Estimating the sampling distribution of the maximum likelihood estimator of the parameters of a series system given a sample of masked system failure times

Alexander Towell
atowell@siue.edu

Abstract

First, the *parametric family* of an abstract *series system* is derived. Then, the *asymptotic sampling distribution* (*multivariate normal*) of a *maximum likelihood estimator* of the true parameter is derived conditioned on the *information* given by a random sample of *masked system failure times*. The asymptotic sampling distributions of statistics that are functions of the parameter follow as a result. Finally, these results are applied to series systems in which component lifetimes are *exponentially* distributed, which lead to closed formulas on a set of *minimally sufficient statistics*.

Contents

Nomenclature

$\mathcal{C}$ The sample space of $\mathbf{C}$, the set of all candidate sets.

$\mathcal{C}_{\boldsymbol{w}}$ The sample space of $\mathbf{C}$ given $\mathrm{W} = w$, the set of all candidate sets of cardinality w.

$\boldsymbol{\Sigma}_{\boldsymbol{n}}(\boldsymbol{\theta}^\star \mid w)$ The *expected* variance-covariance of the asymptotic sampling distribution of $\hat{\boldsymbol{\theta}}(n \mid w)$.

$\boldsymbol{\Sigma}_{\boldsymbol{n}}(\boldsymbol{\theta}^\star)$ The *expected* variance-covariance of the asymptotic sampling distribution of $\hat{\boldsymbol{\theta}}_{\mathbf{n}}$.

T_j The continuous random lifetime of the j^{th} component.

$\ell(\boldsymbol{\theta} \mid w\,,\,\cdot\,)$ The log-likelihood function conditioned on $\mathrm{W} = w$.

$\mathbb{1}_{\mathcal{X}}$ The indicator function $\mathbb{1}_{\mathcal{X}} \colon \mathcal{X} \mapsto \{0, 1\}$ outputs 1 if the input is a member of $\mathcal{X}$ and otherwise outputs 0.

$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}^\star \mid w)$ The *expected* information matrix for a random masked system failure times conditioned on w candidates.

$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}^\star)$ The *expected* information matrix for a random masked system failure times.

$\boldsymbol{\mathcal{I}}_{\boldsymbol{n}}(\boldsymbol{\theta}^\star \mid w)$ The *expected* information matrix for a random sample of n masked system failure times conditioned on w candidates.

$\boldsymbol{\mathcal{I}}_{\boldsymbol{n}}(\boldsymbol{\theta}^\star)$ The *expected* information matrix for a random sample of n masked system failure times.

$f_{\mathbf{C},\mathrm{S} \mid \mathrm{W}}$ The joint conditioned probability density function of $\mathbf{C}$, S given some realization of W.

$\mathcal{L}(\boldsymbol{\theta} \mid w\,,\,\cdot\,)$ The likelihood function.

$\mathbf{M}_{\mathbf{n}}$ A random sample of n masked system failure times.

$\mathbf{m}_{\mathbf{n}}$ A realization of a sample of n masked system failure times.

$\mathbf{M}(w, n)$ A random sample of n masked system failure times conditioned on w candidates.

$\hat{\boldsymbol{\theta}}(n \mid w)$ The maximum likelihood estimator conditioned on a sample of n masked system failure times in which only candidate sets of cardinality w are observed .

$\hat{\boldsymbol{\theta}}_{\mathbf{n}}$ The maximum likelihood estimator conditioned on a sample of n masked system failure times.

f_j The probability density function of T_j.

$\mathcal{F}_{\boldsymbol{\Theta}}$ The parametric family of the system.

$\mathcal{F}_{\mathbf{v}}$ The parametric family of the component lifetimes.

$p_{\mathrm{K}\,|\,\mathrm{S,W}}$ The conditional probability mass function of W given some realization of S and W.

p_{W} The marginal probability mass function of W.

$\mathbf{C}$ The discrete random variable (set) denoting the components of a candidate set.

K The discrete random variable denoting the cause of system failure.

W The discrete random variable denoting the cardinality of the candidate set.

$\mathbf{Y}(n\,|\,w)$ The sampling distribution of $\hat{\boldsymbol{\theta}}(n\,|\,w)$.

$\mathbf{Y_n}$ The sampling distribution of $\hat{\boldsymbol{\theta}}_{\mathbf{n}}$.

$\mathrm{s}(\boldsymbol{\theta}\,|\,w\,,\,\cdot\,)$ The score function, the gradient of the log-likelihood.

f_{S} The probability density function of S.

S The continuous random lifetime of the series system.

$\boldsymbol{\Theta}$ A parameter index (matrix) of the parametric family of the system.

$\boldsymbol{\Theta}_j$ A parameter index (vector) of the parametric family of the j^{th} component.

$\boldsymbol{\theta}$ A parameter index (vector) of the parametric family of the system.

$\boldsymbol{\Theta}^\star$ The true parameter index (matrix) of the parametric family of the system.

$\boldsymbol{\theta}^\star$ The true parameter index (vector) of the parametric family of the system.

Chapter 1

Introduction

We consider a series system composed of m components, where each component has an uncertain lifetime. We desire a mathematical model that predicts the failure time of the system and its component causes. However, since the system lifetime and the cause of failure is uncertain, we are interested in probabilistically modeling the system lifetime, e.g., there is a 75% chance that component 3 will cause a system failure in the next 3 years.

We assume no prior information about the system except for a parametric model and a sample of n *masked system failure times*. We employ the frequentist approach of repeated experiments which induces a sampling distribution on the point estimator and assume the true parameter of the model is a fixed unknown value. This is in contrast to the Bayesian approach where the true parameter is treated probabilistically. However, we assume the large sample approximation which overrides any Bayesian prior on the true parameter, in which case the frequentist and Bayesian approaches converge to the same result, a point estimator with a multivariate normal sampling distribution.

