify the failed component if the second component inspected, component $\sigma(2)$, has failed. The probability that the component $\sigma(2)$ has failed given a system failure at time t is $\Pr[K_t = \sigma(2)] = \mathbb{P}_{\boldsymbol{\theta}^\star}(\sigma(2) \mid t)$. Therefore, $\Pr[N_{t,\sigma} = 2] = \mathbb{P}_{\boldsymbol{\theta}^\star}(\sigma(2) \mid t)$ computes the probability that two inspections are required.

In general, n inspections are required to identify the failed component if the n^{th} component inspected, component $\sigma(n)$, has failed. The probability that the component $\sigma(n)$ has failed given a system failure at time t is $\Pr[K_t = \sigma(n)] = \mathbb{P}_{\boldsymbol{\theta}^\star}(\sigma(n) \mid t)$. Therefore, $\Pr[N_{t,\sigma} = n] = \mathbb{P}_{\boldsymbol{\theta}^\star}(\sigma(n) \mid t)$ computes the probability that n inspections are required. □

An application of knowing the distribution of random variable $N_{t,\pi}$ is computing the expected number of component inspections required to identify the failed component given a system failure at time t, $\mathbb{E}[N_{t,\pi}]$. All else being equal, we prefer to inspect components using an order policy that minimizes this expectation, denoted optimal order policy π^*. Mathematically,

$$\pi^* = \arg\min_\pi \mathbb{E}[N_{t,\pi}] = \arg\min_\pi \sum_{n=1}^{m} n \cdot \mathbb{P}_\pi(n) = \arg\min_\pi \sum_{n=1}^{m} n \cdot \mathbb{P}_{\boldsymbol{\theta}^\star}(\pi(n) \mid t).$$

Theorem 2.16. *If we know the distribution of the discrete random variable,* $K_t \sim \mathbb{P}_{\boldsymbol{\theta}^\star}(k \mid t)$*, then the*

optimal order policy is $\pi^* = \sigma$. That is, the optimal order policy is inspecting components in the order of highest probability of component failure to lowest probability of component failure.

Consequently, the optimal expected number of inspections given a system failure at time t and a discrete random variable $K_t \sim \mathbb{P}_{\boldsymbol{\theta}^\star}(k)$ is

$$\mathbb{E}[N_{t,\sigma}] = \sum_{n=1}^{m} n \cdot \mathbb{P}_{\boldsymbol{\theta}^\star}(\sigma(n) \mid t). \tag{24}$$

Proof. Suppose $\mathbb{P}_{\boldsymbol{\theta}^\star}(k \mid t) < \mathbb{P}_{\boldsymbol{\theta}^\star}(j \mid t)$ and we inspect component k at order position r and component j at order position $r+n$, $n > 0$. Then, these two inspections have an expectation of $r\mathbb{P}_{\boldsymbol{\theta}^\star}(k \mid t) + (r+n)\mathbb{P}_{\boldsymbol{\theta}^\star}(j \mid t) = r\left(\mathbb{P}_{\boldsymbol{\theta}^\star}(k \mid t) + \mathbb{P}_{\boldsymbol{\theta}^\star}(j \mid t)\right) + n\mathbb{P}_{\boldsymbol{\theta}^\star}(j \mid t)$. However, $n\mathbb{P}_{\boldsymbol{\theta}^\star}(j \mid t)$ must be greater than $n\mathbb{P}_{\boldsymbol{\theta}^\star}(k \mid t)$ since $\mathbb{P}_{\boldsymbol{\theta}^\star}(j \mid t) > \mathbb{P}_{\boldsymbol{\theta}^\star}(k \mid t)$, therefore ordering component j before component k would result in a smaller expectation. Since component k and component j correspond to arbitrary components, this argument applies to every pair of components, and thus the minimum expectation is achieved by order policy σ. □

The lower-bound for the optimal order policy, $\mathbb{LB}$, is the optimal order policy σ for a distribution for the random variable, K_t, satisfying

$$\mathbb{LB} = \min_{\boldsymbol{\theta} \in \Theta} \sum_{n=1}^{m} n \cdot \mathbb{P}_{\boldsymbol{\theta}}(\sigma(n) \mid t)$$

which occurs when $\mathbb{P}_{\boldsymbol{\theta}}(\cdot \mid t)$ denotes a degenerate discrete random variable in which all of the probability is concentrated at a single component index. Thus,

$$\mathbb{LB} = 1.$$

The upper-bound for the optimal order policy, $\mathbb{UB}$, is the optimal order policy σ for a distribution for the random variable, K_t, satisfying

$$\mathbb{UB} = \max_{\boldsymbol{\theta} \in \Theta} \sum_{n=1}^{m} n \cdot \mathbb{P}_{\boldsymbol{\theta}}(\pi(n) \mid t)$$

which occurs when $\mathbb{P}_{\boldsymbol{\theta}}(\cdot \mid t)$ denotes a discrete uniform distribution, i.e., $K_t \sim \mathbb{P}(k \mid t) = \frac{1}{m}$ for $k = 1, \ldots, m$. Thus,

$$\mathbb{UB} = \sum_{n=1}^{m} n\frac{1}{m} = \frac{1}{m}\left(1 + \ldots + m\right) = \frac{m+1}{2}.$$

Thus, given these upper and lower bounds, it must be the case that the optimal order policy σ must result in an expectation $\mathbb{E}[N_{t,\sigma}] \in \left[1, \frac{m+1}{2}\right]$.

In general, the greater the entropy of the distribution of K_t, the greater the expected number of inspec-

tions $\mathbb{E}[N_{t,\sigma}]$, where the entropy is defined as

$$H(K_t) = -\sum_{j=1}^{m} \mathbb{P}_{\boldsymbol{\theta}^\star}(j \mid t) \log \mathbb{P}_{\boldsymbol{\theta}^\star}(j \mid t).$$

The maximum entropy occurs when K_t is uniformly distributed and the minimum entropy is occurs when K_t is degenerate.

Finally, observe that if all components have equal probability of being the cause of the system failure, then we have no reason to prefer inspecting any particular component before any other particular component. Consequently, we may as well choose a random order policy, $\pi^* = \tau$. Thus, $\mathbb{E}[N_{t,\tau}] = \frac{m+1}{2}$. Note that this is equivalent to the situation where we have no information about the distribution of the random variable K_t, thus if $\mathbb{E}[N_{t,\tau}] = \frac{m+1}{2}$ is acceptable, then no statistical study needs to be conducted.

2.7.4 Estimating distribution of inspection counts

If the distribution of the required number of inspections, $N_{t,\sigma} \sim \mathbb{P}_\sigma(n \mid t)$, is not known, then instead we may derive its ML estimator, $N_{t,\hat{\sigma}} \sim \mathbb{P}_{\hat{\sigma}}(n \mid t, \hat{\boldsymbol{\theta}})$.

Theorem 2.17. *The* $(1-\alpha)\cdot 100\%$*-confidence interval for parameter* $\mathbb{E}[N_{t,\sigma}]$ *is* (L_t, U_t)

$$\begin{aligned} L_t &= \min_{\boldsymbol{\theta} \in D} \sum_{n=1}^{m} n \mathbb{P}_{\boldsymbol{\theta}}(\sigma(n) \mid t) \\ U_t &= \max_{\boldsymbol{\theta} \in D} \sum_{n=1}^{m} n \mathbb{P}_{\boldsymbol{\theta}}(\sigma(n) \mid t) \end{aligned} \tag{25}$$

where D *is a Cartesian product of* q $(1-\alpha)\cdot 100\%$*-confidence intervals for the* q *parameters in* $\boldsymbol{\theta}^\star = (\theta_1^\star, \ldots, \theta_q^\star)$.