Chapter 2

Probabilistic model

A probabilistic model is specified by equations involving random variables which make assumptions about how observable data about the system is generated. In what follows, we describe our model and observable data.

2.1 Components

The system consists of m components. We make the following assumption about the states of the components.

Assumption 2.1. *The j^{th} component is initially in a non-failed state and permanently enters a failed state at some uncertain time $t_j > 0$ for $j = 1, \ldots, m$.*

Components have lifetimes as given by the following definition.

Definition 2.1. *The uncertain lifetime of the j^{th} component is given by the continuous random variable $\mathrm{T}_j \in (0, \infty)$ for $j = 1, \ldots, m$.*

We make the following assumption about the joint distribution of the component lifetimes.

Assumption 2.2. *The random variables $\mathrm{T}_1, \ldots, \mathrm{T}_m$ are mutually independent.*

The cumulative distribution function describes the distribution of random variables and is given by the following definition.

Definition 2.2 (Cumulative distribution function)**.** *Let X be some random variable. The probability that $\mathrm{X} \leq x$ is given by its cumulative distribution function, denoted by*

$$F_{\mathrm{X}}(x) = \Pr[\mathrm{X} \leq x]\,. \tag{2.1}$$

The component lifetimes have cumulative distribution functions given by the following definition.

Definition 2.3. *The probability that the j^{th} component will fail before time t, $\mathrm{T}_j \leq t$, is given by*

$$F_j(t) \equiv F_{\mathrm{T}_j}(t)\ . \tag{2.2}$$

The probability density function is given by the following definition.

Definition 2.4. *Let* X *be some continuous random variable. The probability density function is given by*

$$f_{\mathrm{X}}(x) = \frac{\mathrm{d}}{\mathrm{d}t} F_{\mathrm{X}}(t)\,. \tag{2.3}$$

By the Second Fundamental Theorem of Calculus, the probability that a continuous random variable $\mathrm{X} \in (a, b]$, $a \leq b$, is given by

$$\Pr[a < \mathrm{X} \leq b] = F_{\mathrm{X}}(b) - F_{\mathrm{X}}(a) = \int_a^b f_{\mathrm{X}}(s)\mathrm{d}s\,. \tag{2.4}$$

By eq. (2.4), the probability that a continuous random variable $\mathrm{X} = x$ is given by

$$\Pr[x < \mathrm{X} \leq x] = \int_t^t f_{\mathrm{X}}(s)\mathrm{d}s = 0\,. \tag{2.5}$$

The probability density function evaluated at x is said to calculate the *relative likelihood* of x.

The component lifetimes have probability density functions given by the following definition.

Definition 2.5. *The* relative likelihood *that the j^{th} component will fail at time t, $\mathrm{T}_j = t$, is given by*

$$f_j(t) \equiv f_{\mathrm{T}_j}(t)\,. \tag{2.6}$$

We restrict our attention to parametrized probability density functions.

Notation. Column vectors are denoted by boldface lower case letters, e.g., $\mathbf{x}$, and row vectors are denoted by the transpose of column vectors, e.g., $\mathbf{x}^{\intercal}$.

Assumption 2.3. *The component lifetimes have probability density functions that are members of the parametric family of probability density functions denoted by*

$$\mathcal{F}_{\mathbf{v}} = \left\{ f\big(\,\cdot \,|\, \mathbf{v}\big) : \mathbf{v} \in \boldsymbol{\Omega} \right\}\,, \tag{2.7}$$

where $\boldsymbol{\Omega} \subset \mathbb{R}^q$ is the parameter support.

Particular values of $\mathbf{v}$ index particular probability density functions within the $\mathcal{F}_{\mathbf{v}}$-family. The true probability density function of the j^{th} component lifetime is given by the following definition.

Definition 2.6. *The true parameter index of the parametric family of the j^{th} component lifetime is denoted by $\mathbf{v}_j^{\star}$, therefore*

$$\mathrm{T}_j \sim f_j\Big(\,\cdot \,|\, \mathbf{v}_j^{\star}\Big)\,. \tag{2.8}$$

The complement of the cumulative distribution function is the reliability function and is given by the following definition.

Definition 2.7 (Reliability function)**.** *Let* X *be some random variable. The probability that $\mathrm{X} > x$ is given by the reliability function, denoted by*

$$R_{\mathrm{X}}(x) = \Pr\left[\mathrm{X} > x\right] = 1 - F_{\mathrm{X}}(x)\,. \tag{2.9}$$

The component lifetimes have reliability functions given by the following definition.

Definition 2.8. *The probability that the j^{th} component will not fail before time t, $\mathrm{T}_j > t$, is given by*

$$R_j\big(t \,|\, \mathbf{v}^{\star}\big) \equiv R_{\mathrm{T}_j}\big(t \,|\, \mathbf{v}^{\star}\big)\,. \tag{2.10}$$

2.2 System

The system is a series system as given by the following assumption.

Assumption 2.4. *The system is in a non-failed state if all m components are in a non-failed state, otherwise the system is in a failed state.*

A system in which at least k of the components must be in a non-failed state for the system to be in a non-failed system is denoted a k-out-of-m system, thus a series system is an m-out-of-m system.

The system lifetime of the system is given by the following definition.

Theorem 2.1. *The uncertain lifetime of the system is given by*

$$\mathrm{S} = \min(\mathrm{T}_1\,, \ldots\,, \mathrm{T}_m) \in (0, \infty)\ , \tag{2.11}$$

where $\mathrm{T}_1\,, \ldots\,, \mathrm{T}_m$ *are the component lifetimes.*

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

By theorem 2.1, if a system failure occurs at time t, then one of the components failed at time t (only a single component can cause a system failure since two or more continuous random variables cannot realize the same value).

Definition 2.9. *Let* K *denote the discrete random variable indicating the component responsible for a system failure, e.g.,* $\mathrm{K} = j$ *indicates that* $\mathrm{T}_j < \mathrm{T}_k$ *for* $k \in \{1\,, \ldots\,, m\} \setminus \{j\}$.