2.7.5 Utility theory

Finding an optimal order policy π^* involves finding an optimal decision (the order in which components are inspected) under uncertainty given certain preferences. In section 2.7.3, we discussed the optimal order policy σ to achieve the minimum expected number of component inspections given a system failure at time t. That is, the only preference was to minimize the expected number of component inspections required to identify the failed component. More generally, we can compute the optimal order policy π^* that maximizes the *expected utility*, where utility is characterized by a utility function u that quantifies how preferable each possible outcome is.

Definition 10. *The expected utility of order policy π given a discrete random variable $K_t \sim \mathbb{P}_{\boldsymbol{\theta}^\star}(k)$ and a utility function* u *is*

$$\mathcal{U}[\pi \mid \mathrm{u}] = \sum_{n=1}^{m} \mathrm{u}(\pi(n), n) \cdot \mathbb{P}_{\boldsymbol{\theta}^\star}(\pi(n) \mid t) \tag{26}$$

where u *is a utility function that takes as its first argument a component index and as its second argument an inspection order and returns the value of assigning the given inspection order to the given component index.*

To find the maximum expected utility, the objective is to find the order policy π^* that maximizes eq. (26).

Definition 11. *The optimal order policy π^* given utility function* u *is*

$$\pi^* = \arg\max_{\pi} \mathcal{U}[\pi \mid \mathrm{u}]. \tag{27}$$

In section 2.7.3, the utility function is implicitly $\mathrm{h}(\cdot, n) = -n$. Thus, $\mathbb{E}[N_{t,\sigma}] = -\mathcal{U}[\pi^* \mid \mathrm{h}]$. However, other utility functions are possible. For instance, different components may have different inspection costs, and thus given two components, all things else being equal the component that is less costly to inspect should be inspected before the component that is more costly to inspect.

Another example of a utility function that is similar to h except that it also incorporates the insight that, if the first $m-1$ inspected components have not failed, then we may conclude the last remaining non-inspected component must have failed without needing to inspect it. Mathematically,

$$\mathrm{g}(n) = \begin{cases} -n & n \in \{1, \ldots, m-1\} \\ -m+1 & n = m. \end{cases}$$

In other words, let the cost of inspecting the failed component on the $m-1$-th inspection be the same as the cost of not inspecting the failed component on the first $m-1$ inspections. Thus,

$$-\mathcal{U}[\pi^* = \sigma \mid \mathrm{h}] = \left[\sum_{n=1}^{m} n \cdot \mathbb{P}_{\boldsymbol{\theta}^\star}(\sigma(n) \mid t)\right] - \mathbb{P}_{\boldsymbol{\theta}^\star}(\sigma(m) \mid t).$$

3 Exponential series systems

3.1 Component lifetimes

The Exponential distribution is one of the more popular lifetime distributions. A significant reason for its popularity undoubtedly has to do with its analytical tractability. For instance, it is the only continuous distribution with the *no-memory* property,

$$\Pr[T > t + h \mid T > t] = \Pr[T > h] = R(t),$$

which implies that components with exponentially distributed lifetimes do not wear out over time, i.e., the reliability of the component is independent of its current age. It is often reasonable to assume various electronic devices randomly fail rather than wear out.

Definition 12. *Consider a series system composed of m components with exponentially distributed lifetimes, $T_j \sim \text{EXP}(\lambda_j)$ for $j = 1, \ldots, m$. The lifetime of component j has reliability, CDF, and PDF functions given in the following:*

$$R_j(t \mid \lambda_j) = \begin{cases} \mathrm{e}^{-\lambda_j t} & t \geq 0 \\ 1 & t < 0, \end{cases} \tag{28}$$

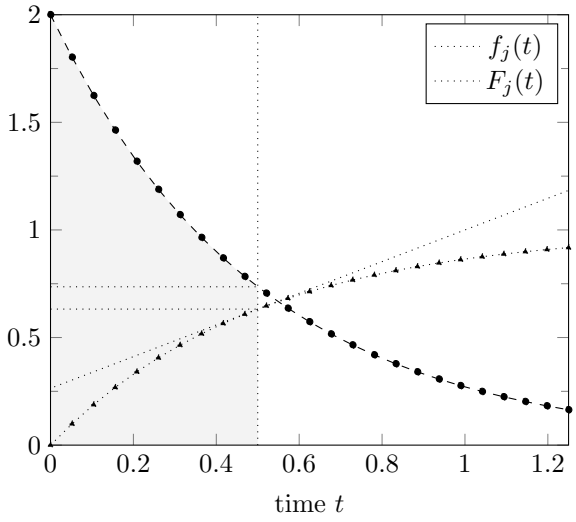
$$F_j(t \mid \lambda_j) = \begin{cases} 1 - \mathrm{e}^{-\lambda_j t} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{29}$$

$$f_j(t \mid \lambda_j) = \begin{cases} \lambda_j \, \mathrm{e}^{-\lambda_j t} & t \geq 0 \\ 0 & t < 0. \end{cases} \tag{30}$$

See Figure 1 for a visualization of the distribution functions Equations (29) and (30) where we show their respective graphs and mathematical relationship for an Exponential distribution with failure rate $\lambda = 2$. Specifically, the shaded region under the curve of the PDF $f_j(t)$ represents the probability that component j fails before its expected lifetime $\mathbb{E}[T_j] = \mu_j = 0.5$. This shaded region has an area equal to the height of the CDF at $t = \mu_j = 0.5$, $F_j(t = 0.5) \approx 0.63$. The line tangent to the CDF at $t = 0.5$ has a slope equal to the height of the PDF at $t = 0.5$, $f_j(0.5) \approx 0.74$.

Recall that exponentially distributed lifetimes are an appropriate model when components do not wear out over time. A component that does not wear out has a constant failure rate. Indeed, the Exponential

Figure 1: Exponentially distributed lifetime
with failure rate $\lambda_j = 2$



distribution is the only distribution with a constant failure rate.

Theorem 3.1. *A component j with an exponentially distributed lifetime has a constant failure rate*

$$\lambda_j(t) = \lambda_j. \tag{31}$$

Proof. According to Equation (4), $\lambda_j(t) = \frac{f_j(t)}{R_j(t)}$. Substituting the Exponential distribution functions for $f_j(t)$ and $R_j(t)$, we get

$$\lambda_j(t) = \frac{\lambda_j \, \mathrm{e}^{-\lambda_j t}}{\mathrm{e}^{-\lambda_j t}} = \lambda_j.$$

□

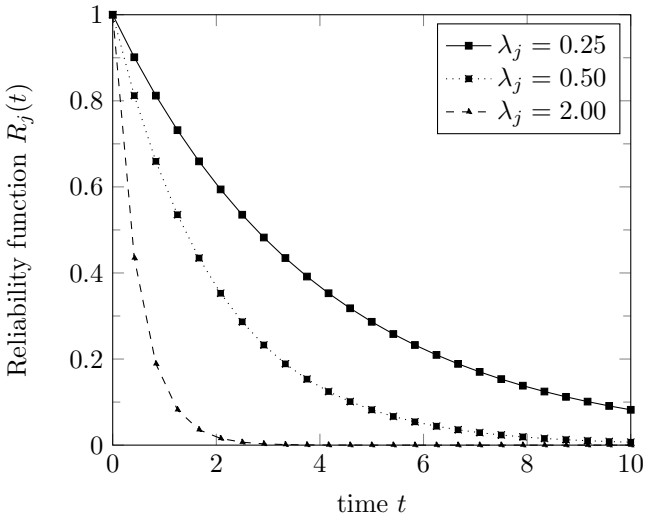
Figure 2 shows that as the failure rate λ_j increases, the reliability $R_j(t) = \Pr[T_j > t]$ converges more rapidly to the asymptotic limit of zero probability of survival.