Notation. Matrices are denoted by boldface upper case greek letters, e.g., $\boldsymbol{A}$. The j^{th} column of $\boldsymbol{A}$ is denoted by $\boldsymbol{A}_j$ (or $[\boldsymbol{A}]_j$) and the (j,k)-th element of $\boldsymbol{A}$ is denoted by $\boldsymbol{A}_{j\,k}$ (or $[\boldsymbol{A}]_{j\,k}$).

The system lifetime has a probability density function given by the following definition.

Definition 2.10. *The system lifetime has a probability density function that is a member of the parametric family of probability density functions denoted by*

$$\mathcal{F}_{\boldsymbol{\Theta}} = \left\{ f\left(\,\cdot\,|\,\boldsymbol{\Theta}\right) : \boldsymbol{\Theta} \in \mathbb{R}^{q \times m} \right\}\ , \tag{2.12}$$

where $\boldsymbol{\Theta}_j \in \boldsymbol{\Omega}$ *corresponds to the parameter index of the* j^{th} *component lifetime.*

Particular values of $\boldsymbol{\Theta}$ index particular probability density functions within the $\mathcal{F}_{\boldsymbol{\Theta}}$-family. The true probability density function of the system lifetime is given by the following definition.

Definition 2.11. *The true parameter index of the parametric family of system lifetime is denoted by* $\boldsymbol{\Theta}^\star$*, therefore*

$$\mathrm{S} \sim f_{\mathrm{S}}\big(\,\cdot\,|\,\boldsymbol{\Theta}^\star\big)\ . \tag{2.13}$$

2.3 Candidate sets

Each time a system failure occurs, one of the components failed. Given a set of them, we may compute the probability that one of them was the cause. The set of all component sets is given by the following definition.

Definition 2.12. *We denote the set of all candidate sets by*

$$\boldsymbol{\mathcal{C}} \equiv \Big\{ \mathbf{c} \in \mathcal{P}\left(\{1\,,\ldots\,,m\}\right) : \mathbf{c} \neq \emptyset \cap \mathbf{c} \neq \{1\,,\ldots\,,m\} \Big\}\,, \tag{2.14}$$

where $\mathcal{P}$ *is the power set function.*

The probability that a component in a set $\mathbf{c}$ is the cause of a system failure follows from the definition of K, i.e., $\Pr\left[\mathrm{K} \in \mathbf{c}\right]$. The total probability over $\boldsymbol{\mathcal{C}}$ does not sum to unity since candidate sets are not disjoint events with respect to the sample space of K. We consider another random variable (random set) in which the sample space is $\boldsymbol{\mathcal{C}}$.

The uncertain candidate sets are given by the following definition.

Definition 2.13. *Let* $\mathbf{C}$ *denote the random set of components which are the most likely causees of a corresponding system failure. The sample space of* $\mathbf{C}$ *is* $\boldsymbol{\mathcal{C}}$*.*

Assumption 2.5. *The cardinality of the random set* $\mathbf{C}$*, denoted by the discrete random variable*

$$\mathrm{W} = |\mathbf{C}| \in \{1\,,\ldots\,,m-1\}\,, \tag{2.15}$$

is independent of the random system lifetime S *and does not carry information about the true parameter index* $\boldsymbol{\Theta}^\star$*.*

2.4 Masked system failure time

Definition 2.14. *A* masked system failure time *consists of a system failure time* t *and a corresponding set of components* $\mathbf{c}$ *that* plausibly *caused the system failure.*

Assumption 2.6. *A random* masked system failure time *is a jointly distributed random system lifetime* S *and random candidate set* $\mathbf{C}$*.*

Assumption 2.7. *A random sample of* n masked system failure times*, denoted by*

$$\mathbf{M_n} = \left(\mathrm{S}_1,\, \mathbf{C}_1\right)\,,\,\ldots\,,\,\left(\mathrm{S}_n,\, \mathbf{C}_n\right)\,, \tag{2.16}$$

consists of n *independent and identically jointly distributed pairs of system failure times and candidate sets.*

Definition 2.15. *A particular realization of* $\mathbf{M_n}$ *is denoted by*

$$\mathbf{m_n} = \left(\mathrm{t}_1,\, \mathbf{c}_1\right)\,,\,\ldots\,,\,\left(\mathrm{t_n},\, \mathbf{c}_\mathrm{n}\right)\,. \tag{2.17}$$

Assumption 2.8. *The only information we have about the system and its true parameter index* $\boldsymbol{\Theta}^\star$ *is given by* $\mathbf{m_n}$*, a sample of* n *masked system failure times.*

Given the assumed model, the object of statistical interest is the (unknown) true parameter index $\boldsymbol{\Theta}^\star$. Provided an estimate of $\boldsymbol{\Theta}^\star$, any characteristic that is a function of $\boldsymbol{\Theta}^\star$ can be estimated as a result, thus the primary objective is to use the *information* in a sample of n masked system failure times to estimate $\boldsymbol{\Theta}^\star$ with some quantifiable uncertainty.

Chapter 3

Parametric functions

Chapter 2 described the probabilistic model. In what follows, we derive the parametric functions that are entailed by the model.

3.1 System lifetime

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

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

$$R_{\mathrm{S}}\left(\mathrm{t} \mid \boldsymbol{\Theta}^{\star}\right) = \prod_{j=1}^{m} R_j\left(\mathrm{t} \mid \boldsymbol{\Theta}^{\star}_{\mathrm{j}}\right) , \tag{3.1}$$

where $R_j(\,\cdot\,)$ is the reliability function of the j^{th} component lifetime.