3.2 System lifetime

The system's lifetime depends on the lifetime of its components.

Theorem 3.2. *Consider a series system composed of m components with exponentially distributed lifetimes, $T_j \sim \mathrm{EXP}(\lambda_j)$ for $j = 1, \ldots, m$. The system has a lifetime $T = \min(T_1, \ldots, T_m)$ with reliability, CDF, and*

Figure 2: Reliability function $R_j(t\cdot)$
with respect to failure rate λ_j



PDF functions given in the following.

$$R(t \mid \boldsymbol{\lambda}) = \begin{cases} \mathrm{e}^{-\left[\sum_{j=1}^{m} \lambda_j\right] t} & t \geq 0 \\ 1 & t < 0, \end{cases} \tag{32}$$

$$F(t \mid \boldsymbol{\lambda}) = \begin{cases} 1 - \mathrm{e}^{-\left[\sum_{j=1}^{m} \lambda_j)\right] t} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{33}$$

$$f(t \mid \boldsymbol{\lambda}) = \begin{cases} \left[\sum_{j=1}^{m} \lambda_j\right] \mathrm{e}^{-\left[\sum_{j=1}^{m} \lambda_j\right] t} & t \geq 0 \\ 0 & t < 0. \end{cases} \tag{34}$$

Proof. To prove Equation (32), note that Theorem 2.2 shows that $R(t) = \prod_{j=1}^{m} R_j(t)$. It is straightforward to plug in the component reliability functions $R_j(t)$ for $j = 1, \ldots, m$ to derive the result.

To prove Equation (33), by definition the CDF is the complement of the reliability function, i.e., $F(t) = 1 - R(t)$.

To prove Equation (34), note that Theorem 2.4 shows that $f(t) = \sum_{j=1}^{m} \left(f_j(t) \prod_{k \neq j} R_k(t) \right)$. Substituting

in the exponential distribution functions for $f_j(t)$ and $R_k(t)$, we get

$$\begin{aligned} f(t) &= \sum_{j=1}^{m} \left(\lambda_j \, \mathrm{e}^{-\lambda_j t} \prod_{k \neq j} \mathrm{e}^{-\lambda_k t} \right) \\ &= \sum_{j=1}^{m} \left(\lambda_j \, \mathrm{e}^{-\lambda_j t} \, \mathrm{e}^{-\sum_{k \neq j} \lambda_k t} \right) \\ &= \sum_{j=1}^{m} \left(\lambda_j \, \mathrm{e}^{-\sum_{k=1}^{m} \lambda_k t} \right) \\ &= \left(\sum_{j=1}^{m} \lambda_j \right) \mathrm{e}^{-\sum_{j=1}^{m} \lambda_j t} \end{aligned}$$

□

Examining the CDF function of the system's lifetime Equation (33), we see that this function fits the pattern of the CDF for an exponential distribution.

Theorem 3.3. *A series system composed of m components with exponentially distributed lifetimes, $T_j \sim EXP(\lambda_j)$ for $j = 1, \ldots, m$, has an exponentially distributed lifetime*

$$T \sim EXP\big(\lambda = \sum_{j=1}^{m} \lambda_j\big) \tag{35}$$

That is, for a series system exclusively composed of exponentially distributed component lifetimes, the system lifetime is also exponentially distributed with a failure rate equal to the sum of the failure rates of its components.

Theorem 3.3 is another example of the exponential distribution's analytical tractability.

3.3 Point estimation

If the parameters of $T_j \sim \mathrm{EXP}(\lambda_j)$ for $j = 1, \ldots, m$ are not known but we have a masked system failure sample $\mathbb{F}$, then this information may be used to find the maximum likelihood estimate of the individual failure rates of the m components, $\hat{\boldsymbol{\lambda}} = \left(\hat{\lambda}_1, \ldots, \hat{\lambda}_m \right)$.

Remark. The characteristic of interest may be unrelated to the failure rates of the m individual components. In this case, since an m-out-of-m system composed of exponentially distributed component lifetimes has an exponentially distributed system lifetime, the model can be straightforwardly simplified. Model the entire system as a 1-out-of-1 system, and thus only the system's failure rate, $\lambda^\star$, needs to be estimated (rather

than the m failure rates of the components). Estimating $m-1$ fewer parameters simplifies the estimation problem, i.e., smaller samples are needed for a given standard error target. In general, the characteristic of interest may only apply to a subset of the m components, $\{j_1, \ldots, j_k\}$, where $k < m$ and $j_i \in \{1, \ldots, m\}$ for $i = 1, \ldots, k$. In this case, we can simplify the model by only estimating the failure rates for components $j_1, \ldots, j_k$ and model the remaining $m-k$ components as a single component with a single failure rate. This reduces the number of failure rate parameters that need to be estimated from m to $k+1$.

Likelihood To find the maximum likelihood estimator, the $\mathbb{F}_n$ likelihood function $\mathcal{L}$ and its log-likelihood function ℓ must be derived.

Theorem 3.4. *The likelihood function $\mathcal{L}$ with respect to $\boldsymbol{\lambda}$ and $\mathbb{F}_n$ is*

$$\mathcal{L}(\boldsymbol{\lambda} \mid \mathbb{F}_n) = \exp\left[-\left(\sum_{j=1}^{m} \lambda_j\right)\left(\sum_{i=1}^{n} t_i\right)\right]\left[\prod_{i=1}^{n}\left(\sum_{j \in \mathrm{c}_i} \lambda_j\right)\right]. \tag{36}$$

Proof. By **??**, the likelihood function is

$$\mathcal{L} = \prod_{i=1}^{n}\left[\sum_{j \in \mathrm{c}_i}\left\{f_j(t_i \mid \lambda_j) \prod_{\substack{p=1 \\ p \neq j}}^{m} R_p(t_i \mid \lambda_p)\right\}\right].$$

Substituting in the functions for the Exponential distribution, we get

$$\mathcal{L} = \prod_{i=1}^{n}\left[\sum_{j \in \mathrm{c}_i}\left\{\lambda_j \exp(-\lambda_j t_i) \prod_{\substack{p=1 \\ p \neq j}}^{m} \exp(-\lambda_j t_i)\right\}\right].$$

Converting the product of the exponential functions into an exponential of the sum, we get

$$\mathcal{L} = \prod_{i=1}^{n}\left[\sum_{j \in \mathrm{c}_i} \lambda_j \exp\left(-t_i \sum_{p=1}^{m} \lambda_p\right)\right].$$