Proof. By definition 2.7, the reliability function is given by

$$R_{\mathrm{S}}\left(\mathrm{t} \mid \boldsymbol{\Theta}^{\star}\right) = \Pr\left[\mathrm{S} > \mathrm{t}\right] . \tag{a}$$

By eq. (2.11), $\mathrm{S} = \min(\mathrm{T}_1, \,, \ldots, \mathrm{T}_m)$. Performing this substitution yields

$$R_{\mathrm{S}}(\mathrm{t} \mid \boldsymbol{\Theta}^{\star}) = \Pr[\min(\mathrm{T}_1, \,, \ldots, \mathrm{T}_{\mathrm{m}}) > \mathrm{t}] . \tag{b}$$

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

$$R_{\mathrm{S}}(\mathrm{t} \mid \boldsymbol{\Theta}^{\star}) = \Pr[\mathrm{T}_1 > \mathrm{t}, \ldots, \mathrm{T}_{\mathrm{m}} > \mathrm{t}] . \tag{c}$$

By assumption 2.2, the component lifetimes are independent, thus by the axioms of probability

$$R_{\mathrm{S}}\left(\mathrm{t} \mid \boldsymbol{\Theta}^{\star}\right) = \Pr[\mathrm{T}_1 > \mathrm{t}] \cdots \Pr[\mathrm{T}_{\mathrm{m}} > \mathrm{t}] . \tag{d}$$

By **??**, the probability that $\mathrm{T}_j > \mathrm{t}$ is equivalent to the reliability function of the j^{th} component lifetime evaluated at t. Performing these substitutions results in

$$R_{\mathrm{S}}\left(t \mid \boldsymbol{\Theta}^{\star}\right) = \prod_{j=1}^{m} R_j(t \mid \boldsymbol{\Theta}^{\star}_j) . \tag{e}$$

□

The probability density function of the system's lifetime is given by the following theorem.

Theorem 3.2. *The lifetime of the m-out-of-m system has a probability density function given by*

$$f_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^{\star}) = \sum_{j=1}^{m} \left(f_j(t \mid \boldsymbol{\Theta}_j^{\star}) \prod_{\substack{k=1 \\ k \neq j}}^{m} R_k(t \mid \boldsymbol{\Theta}_k^{\star}) \right) , \tag{3.2}$$

where $R_k(\,\cdot\,)$ and $f_j(\,\cdot\,)$ are respectively the reliability function and probability density function of the j^{th} component lifetime.

Proof. By **??**,

$$f_{\mathrm{S}}\left(t \mid \boldsymbol{\Theta}^{\star}\right) = -\frac{\mathrm{d}}{\mathrm{d}t} R_{\mathrm{S}}\left(t \mid \boldsymbol{\Theta}^{\star}\right) . \tag{a}$$

By theorem 3.1, this is equivalent to

$$f_{\mathrm{S}}\left(t \mid \boldsymbol{\Theta}^{\star}\right) = -\frac{\mathrm{d}}{\mathrm{d}t} \prod_{j=1}^{m} R_j\left(t \mid \boldsymbol{\Theta}_j^{\star}\right) . \tag{b}$$

By the product rule, this is equivalent to

$$\begin{aligned} f_{\mathrm{S}}(t) = & \frac{\mathrm{d}}{\mathrm{d}t} R_1\left(t \mid \boldsymbol{\Theta}_1^{\star}\right) \prod_{j=2}^{m} R_j\left(t \mid \boldsymbol{\Theta}_j^{\star}\right) \\ & - R_1\left(t \mid \boldsymbol{\Theta}_1^{\star}\right) \frac{\mathrm{d}}{\mathrm{d}t} \prod_{j=2}^{m} R_j\left(t \mid \boldsymbol{\Theta}_j^{\star}\right) . \end{aligned} \tag{c}$$

By **??**, we can substitute $-\frac{\mathrm{d}}{\mathrm{d}t} R_1\left(t \mid \boldsymbol{\Theta}_1^{\star}\right)$ with $f_1\left(t \mid \boldsymbol{\Theta}_1^{\star}\right)$, resulting in

$$\begin{aligned} f_{\mathrm{S}}(t) = & f_1(t) \prod_{j=2}^{m} R_j\left(t \mid \boldsymbol{\Theta}_j^{\star}\right) \\ & - R_1\left(t \mid \boldsymbol{\Theta}_1^{\star}\right) \frac{\mathrm{d}}{\mathrm{d}t} \prod_{j=2}^{m} R_j\left(t \mid \boldsymbol{\Theta}_j^{\star}\right) . \end{aligned} \tag{d}$$

Applying the product rule again results in

$$\begin{aligned} f_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^{\star}) = & f_1\left(t \mid \boldsymbol{\Theta}_1^{\star}\right) \prod_{j=2}^{m} R_j\left(t \mid \boldsymbol{\Theta}_j^{\star}\right) \\ & + f_2\left(t \mid \boldsymbol{\Theta}_2^{\star}\right) \prod_{\substack{j=1 \\ j \neq 2}}^{m} R_j\left(t \mid \boldsymbol{\Theta}_j^{\star}\right) \\ & - R_1\left(t \mid \boldsymbol{\Theta}_1^{\star}\right) R_2\left(t \mid \boldsymbol{\Theta}_2^{\star}\right) \frac{\mathrm{d}}{\mathrm{d}t} \prod_{j=3}^{m} R_j\left(t \mid \boldsymbol{\Theta}_j^{\star}\right) . \end{aligned} \tag{e}$$

We see a pattern emerge. Applying the product rule $m-1$ times results in

$$\begin{aligned} f_{\mathrm{S}}\left(t \mid \boldsymbol{\Theta}^{\star}\right) &= \sum_{j=1}^{m-1} f_j\left(t \mid \boldsymbol{\Theta}_j^{\star}\right) \prod_{\substack{k=1 \\ k \neq j}}^{m} R_k\left(t \mid \boldsymbol{\Theta}_k^{\star}\right) \\ &\quad - \prod_{j=1}^{m-1} R_j\left(t \mid \boldsymbol{\Theta}_j^{\star}\right) \frac{\mathrm{d}}{\mathrm{d}t} R_m\left(t \mid \boldsymbol{\Theta}_m^{\star}\right) \qquad \text{(f)} \\ &= \sum_{j=1}^{m} f_j\left(t \mid \boldsymbol{\Theta}_j^{\star}\right) \prod_{\substack{k=1 \\ k \neq j}}^{m} R_k\left(t \mid \boldsymbol{\Theta}_k^{\star}\right) . \end{aligned}$$

□

The failure rate function simplifies some of the subsequent material and is given by the following definition.

Definition 3.1. *A random variable* X *with a probability density function* $f_{\mathrm{X}}(\,\cdot\,)$ *and reliability function* $R_{\mathrm{X}}(\,\cdot\,)$ *has a failure rate function given by*

$$h_{\mathrm{X}}(t) = \frac{f_{\mathrm{X}}(t)}{R_{\mathrm{X}}(t)} . \tag{3.3}$$

The failure rate function of the system lifetime is given by the following theorem.

Theorem 3.3. *The system lifetime has a failure rate function given by*

$$h_{\mathrm{S}}(\mathrm{t} \mid \boldsymbol{\Theta}^{\star}) = \sum_{\mathrm{j}=1}^{\mathrm{m}} \mathrm{h_j}(\mathrm{t} \mid \boldsymbol{\Theta}_{\mathrm{j}}^{\star}) \tag{3.4}$$

where $h_j(\,\cdot\,)$ *is the failure rate function of the* j^{th} *component lifetime.*

Proof. By eq. (3.3), the failure rate function for the system is given by

$$h_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^{\star}) = \frac{f_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^{\star})}{R_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^{\star})} . \tag{a}$$

Plugging in these parametric functions results in

$$h_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^{\star}) = \frac{\sum_{j=1}^{m}\left(f_j(t \mid \boldsymbol{\Theta}_j^{\star}) \prod_{\substack{k=1 \\ k \neq j}}^{m} R_k(t \mid \boldsymbol{\Theta}_k^{\star})\right)}{\prod_{j=1}^{m} R_j(t \mid \boldsymbol{\Theta}_j^{\star})} , \tag{b}$$

which can be simplified to

$$\begin{aligned} h_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^{\star}) &= \sum_{j=1}^{m} f_j(t \mid \boldsymbol{\Theta}_j^{\star}) / R_j(t \mid \boldsymbol{\Theta}_j^{\star}) & \text{(c)} \\ &= \sum_{j=1}^{m} h_j(t \mid \boldsymbol{\Theta}_j^{\star}) . & \text{(d)} \end{aligned}$$

□

3.2 Cause of system failure

According to definition 2.9, the component responsible for a system failure is given by the discrete random variable K. The joint likelihood that component k is the cause of a system failure occurring at time t is given by the following theorem.

Theorem 3.4. *The joint probability density function of random variable* K *and random system lifetime* S *is given by*

$$f_{\mathrm{K,S}}(k, t \mid \boldsymbol{\Theta}^\star) = f_k(t \mid \boldsymbol{\Theta}^\star) \prod_{\substack{j=1 \\ j \neq k}}^{m} R_j(t \mid \boldsymbol{\Theta}^\star)\,, \tag{3.5}$$

where $f_k(\,\cdot \mid \boldsymbol{\Theta}_k^\star)$ is the probability density function of component k and $R_j(\,\cdot \mid \boldsymbol{\Theta}_k^\star)$ is the reliability function of component k.

Proof. See appendix A.1 for a detailed proof. However, a quick justification follows. The likelihood that component k is the cause of a system failure at time t is equivalent to the likelihood that component k fails at time t and the other components do not fail before time t. According to assumption 2.2, the random variables $\mathrm{T}_1, \ldots, \mathrm{T}_m$ are mutually independent, therefore the likelihood is the product of the probability density function $f_k(t \mid \boldsymbol{\Theta}_k^\star)$ and the reliability functions $R_j(t \mid \boldsymbol{\Theta}_j^\star)$ for $j \in \{1, \ldots, m\} \setminus \{k\}$. $\square$

By definition 2.9, if a system failure occurs at time t, then each component has a particular probability of being the cause as given by the following theorem.

Theorem 3.5. *The conditional probability mass function $p_{\mathrm{K}\,|\,\mathrm{S}}(k \mid t, \boldsymbol{\Theta}^\star)$ maps each component $k \in \{1, \ldots, m\}$ to the probability that component k is the cause of the system failure at time t and is given by*

$$p_{\mathrm{K}\,|\,\mathrm{S}}\left(k \mid t, \boldsymbol{\Theta}^\star\right) = \frac{h_k\left(t \mid \boldsymbol{\Theta}_k^\star\right)}{h_{\mathrm{S}}\left(t \mid \boldsymbol{\Theta}^\star\right)}\,, \tag{3.6}$$

where $h_j(\,\cdot \mid \boldsymbol{\Theta}_j^\star)$ is the failure rate function of the j^{th} component and $h_{\mathrm{S}}(\,\cdot \mid \boldsymbol{\Theta}^\star)$ is the failure rate function of the system.