Since the exponential function does not depend on the index j we may pull it out of its sum and get

$$\begin{aligned}\mathcal{L} &= \prod_{i=1}^{n}\left[\exp\left(-t_i\sum_{p=1}^{m}\lambda_p\right)\sum_{j\in \mathrm{c}_i}\lambda_j\right] \\ &= \prod_{i=1}^{n}\left[\exp\left(-t_i\sum_{p=1}^{m}\lambda_p\right)\right]\prod_{i=1}^{n}\left[\sum_{j\in \mathrm{c}_i}\lambda_j\right].\end{aligned}$$

Converting the product of the exponential functions into an exponential of the sum, we get

$$\mathcal{L} = \exp\left[-\sum_{i=1}^{n}\left(t_i\sum_{p=1}^{m}\lambda_p\right)\right]\prod_{i=1}^{n}\left[\sum_{j\in \mathrm{c}_i}\lambda_j\right].$$

Since the sum over $\boldsymbol{\lambda}$ in the exponential function does not depend on index i, we can pull it over its sum and get

$$\mathcal{L} = \exp\left[-\left(\sum_{p=1}^{m}\lambda_p\right)\left(\sum_{i=1}^{n}t_i\right)\right]\prod_{i=1}^{n}\left[\sum_{j\in \mathrm{c}_i}\lambda_j\right].$$

□

The log-likelihood ℓ is the logarithm of $\mathcal{L}$.

Theorem 3.5. *The log-likelihood function ℓ with respect to $\boldsymbol{\lambda}$ and $\mathbb{F}_n$ is*

$$\ell(\boldsymbol{\lambda} \mid \mathbb{F}_n) = \sum_{i=1}^{n}\log\left[\sum_{j\in \mathrm{c}_i}\lambda_j\right] - \left[\sum_{i=1}^{n}t_i\right]\left[\sum_{j=1}^{m}\lambda_j\right] \tag{37}$$

where $\lambda_j > 0$ for $j = 1, \ldots, m$ is the failure rate of the j^{th} exponentially distributed component.

Proof. Taking the logarithm of eq. (36), we get

$$\ell = \log\left\{\exp\left[-\left(\sum_{j=1}^{m}\lambda_j\right)\left(\sum_{i=1}^{n}t_i\right)\right]\left[\prod_{i=1}^{n}\left(\sum_{j\in \mathrm{c}_i}\lambda_j\right)\right]\right\}.$$

The logarithm of a product $\log(A \cdot B)$ is the sum of the logarithms $\log A + \log B$, so the above is equivalent to

$$\ell = \log\left\{\exp\left[-\left(\sum_{j=1}^{m}\lambda_j\right)\left(\sum_{i=1}^{n}t_i\right)\right]\right\} + \log\left[\prod_{i=1}^{n}\left(\sum_{j\in \mathrm{c}_i}\lambda_j\right)\right].$$

Since the exponential function and the logarithm function are inverses, the above simplifies to

$$\ell = -\left[\sum_{j=1}^{m} \lambda_j\right]\left[\sum_{i=1}^{n} t_i\right] + \sum_{i=1}^{n} \log\left[\sum_{j \in \mathrm{c}_i} \lambda_j\right].$$

□

Minimal complete sufficient statistic We can simplify eq. (36) by finding a sufficient statistic for $\hat{\boldsymbol{\lambda}}_n$ that is simpler than the $\mathbb{F}_n$ sample.

Theorem 3.6. *A minimal complete sufficient statistic for $\hat{\boldsymbol{\lambda}}_n$ is $\bar{t}$ and $\mathtt{n}_\alpha$. The statistic $\bar{t} = \sum_{i=1}^{n} t_i$ denotes the mean of the system lifetimes in the $\mathbb{F}_n$ sample. The statistic $\mathtt{n}_\alpha$ denotes the number of candidate sets in the $\mathbb{F}_n$ sample that equals the specified set α for each $\alpha \in \mathcal{P}(\{1, \ldots, m\})$, where $\mathcal{P}(\{1, \ldots, m\})$ is the power set of the set $\{1, \ldots, m\}$.*

Proof. Furthermore, note that the $\mathbb{F}_n$ sample contains sets of candidate components of the form $\mathrm{c}_i \in \mathcal{P}$ for $i = 1, \ldots, n$, where $\mathcal{P}$ is the power set of $\{1, \ldots, m\}$, e.g., $\mathcal{P}(\{1,2,3\}) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$. Thus, a candidate set of the form $\alpha \in \mathcal{P}$ may occur multiple times in the $\mathbb{F}_n$ sample, and for each such occurrence we are multiplying the likelihood function $\mathcal{L}$ by $\sum_{j \in \alpha} \lambda_j$.

By the Fisher–Neyman factorization theorem, since $\mathcal{L}$ can be factored as $\mathcal{L} = h(T)g(\mathtt{n}, \bar{t})$, $\mathtt{n}$ and $\bar{t}$ are sufficient statistics for $\boldsymbol{\lambda}$. No other information about the sample needs to be stored. Indeed, the form of the PDF is an exponential family, so these are complete sufficient statistics. By the Lehmann–Scheffé theorem, an unbiased estimator of $\boldsymbol{\lambda}^\star$ that is a function of a complete sufficient statistic is the UMVUE estimator of $\boldsymbol{\lambda}^\star$. □

Maximizing the log-likelihood By **??**, we wish to find the MLE $\hat{\boldsymbol{\lambda}}_n$ by maximizing ℓ with respect to $\boldsymbol{\theta}$.

Definition 13. *The MLE estimator $\hat{\boldsymbol{\lambda}}_n$ is*

$$\hat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}} \ell(\boldsymbol{\lambda} \mid \mathbb{F}_n) \tag{38}$$

This can be found by finding the stationary points and letting $\hat{\boldsymbol{\lambda}}$ be the stationary point with the maximum value.

$$0 = -n\bar{t} + \sum_{\{i \mid i \in N \cap k \in \mathrm{c}_i\}} \left[\left(\sum_{j \in \mathrm{c}_i} \lambda_j\right)^{-1}\right]; k = 1, \ldots, m \tag{39}$$

The stationary points for $\lambda_1, \ldots, \lambda_m$ are

$$\begin{aligned} 0 &= \frac{\partial \ell}{\partial \lambda_k}; k = 1, \ldots, m \\ &= -n\bar{t} + \sum_{\{\alpha | \alpha \in \mathcal{P}(1,\ldots,m) \cap k \in \alpha\}} \mathtt{n}_\alpha \left(\sum_{j \in \alpha} \lambda_j \right)^{-1} ; k = 1, \ldots, m \end{aligned} \tag{40}$$

3.4 Confidence intervals

The MLE $\hat{\boldsymbol{\lambda}}$ is a point estimate of the true parameter value $\boldsymbol{\lambda}^\star$. As discussed in **??**, we would like to estimate how accurate this point estimate is by deriving its $(1-\alpha) \cdot 100\%$-confidence interval.

Theorem 3.7. *By definition 9, the asymptotic confidence interval for the failure rate $\lambda_j{}^\star$ of the j^{th} component is*

$$\hat{\lambda}_j \pm z_{\alpha/2} \sqrt{\left[\boldsymbol{\mathcal{I}}_n^{-1}(\boldsymbol{\lambda}^\star) \right]_{j,j}}. \tag{41}$$

where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal.