Proof. By the axioms of probability,

$$p_{\mathrm{K}\,|\,\mathrm{S}}(k, \mid t, \boldsymbol{\Theta}^\star) = \frac{f_{\mathrm{K,S}}(k, t \mid \boldsymbol{\Theta}^\star)}{f_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^\star)}\,, \tag{a}$$

thus we wish to show that

$$\frac{h_k(t \mid \boldsymbol{\Theta}_k^\star)}{h_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^\star)} = \frac{f_{\mathrm{K,S}}(k, t \mid \boldsymbol{\Theta}^\star)}{f_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^\star)}\,. \tag{b}$$

By theorem 3.3, the system's failure rate function is given by

$$h_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^\star) = \sum_{j=1}^{m} h_j(t \mid \boldsymbol{\Theta}_j^\star)\,. \tag{3.4 revisited}$$

Performing this substitution results in

$$\frac{h_k(t \mid \boldsymbol{\Theta}_k^\star)}{h_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^\star)} = \frac{h_k(t \mid \boldsymbol{\Theta}_k^\star)}{\sum_{j=1}^{m} h_j(t \mid \boldsymbol{\Theta}_j^\star)}\,. \tag{c}$$

By definition 3.1, the failure rate function of the p^{th} component is given by

$$h_p(t \mid \boldsymbol{\Theta}_p^\star) = \frac{f_p(t \mid \boldsymbol{\Theta}_p^\star)}{R_p(t \mid \boldsymbol{\Theta}_p^\star)} \,. \tag{d}$$

Performing this substitution results in

$$\frac{h_k(t \mid \boldsymbol{\Theta}_k^\star)}{h_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^\star)} = \frac{f_k(t \mid \boldsymbol{\Theta}_k^\star)}{R_k(t \mid \boldsymbol{\Theta}_k^\star)} \cdot \left[\underbrace{\frac{f_1(t \mid \boldsymbol{\Theta}_1^\star)}{R_1(t \mid \boldsymbol{\Theta}_1^\star)} + \cdots + \frac{f_m(t \mid \boldsymbol{\Theta}_m^\star)}{R_m(t \mid \boldsymbol{\Theta}_m^\star)}}_{A} \right]^{-1} \,. \tag{e}$$

To combine the fractions labeled A in the above equation, we make their respective denominators the same, resulting in

$$\frac{h_k(t \mid \boldsymbol{\Theta}_k^\star)}{h_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^\star)} = \underbrace{\frac{f_k(t \mid \boldsymbol{\Theta}_k^\star)}{R_k(t \mid \boldsymbol{\Theta}_k^\star)}}_{B} \cdot \frac{\overbrace{\prod_{j=1}^{m} R_j(t \mid \boldsymbol{\Theta}^\star)}^{C}}{\sum_{j=1}^{m} \left[\prod_{\substack{p=1 \\ p \neq j}}^{m} R_p(t \mid \boldsymbol{\Theta}_p^\star) \right] f_j(t \mid \boldsymbol{\Theta}_j^\star)} \,. \tag{f}$$

Dividing the part of the above equation labeled C in the numerator by the part labeled B in the denominator results in

$$\frac{h_k(t \mid \boldsymbol{\Theta}_k^\star)}{h_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^\star)} = \frac{f_k(t \mid \boldsymbol{\Theta}_k^\star) \prod_{\substack{j=1 \\ j \neq k}}^{m} R_j(t \mid \boldsymbol{\Theta}_j^\star)}{\sum_{j=1}^{m} f_j(t \mid \boldsymbol{\Theta}_j^\star) \prod_{\substack{p=1 \\ p \neq j}}^{m} R_p(t \mid \boldsymbol{\Theta}_p^\star)} \,. \tag{g}$$

By eq. (3.2), the denominator in the above equation is the density $f_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^\star)$ and, by eq. (3.5), the numerator is the joint density $f_{\mathrm{K,S}}(k, t \mid \boldsymbol{\Theta}^\star)$. Performing these substitutions results in

$$\frac{h_k(t \mid \boldsymbol{\Theta}_k^\star)}{h_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^\star)} = \frac{f_{\mathrm{K,S}}(k, t \mid \boldsymbol{\Theta}^\star)}{f_{\mathrm{S}}(t \mid \boldsymbol{\Theta}^\star)} \,. \tag{h}$$

□

Corollary 3.5.1. *The joint probability density function of random variable* K *and random system lifetime* S *is given by*

$$f_{\mathrm{K,S}}\left(k, t \mid \boldsymbol{\Theta}^\star\right) = h_k\left(t \mid \boldsymbol{\Theta}_k^\star\right) R_{\mathrm{S}}\left(t \mid \boldsymbol{\Theta}^\star\right) \,. \tag{3.7}$$

The proof of eq. (3.7) immediately follows from eqs. (3.5) and (3.6).

3.3 Candidate sets

According to definition 2.13, the random set $\mathbf{C}$ denotes a subset of components in which one of the components is responsible for the system failure.

By assumption 2.5, the distribution of W is not dependent on the true parameter index $\boldsymbol{\Theta}^\star$.

Definition 3.2. *The probability that* $\mathrm{W} = |\mathbf{C}| = w$ *is given by the marginal probability mass function* $p_{\mathrm{W}}(w)$.

By assumption 2.5, the random lifetime S and the cardinality of the random candidate set $\mathrm{W} = |\mathbf{C}|$ are independent, thus by the axioms of probability the joint density of $\mathbf{C}$, S, and W is given by

$$f_{\mathbf{C},\mathrm{S},\mathrm{W}}(\mathbf{c}, t, w \,|\, \boldsymbol{\Theta}^\star) = p_{\mathbf{C}\,|\,\mathrm{S},\mathrm{W}}(\mathbf{c} \,|\, t, w, \boldsymbol{\Theta}^\star) p_{\mathrm{W}}(w) f_{\mathrm{S}}(t \,|\, \boldsymbol{\Theta}^\star)\,. \tag{3.8}$$

If it is given that $\mathrm{W} = w$, by the axioms of probability

$$f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}(\mathbf{c}, t \,|\, w, \boldsymbol{\Theta}^\star) = p_{\mathbf{C}\,|\,\mathrm{S},\mathrm{W}}(\mathbf{c} \,|\, t, w, \boldsymbol{\Theta}^\star) f_{\mathrm{S}}(t \,|\, \boldsymbol{\Theta}^\star)\,. \tag{3.9}$$

The indicator function is needed for subsequent material and is given by the following definition.

Definition 3.3 (Indicator function)**.** *Given a subspace* $\mathbf{E}$ *of a space* $\mathbf{A}$*, the indicator function* $\mathbb{1}_{\mathbf{E}} \colon \mathbf{A} \mapsto \{\mathbf{0}, \mathbf{1}\}$ *is equal to* 1 *if* $x \in \mathbf{E}$ *and* 0 *otherwise.*

Definition 3.4. *The conditional random variable* $\mathbf{C} \,|\, \mathrm{W} = w$ *has a sample space given by*

$$\boldsymbol{\mathcal{C}}_{\boldsymbol{w}} \equiv \left\{ \mathbf{c} \in \boldsymbol{\mathcal{C}} \,|\, |\mathbf{c}| = w \right\}\,, \tag{3.10}$$

the set of all candidate sets of cardinality w*.*

The conditional probability density of $\mathbf{C}$ and S given $\mathrm{W} = w$ is given by the following theorem.

Theorem 3.6. *The joint conditional probability density of* $\mathbf{C}$ *and* S *given* $\mathrm{W} = w$ *is given by*

$$f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}(\mathbf{c}, t \,|\, w, \boldsymbol{\Theta}^\star) = \eta R_{\mathrm{S}}\left(t \,|\, \boldsymbol{\Theta}^\star\right) \sum_{k \in \mathbf{c}} h_k\left(t \,|\, \boldsymbol{\Theta}^\star\right) \mathbb{1}_{\boldsymbol{\mathcal{C}}_{\boldsymbol{w}}}(\mathbf{c})\,. \tag{3.11}$$

where η *is the normalizing constant that makes the joint probability density function integrate*[1] *to unity over the sample space* $\boldsymbol{\mathcal{C}}_{\boldsymbol{w}} \times (0, \infty)$*.*

Proof. Given a candidate set $\mathbf{c} \in \boldsymbol{\mathcal{C}}_{\boldsymbol{w}}$, the event E we wish to compute the likelihood of is given by

$$E = \bigcup_{j \in \mathbf{c}} (\mathrm{K} = j \cap \mathrm{S} = t)\;. \tag{a}$$

We would like to compute the likelihood $L(E)$. The likelihood that $\mathrm{K} = j$ and $\mathrm{S} = t$ is given by the probability density function

$$f_{\mathrm{K},\mathrm{S}}\left(k, t \,|\, \boldsymbol{\Theta}^\star\right) = h_k\left(t \,|\, \boldsymbol{\Theta}_k^\star\right) R_{\mathrm{S}}\left(t \,|\, \boldsymbol{\Theta}^\star\right)\;. \tag{3.7 revisited}$$

Performing this substitution yields

$$L(E) = \bigcup_{j \in \mathbf{c}} h_j\left(t \,|\, \boldsymbol{\Theta}_j^\star\right) R_{\mathrm{S}}\left(t \,|\, \boldsymbol{\Theta}^\star\right)\;. \tag{b}$$

Since the events $\mathrm{K} = j, \mathrm{S} = t$ for $j \in \mathbf{c}$ are mutually exclusive, the likelihood is given by

$$L(E) = \sum_{j \in \mathbf{c}} h_j\left(t \,|\, \boldsymbol{\Theta}_j^\star\right) R_{\mathrm{S}}\left(t \,|\, \boldsymbol{\Theta}^\star\right)\;. \tag{c}$$

Since outcomes in the sample space $\boldsymbol{\mathcal{C}}_{\boldsymbol{w}}$ are not atomic with respect to the sample space of $\mathrm{K} \in \{1, \ldots, m\}$, we multiply by a normalizing constant η to make it a proper joint probability density function. □

[1] Integration refers to both summations and integrations.

The normalizing constant η indicated in eq. (3.11) is given by the following theorem.

Theorem 3.7. *The normalizing constant* η *that makes the joint probability density* $f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}(\mathbf{c}, t \,|\, w, \boldsymbol{\Theta}^\star)$ *integrate to unity over the sample space* $\boldsymbol{\mathcal{C}_w} \times (0, \infty)$ *is given by*

$$\eta = \frac{1}{\binom{m-1}{w-1}} \,. \tag{3.12}$$

Proof. Integrating over the sample space $\boldsymbol{\mathcal{C}_w} \times (0, \infty)$ must be equal to unity. That is,

$$\eta \sum_{\mathbf{c} \in \boldsymbol{\mathcal{C}_w}} \int_0^\infty f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}(\mathbf{c}, t \,|\, \boldsymbol{\Theta}^\star)\, \mathrm{d}t = 1 \,. \tag{a}$$

Substituting the joint probability density with the definition given by eq. (3.11) results in

$$\eta \sum_{\mathbf{c} \in \boldsymbol{\mathcal{C}_w}} \int_0^\infty \left[\sum_{j \in \mathbf{c}} f_{\mathrm{K},\mathrm{S}}(j, t \,|\, \boldsymbol{\Theta}^\star) \right] \mathrm{d}t = 1 \,. \tag{b}$$

We can switch the summation over $\boldsymbol{\mathcal{C}_w}$ and the integration over t without changing the equation, resulting in

$$\eta \sum_{\mathbf{c} \in \boldsymbol{\mathcal{C}_w}} \sum_{j \in \mathbf{c}} \left[\int_0^\infty f_{\mathrm{K},\mathrm{S}}(j, t \,|\, \boldsymbol{\Theta}^\star) \mathrm{d}t \right] = 1 \,. \tag{c}$$

By the axioms of probability, the inner integral is the marginal probability mass function of K. Performing this substitution results in

$$\eta \sum_{\mathbf{c} \in \boldsymbol{\mathcal{C}_w}} \sum_{j \in \mathbf{c}} p_{\mathrm{K}}(j \,|\, \boldsymbol{\Theta}^\star) = 1 \,. \tag{d}$$

When constructing a candidate set of cardinality w, suppose we restrict our attention to only those sets in which component r is present. Then, we have $w - 1$ additional components to choose from of the $m - 1$ remaining. There are $\binom{m-1}{w-1}$ such choices possible, thus component r is present in $\binom{m-1}{w-1}$ of the sets. Consequently, each component has $\binom{m-1}{w-1}$ occurrences. Performing this substitution results in

$$\eta \binom{m-1}{w-1} \sum_{j=1}^{m} p_{\mathrm{K}}(j \,|\, \boldsymbol{\Theta}^\star) = 1 \,. \tag{e}$$