Therefore, to derive the confidence interval, we need to derive the information matrix $\boldsymbol{\mathcal{I}}_n(\boldsymbol{\lambda})$.

Theorem 3.8. *The (j,k)-th element of the information matrix $\boldsymbol{\mathcal{I}}_n(\boldsymbol{\lambda})$ is*

$$\left[\boldsymbol{\mathcal{I}}_n(\boldsymbol{\lambda}) \right]_{j,k} = \sum_{i \in \mathbb{I}} \left[\left(\sum_{p \in \mathrm{c}_i} \lambda_p \right)^{-2} \right] \tag{42}$$

where $\mathbb{I} = \{i : i \in \{1, \ldots, n\} \wedge j \in \mathrm{c}_i \wedge k \in \mathrm{c}_i\}$ is the set of indexed observations in the $\mathbb{F}_n$ sample in which corresponding candidate sets contain component indexes j and k.

Recall that we also developed a minimal sufficient statistic for the MLE $\hat{\boldsymbol{\lambda}}$.

Theorem 3.9. *The (j,k)-th element of the sufficient statistic information matrix $\boldsymbol{\mathcal{I}}_n(\boldsymbol{\lambda})$ is*

$$\left[\boldsymbol{\mathcal{I}}_n(\boldsymbol{\lambda}) \right]_{j,k} = \sum_{\alpha \in \mathcal{P}_{j,k}} \left[\mathtt{n}_\alpha \left(\sum_{p \in \alpha} \lambda_p \right)^{-2} \right] \tag{43}$$

where $\mathcal{P}_{j,k} = \{\alpha : \alpha \in \mathcal{P}(\{1, \ldots, m\}) \wedge j \in \alpha \wedge k \in \alpha\}$, which are the sets in the powerset of set $\{1, \ldots, m\}$ that includes component indexes j and k, and $\mathtt{n}_\alpha$ counts the number of times α is equal to candidate sets in the vector $(\mathrm{c}_1, \ldots, \mathrm{c}_n)$.

Proof.

$$
\begin{aligned}
\left[\boldsymbol{\mathcal{I}}_{j,k}\right]_{j,k} &= -\,\mathbb{E}\left[\frac{\partial^2}{\partial\lambda_j\partial\lambda_k}\ell(\boldsymbol{\lambda})\right] \\
&= -\,\mathbb{E}\left[\frac{\partial}{\partial\lambda_j}\left\{-\sum_{i=1}^{n}t_i + \sum_{\{i|i\in\{1,\ldots,n\}\wedge k\in \mathrm{c}_i\}}\left[\left(\sum_{p\in\mathrm{c}_i}\lambda_p\right)^{-1}\right]\right\}\right] \\
&= -\,\mathbb{E}\left\{\sum_{\{i|i\in\{1,\ldots,n\}\wedge j\in\mathrm{c}_i\wedge k\in\mathrm{c}_i\}}\left[-\left(\sum_{p\in\mathrm{c}_i}\lambda_p\right)^{-2}\right]\right\} \\
&= \sum_{\{i|i\in\{1,\ldots,n\}\wedge j\in\mathrm{c}_i\wedge k\in\mathrm{c}_i\}}\left[\left(\sum_{l\in\mathrm{c}_i}\lambda_p\right)^{-2}\right] \\
&= \sum_{\{\alpha|\alpha\in\mathcal{P}(\{1,\ldots,m\})\wedge j\in\alpha\wedge k\in\alpha\}}\left[\mathtt{n}_\alpha\left(\sum_{l\in\alpha}\lambda_p\right)^{-2}\right]
\end{aligned}
$$

□

3.5 Applications

3.5.1 Probability of component failure

In **??**, we derived functions that calculate the conditional probability that component k is responsible for a system failure given that a system failure occurred at time t, $\mathbb{P}_{\boldsymbol{\lambda}^\star}(k \mid t)$.

In the case of a series system with exponentially distributed component lifetimes, this conditional probability is independent of time t due to the constant failure rate of the component lifetimes.

Theorem 3.10. *The probability that component k is the cause of the system failure of an m-out-of-m system composed of m components with exponentially distributed lifetimes is*

$$\mathbb{P}_{\boldsymbol{\lambda}^\star}(k) = \frac{\lambda_k^\star}{\sum_{j=1}^{m}\lambda_j^\star} \tag{44}$$

Proof. The instantaneous failure rate of component j is ${\lambda_j}^\star$ for $j = 1, \ldots, m$. According to **??**, the probability that component k is responsible for a system failure given a system failure at time t is

$$\mathbb{P}_{\boldsymbol{\lambda}^\star}(k \mid t) = \frac{\lambda_k(t)}{\sum_{j=1}^{m}\lambda_j(t)} = \frac{\lambda_k^\star}{\sum_{j=1}^{m}\lambda_j^\star}.$$

Observe that this conditional probability is independent of time t. □

In fact, we claim without proof that eq. (44) holds true for component j being the cause of a system failure given a system failure at some uncertain time $t \in (t_a, t_b)$.

Note that if $\boldsymbol{\lambda}^\star$ is not known, then it must be estimated as described in section 3.3.

3.5.2 Distribution of inspection counts

The optimal order policy σ is inspecting the components in the order of decreasing failure rates,

$$\sigma(j) \leq \sigma(k) \iff \lambda^\star_{\sigma(j)} \geq \lambda^\star_{\sigma(k)} \tag{45}$$

Theorem 3.11. *The discrete conditional random variable N_σ has a probability mass function*

$$\mathbb{P}_\sigma(n) = \begin{cases} \frac{\lambda^\star_{\sigma(n)}}{\sum_{j=1}^m \lambda^\star_j} & n \in \{1, \ldots, m\} \\ 0 & n \notin \{1, \ldots, m\} \end{cases} \tag{46}$$

where n denotes the number of inspections required to identify the failed component.

Theorem 3.12. *The expected number of inspections with respect to the optimal order policy σ and discrete random variable $K \sim \mathbb{P}_{\boldsymbol{\lambda}^\star}(k)$ is*

$$\mathbb{E}[N_\sigma] = \frac{\sum_{n=1}^m n \cdot \lambda^\star_{\sigma(n)}}{\sum_{j=1}^m \lambda^\star_j} \tag{47}$$

Proof. The expectation of the random variable N_σ is

$$\mathbb{E}[N_\sigma] = \sum_{n=1}^m n \cdot \mathbb{P}\sigma(n) = \sum_{n=1}^m \left[n \cdot \frac{\lambda^\star_{\sigma(n)}}{\sum_{j=1}^m \lambda^\star_j} \right]$$

Notice that $\frac{1}{\sum_{i=1}^m \lambda_i}$ is a common factor of each term, so it can be pulled outside of the summation. □

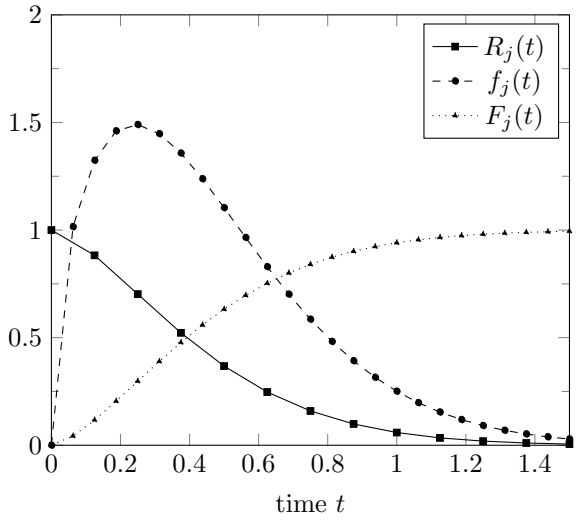
4 Weibull series system

4.1 Component lifetimes

The Weibull distribution is a popular continuous distribution for modeling lifetimes. Consider a series system composed of m components with Weibull distributed lifetimes, $T_j \sim \text{WEI}(\alpha_j, \beta_j)$ for $j = 1, \ldots, m$.

Recall that a component j with an exponentially distributed lifetime has a constant failure rate λ_j. Conversely, a component j with a Weibull distributed lifetime has a failure rate $\lambda_j(t) = \beta_j {\alpha_j}^{\beta_j} t^{\beta_j - 1})$. Thus,

Figure 3: Weibull distributed lifetime with $\alpha_j = 2.0$, $\beta_j = 1.5$



component j has a failure rate that increases with respect to time t (e.g., wears out) if $\beta_j > 1$, decreases with respect to time t (e.g., has a high infant mortality rate.) if $0 < \beta_j < 1$, and remains constant with respect to time t if $\beta_j = 1$. Consequently, the Weibull is a generalization of the Exponential in which $T_j \sim \text{WEI}(\alpha_j, \beta_j = 1) \iff T_j \sim \text{EXP}(\lambda_j = \alpha_j)$.

Definition 14. *Consider a series system composed of m components with Weibull distributed lifetimes, $T_j \sim \text{WEI}(\alpha_j, \beta_j)$ for $j = 1, \ldots, m$. The lifetime of component j has reliability, CDF, and PDF functions*

$$R_j(t \mid \alpha_j, \beta_j) = \begin{cases} \mathrm{e}^{-(\alpha_j t)^{\beta_j}} & t \geq 0 \\ 1 & t < 0 \end{cases} \tag{48}$$

$$F_j(t \mid \alpha_j, \beta_j) = \begin{cases} 1 - \mathrm{e}^{-(\alpha_j t)^{\beta_j}} & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{49}$$

$$f_j(t \mid \alpha_j, \beta_j) = \begin{cases} \beta_j \alpha_j^{\beta_j} t^{\beta_j - 1} \, \mathrm{e}^{-(\alpha_j t)^{\beta_j}} & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{50}$$

where $\alpha_j > 0$ ($\frac{1}{\alpha_j}$ is the scale) and $\beta_j > 0$ is the shape parameter.

4.2 System lifetime

Theorem 4.1. *System lifetime. Consider a series system composed of m components with Weibull distributed lifetimes, $T_j \sim \text{WEI}(\alpha_j, \beta_j)$ for $j = 1, \ldots, m$. The system has a lifetime $T = \min(T_1, \ldots, T_m)$ with reliability, CDF, and PDF functions*

$$R(t \mid \boldsymbol{\theta}) = \begin{cases} \mathrm{e}^{-\sum_{j=1}^{m} (\alpha_j t)^{\beta_j}} & t \geq 0 \\ 1 & t < 0, \end{cases} \tag{51}$$

$$F(t \mid \boldsymbol{\theta}) = \begin{cases} 1 - \mathrm{e}^{-\sum_{j=1}^{m} (\alpha_j t)^{\beta_j}} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{52}$$

$$f(t \mid \boldsymbol{\theta}) = \begin{cases} \left[\mathrm{e}^{-\sum_{j=1}^{m} (\alpha_j t)^{\beta_j}}\right] \left[\sum_{j=1}^{m} \beta_j \alpha_j^{\beta_j} t^{\beta_j - 1}\right] & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{53}$$

where $\boldsymbol{\theta} = (\alpha_1, \beta_1, \ldots, \alpha_m, \beta_m)$.

Proof. To prove Equation (51), by Equation (5), $R(t) = \prod_{j=1}^{m} R_j(t)$. Plugging in the component reliability functions $R_j(t)$ for $j = 1, \ldots, m$ for this system, we get

$$R(t) = \prod_{j=1}^{m} \mathrm{e}^{-(\alpha_j t)^{\beta_j}} = \mathrm{e}^{-\sum_{j=1}^{m} (\alpha_j t)^{\beta_j}}.$$

To prove Equation (52), the CDF is the complement of the reliability function, $F(t) = 1 - R(t)$.

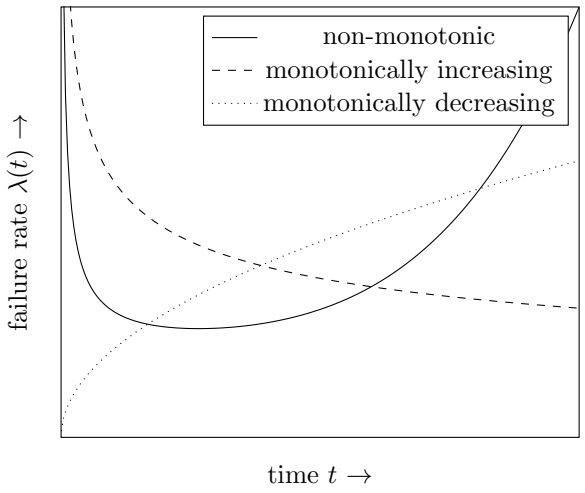
To prove Equation (53), by Equation (7),

$$f(t) = \sum_{j=1}^{m} \left(f_j(t) \prod_{\substack{k=1 \\ k \neq j}}^{m} R_k(t) \right).$$

Plugging in the component reliability and PDF functions $R_j(t)$ and $f_j(t)$ respectively for $j = 1, \ldots, m$ for this system, we get

$$f(t) = \sum_{j=1}^{m} \left(\beta_j \alpha_j^{\beta_j} t^{\beta_j - 1} \, \mathrm{e}^{-(\alpha_j t)^{\beta_j}} \prod_{\substack{k=1 \\ k \neq j}}^{m} \mathrm{e}^{-(\alpha_k t)^{\beta_k}} \right).$$

Figure 4: Failure rate monotonicity



Converting the product into a sum,

$$f(t) = \sum_{j=1}^{m} \left\{ \beta_j \alpha_j^{\beta_j} t^{\beta_j - 1} \exp\left(-(\alpha_j t)^{\beta_j}\right) \exp\left(- \sum_{\substack{k=1 \\ k \neq j}}^{m} (\alpha_k t)^{\beta_k} \right) \right\}.$$

Combining like terms,

$$f(t) = \sum_{j=1}^{m} \left\{ \beta_j \alpha_j^{\beta_j} t^{\beta_j - 1} \exp\left(- \sum_{k=1}^{m} (\alpha_k t)^{\beta_k} \right) \right\}.$$

Pulling the exponential out of the summation,

$$f(t) = \exp\left(- \sum_{k=1}^{m} (\alpha_k t)^{\beta_k} \right) \sum_{j=1}^{m} \left\{ \beta_j \alpha_j^{\beta_j} t^{\beta_j - 1} \right\}.$$

□

Recall that a series system that is strictly composed of components with exponentially distributed lifetimes also has an exponentially distributed lifetime. This is not true for any other continuous lifetime distribution. A series system composed of components with Weibull distributed lifetimes does not have a Weibull distributed lifetime[6].

[6]Excepting the special case when the component lifetimes are identically distributed, i.e., if component j has a lifetime distributed $T_j \sim \text{WEI}(\alpha, \beta)$ for $j = 1, \ldots, m$, then the system has a lifetime distributed $T \sim \text{WEI}(\alpha m^{1/\beta}, \beta)$.

The failure rate of this system, given by theorem 2.5, is

$$\lambda(t) = \sum_{j=1}^{m} \beta_j \alpha_j^{\beta_j} t^{\beta_j - 1}, \tag{54}$$

, thus, the failure rate $\lambda(t)$ is monotonically increasing if $\beta_j > 1$ for $j = 1, \ldots, m$, monotonically decreasing if $\beta_j < 1$ for $j = 1, \ldots, m$, and non-monotonic otherwise[7]. See *Figure* 4 for an example of each.

5 Pareto series system

5.1 Component lifetimes

Previously, we explored systems composed of components with Exponential and Weibull distributed lifetimes. The Exponential is a specialization of the Weibull, so essentially we have so far only explored a class of distribution in which the probability of survival decays at an exponential rate. (The Weibull is a product of a power growth rate and an exponential decay rate, but the exponential dominates the power for large t.) In contrast, the Pareto distribution is a continuous power law distribution in which the probability of survival decays at a power rate.

Furthermore, the Pareto distribution is heavy-tailed since it decays at a power rate rather than an exponential rate. Related to the Pareto's heavy tail, its failure rate $\lambda_j(t)$ is monotonically decreasing with respect to time t, e.g., components with a Pareto lifetime distribution become more reliable with time. This is an appropriate model for components which have a *burning in* period.

The Pareto has been empirically shown to usefully model many types of observable phenomena. It is related to the Pareto principle, a rule of thumb that observes that for many events, approximately 80% of the effects come from 20% of the causes. For example, in economics it is often observed that around 80% of the wealth is concentrated in 20% of the population.

Definition 15. *Consider a series system composed of m components with Pareto distributed lifetimes, $T_j \sim \mathrm{PAR}(\theta_j, \kappa_j)$ for $j = 1, \ldots, m$. The lifetime of component j has reliability, CDF, and PDF functions*

$$R_j(t \mid \theta_j, \kappa_j) = \begin{cases} \left(1 + \frac{t}{\theta_j}\right)^{-\kappa_j} & t \geq 0 \\ 1 & t < 0, \end{cases} \tag{55}$$

[7] If $\beta_j = 1$ for $j = 1, \ldots, m$, then the system lifetime is exponentially distributed and has a constant failure rate.

$$F_j(t \mid \theta_j, \kappa_j) = \begin{cases} 1 - \left(1 + \frac{t}{\theta_j}\right)^{-\kappa_j} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{56}$$

$$f_j(t \mid \theta_j, \kappa_j) = \begin{cases} \frac{\kappa_j}{\theta_j} \left(1 + \frac{t}{\theta_j}\right)^{-(\kappa_j+1)} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{57}$$

where scale parameter $\theta_j > 0$ and shape parameter $\kappa_j > 0$.

Theorem 5.1. *A component j with a Pareto distributed lifetime has a monotonically decreasing failure rate,*

$$\lambda_j(t) = \frac{\kappa_j}{\theta_j + t}. \tag{58}$$

Proof. According to Equation (4), $\lambda_j(t) = \frac{f_j(t)}{R_j(t)}$. Substituting the Pareto distribution functions for $f_j(t)$ and $R_j(t)$, we get

$$\lambda_j(t) = \frac{\kappa_j}{\theta_j} \left(1 + \frac{t}{\theta_j}\right)^{-(\kappa_j+1)} \cdot \left\{ \left(1 + \frac{t}{\theta_j}\right)^{\kappa_j} \right\}^{-1} = \frac{\kappa_j}{\theta_j} \left(\frac{\theta_j + t}{\theta_j}\right)^{-1} = \frac{\kappa_j}{\theta_j + t}.$$

□

5.2 System lifetime

Theorem 5.2. *Consider a series system composed of m components with Pareto distributed lifetimes, $T_j \sim \text{PAR}(\alpha_j, \beta_j)$ for $j = 1, \ldots, m$. The system has a lifetime $T = \min(T_1, \ldots, T_m)$ with reliability, CDF, and PDF functions*

$$R(t \mid \boldsymbol{\theta}) = \prod_{j=1}^{m} \left(1 + \frac{t}{\theta_j}\right)^{-\kappa_j}, \tag{59}$$

$$F(t \mid \boldsymbol{\theta}) = 1 - \prod_{j=1}^{m} \left(1 + \frac{t}{\theta_j}\right)^{-\kappa_j}, \tag{60}$$

$$f(t \mid \boldsymbol{\theta}) = f(t) = \left[\prod_{j=1}^{m} \left(1 + \frac{t}{\theta_j}\right)^{-\kappa_j}\right] \left[\sum_{j=1}^{m} \frac{\kappa_j}{\theta_j + t}\right] \tag{61}$$

where $\boldsymbol{\theta} = (\theta_1, \kappa_1, \ldots, \theta_m, \kappa_m)$.

Proof. To prove Equation (59), by Equation (5), $R(t) = \prod_{j=1}^{m} R_j(t)$. Plugging in the component reliability

functions $R_j(t)$ for $j = 1, \ldots, m$ for this system, we get

$$R(t) = \prod_{j=1}^{m} \left(1 + \frac{t}{\theta_j}\right)^{-\kappa_j}.$$

To prove Equation (60), the CDF is the complement of the reliability function, $F(t) = 1 - R(t)$.

To prove Equation (61), by Equation (7),

$$f(t) = \sum_{j=1}^{m} \left(f_j(t) \prod_{\substack{k=1 \\ k \neq j}}^{m} R_k(t) \right).$$

Plugging in the component reliability and PDF functions $R_j(t)$ and $f_j(t)$ respectively for $j = 1, \ldots, m$ for this system, we get

$$f(t) = \sum_{j=1}^{m} \left\{ \frac{\kappa_j}{\theta_j} \left(1 + \frac{t}{\theta_j}\right)^{-(\kappa_j+1)} \prod_{\substack{k=1 \\ k \neq j}}^{m} \left(1 + \frac{t}{\theta_k}\right)^{-\kappa_k} \right\}.$$

$$f(t) = \sum_{j=1}^{m} \left\{ \frac{\kappa_j}{\theta_j} \left(1 + \frac{t}{\theta_j}\right)^{-1} \prod_{k=1}^{m} \left(1 + \frac{t}{\theta_k}\right)^{-\kappa_k} \right\}.$$

$$f(t) = \left\{ \prod_{j=1}^{m} \left(1 + \frac{t}{\theta_j}\right)^{-\kappa_j} \right\} \sum_{j=1}^{m} \left\{ \frac{\kappa_j}{\theta_j} \left(1 + \frac{t}{\theta_j}\right)^{-1} \right\}.$$

$$f(t) = \left\{ \prod_{j=1}^{m} \left(1 + \frac{t}{\theta_j}\right)^{-\kappa_j} \right\} \sum_{j=1}^{m} \left\{ \frac{\kappa_j}{\theta_j} \left(\frac{\theta_j + t}{\theta_j}\right)^{-1} \right\}.$$

$$f(t) = \left[\prod_{j=1}^{m} \left(1 + \frac{t}{\theta_j}\right)^{-\kappa_j} \right] \left[\sum_{j=1}^{m} \frac{\kappa_j}{\theta_j + t} \right]$$

□

The failure rate of this system, given by theorem 2.5, is

$$\lambda(t) = \sum_{j=1}^{m} \frac{\kappa_j}{\theta_j + t}, \tag{62}$$

thus the failure rate $\lambda(t)$ is monotonically decreasing.

A Appendix

A.1 Numerically solving the MLE $\hat{\boldsymbol{\theta}}_n$

The function $\ell(\boldsymbol{\theta} \mid \mathbb{F}_n)$ is the log-likelihood of the sample $\mathbb{F}_n$ with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ where $\boldsymbol{\Theta} \subset \mathbb{R}^q$. This function has a surface in an $\mathbb{R}^{q+1}$ dimensional space, where a particular point on this surface represents the log-likelihood of observing $\mathbb{F}_n$ with respect to $\boldsymbol{\theta}$.

The point $\hat{\boldsymbol{\theta}}_n \in \boldsymbol{\Theta}$ on this surface which is at the maximum height,

$$\hat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ell(\boldsymbol{\theta} \mid \mathbb{F}_n),$$

is denoted the maximum likelihood estimate. There is in general no closed-form solution that solves the MLE $\hat{\boldsymbol{\theta}}_n$. Instead, local search methods are typically used to numerically approximate the solution.

Let us assume that $\boldsymbol{\Theta}$ is the intersection of a p linear inequality constraints (half-planes), $\sum_{j=1}^{q} \alpha_{i,j}\theta_{i,j} \geq 0$ for $i = 1, \ldots, p$. Thus, since $\boldsymbol{\Theta}$ is an intersection of half-spaces, it denotes a convex polyhedron in a q-dimensional space.

The general version of the local search function subject to the constraint $\hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$, MLE_SEARCH_PROJECTION is shown in algorithm 2. It returns an approximation of the MLE $\hat{\boldsymbol{\theta}}_n$. Specifically, since MLE_SEARCH_PROJECTION is a local search method, it finds a local maximum. If the log-likelihood function ℓ is non-convex, then in general this local maximum cannot be guaranteed to be the global maximum.[8]

function MLE_SEARCH_PROJECTION($\boldsymbol{\theta}^{(0)}$: *initial starting point*, ℓ: *log-likelihood function*)
- $\boldsymbol{\theta}^{(0)} \leftarrow \text{proj}(\boldsymbol{\theta}^{(0)}, \boldsymbol{\Theta})$;
- $i \leftarrow 0$;
- **while** stopping criteria *not satisfied* **do**
  - $\boldsymbol{\theta}^{(i+1)} \leftarrow \text{proj}(\boldsymbol{\theta}^{(i)} + \alpha^{(i)}\mathbf{d}^{(i)})$;
  - $i \leftarrow i + 1$;
- **end**
- **return** $\boldsymbol{\theta}^{(i)}$;

Algorithm 2: General MLE local search projection

The function $\text{proj}(\boldsymbol{\theta}, \boldsymbol{\Theta})$ depicted in algorithm 2 projects any point $\boldsymbol{\theta}$ to the nearest point in $\boldsymbol{\Theta}$, restricting the parameter search to the feasiable parameter space. See appendix A.1 for an illustration. However, intermediate values $\boldsymbol{\theta}^{(i)} + \alpha^{(i)} \cdot \mathbf{d}^{(i)}$ can be outside of the feasible parameter space, and since some of these infeasible values may not be defined for ℓ or $\nabla\ell$, caution must be exercised.

[8] MLE_SEARCH_PROJECTION may be a part of a larger search function that invokes MLE_SEARCH_PROJECTION multiple times such that the best local maximum found is expected to converge to $\hat{\boldsymbol{\theta}}_n$ over time.

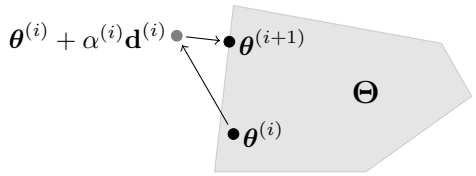


Figure 5: Projection onto convex $\boldsymbol{\Theta}$.

The search function depends on a few unspecified parameters. (i) The unit vector $\mathbf{d}^{(i)}$ is a *promising* direction in which to search for a better solution than $\boldsymbol{\theta}^{(i)}$. (ii) The scalar $\alpha^{(i)} > 0$ is the distance to move from point $\boldsymbol{\theta}^{(i)}$ to generate the next point, $\boldsymbol{\theta}^{(i+1)}$. (iii) The iterations are repeated until some *stopping criteria* is satisfied.

Gradient ascent provides straightforward choices for these parameters. The direction vector $\mathbf{d}^{(i)}$ is chosen to be the normalized gradient, $\frac{\nabla \ell(\boldsymbol{\theta}^{(i)})}{\|\nabla \ell(\boldsymbol{\theta}^{(i)})\|}$, the scalar $\alpha^{(i)}$ is chosen to be magnitude of the gradient, $\|\nabla \ell\left(\boldsymbol{\theta}^{(i)}\right)\|$, and the *stopping criteria* is chosen to be when the magnitude of the gradient is less than some upper threshold, $\|\nabla \ell\left(\boldsymbol{\theta}^{(i)}\right)\| \leq \epsilon$.

The motivation behind these choices in gradient ascent is to make *greedy* choices at each step in the algorithm. It greedily chooses $\mathbf{d}^{(i)}$ to be the direction of steepest ascent. It greedily chooses $\alpha^{(i)}$ to be the magnitude of the gradient, thus taking small steps when approaching a local maximum so that overshooting is mitigated. Finally, the *stopping criteria* is chosen to be when the gradient is *approximately* the desired zero vector, signifying that a stationary point has at least been approximately reached.

Implementation We implemented a variation of gradient ascent where $\mathbf{d^{(i)}}$ is the unit vector in the direction of the gradient and $\alpha^{(i)}$ is determined by a *Golden Section* line search. Specifically, we chose a value of $\alpha^{(i)} \in (0, 1)$ such that $\alpha_{(i)} = \arg\max_{\alpha \in (0,1)} \ell(\boldsymbol{\theta}^{(i)} + \alpha^{(i)} \cdot \mathbf{d}^{(i)})$ is approximated.

In addition, we invoked MLE_SEARCH_PROJECTION multiple times, each time using a randomized starting point $\boldsymbol{\theta}^{(0)}$, and chose the solution that resulted in the maximum value as the MLE $\hat{\boldsymbol{\theta}}$.

We explored parametric distribution families in which the parameters are restricted to positive values greater than 0. Thus, the projection function in algorithm 2, proj, is making any negative parameter values to be ϵ, where ϵ is a small positive scalar value near 0.

Prior information We assume that the m-out-of-m system has a lifetime distribution $T \sim f(t \mid \boldsymbol{\theta}^{\star} \in \boldsymbol{\Theta})$. If there is additional prior information about the value of $\boldsymbol{\theta}^{\star}$ in the form of linear constraints, e.g., $\theta_1 + \theta_3 \leq \theta_4$, then appropriate projections can be defined for those also.

References

[1] J. Doe. Placeholder.