Observe that $\sum_{j=1}^{m} p_{\mathrm{K}}(j \,|\, \boldsymbol{\Theta}^\star)$ must be equal to unity since we are summing over the entire sample space of K. Performing this substitution and solving for η yields

$$\eta = \frac{1}{\binom{m-1}{w-1}} \,. \tag{f}$$

□

Corollary 3.7.1. *The conditional probability mass function* $p_{\mathbf{C}\,|\,\mathrm{S},\mathrm{W}}(\mathbf{c} \,|\, t, w, \boldsymbol{\Theta}^\star)$ *maps each candidate set* $\mathbf{c}$ *to the probability that the given candidate set contains the component responsible for the system failure at time* t*. By the axioms of probability,*

$$p_{\mathbf{C}\,|\,\mathrm{S},\mathrm{W}}(\mathbf{c} \,|\, t, w, \boldsymbol{\Theta}^\star) = \frac{\sum_{j \in \mathbf{c}} h_j \left(t \,|\, \boldsymbol{\Theta}_j^\star \right)}{h_{\mathrm{S}} \left(t \,|\, \boldsymbol{\Theta}^\star \right)} \mathbb{1}_{\boldsymbol{\mathcal{C}_w}}(\mathbf{c}) \,, \tag{3.13}$$

where $h_{\mathrm{S}}(\,\cdot \,|\, \boldsymbol{\Theta}^\star)$ *and* $h_j(\,\cdot \,|\, \boldsymbol{\Theta}_j^\star)$ *are the failure rate functions of the system and the* j^{th} *component respectively.*

Proof. By the axioms of probability,

$$p_{\mathbf{C}\,|\,\mathrm{S},\mathrm{W}}(\mathbf{c}\,|\,t, w, \mathbf{\Theta}^\star) = \frac{f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}(\mathbf{c}, t\,|\,w, \mathbf{\Theta}^\star)}{f_{\mathrm{S}}(t\,|\,\mathbf{\Theta}^\star)}\,, \tag{a}$$

where $f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}(\,\cdot\,|\,\mathbf{\Theta}^\star)$ is the joint probability density function given by theorem 3.6. Making this substitution immediately yields the result. □

3.4 Masked system failure times

Definition 3.5. *A random sample of n masked system failure times in which it is given that each system failure has w candidates, i.e.,* $\mathrm{S}_i, \mathbf{C}_i\,|\,\mathrm{W} = w$ *for* $i = 1, \ldots, n$*, is denoted by* $\mathbf{M}(w, n)$*. That is, if given a random sample of masked system failure times, the conditional distribution* $\mathbf{M}(w, n)$ *is the subset of the random sample with w candidates.*

The likelihood of observing a particular realization, $\mathbf{M}(w, n) = \mathbf{m_n}$, is given by the following definition.

Theorem 3.8. *A random sample of n masked system failure times conditioned on candidate sets of cardinality w has a joint probability density given by*

$$f_{\mathbf{M}(w,n)}\left(\mathbf{m_n}\,|\,w, \mathbf{\Theta}^\star\right) = \prod_{i=1}^{n} f_{\mathrm{S}}(t_i\,|\,\mathbf{\Theta}^\star) \cdot p_{\mathbf{C}\,|\,\mathrm{S},\mathrm{W}}\left(\mathbf{c}_i\,|\,t_i, w, \mathbf{\Theta}^\star\right)\,, \tag{3.14}$$

where $f_{\mathrm{S}}(\cdot\,|\,\mathbf{\Theta}^\star)$ *is the marginal density given by theorem 3.2 and* $p_{\mathbf{C}\,|\,\mathrm{S},\mathrm{W}}(\cdot\,|\,t, w, , \mathbf{\Theta}^\star)$ *is the conditional probability mass given by corollary 3.7.1.*

The reason why a candidate set of cardinality w is generated is outside the scope of the statistical model and where necessary, the relative frequency in a sample is used.

The generative model of $\mathbf{M}(w, n)$, which directly follows from eq. (3.14), is given by algorithm 1.

Algorithm 1: Generative model of a masked system failure time with w candidates.

Result: a realization of a masked system failure time with w candidates.

Input:

$\mathbf{\Theta}^\star$, the true parameter value.

w, the cardinality of the candidate sets.

Output:

$\mathbf{m}$, a realization of $\mathbf{C}, \mathrm{S}\,|\,\mathrm{W} = w$.

Model `GenerateMSFT`$(w, \mathbf{\Theta}^\star)$

  draw a system lifetime $t \sim f_{\mathrm{S}}(\,\cdot\,|\,\mathbf{\Theta})$

  draw a candidate set $\mathbf{c} \sim p_{\mathbf{C}\,|\,\mathrm{S},\mathrm{W}}(\,\cdot\,|\,t, w, \mathbf{\Theta})$

  $\mathbf{m} \leftarrow (t, \mathbf{c})$

  return m

Chapter 4

Information about component failure

Knowing the distribution of component lifetimes has many important applications. The canonical use-case is, if a failure occurs in a series system at time t, for any component we know the probability that it is the cause of the system failure.

However, there are other applications as well. For instance, a system designer may desire a system where the component cause of failure is as uncertain as possible to prevent an *adversary* from learning about critical vulnerabilities. Informally, *entropy* is a fundamental measure of uncertainty.

By definition, $\mathrm{K} = j$ given $\mathrm{S} = t$ is the probability that component j fails before the other $m-1$ components given a system failure at time t.

If a diagnostician could ask "yes" or "no" questions about the component cause of failure, an *expected lower-bound* on the number of questions required to determine which component caused the failure is given by

$$\mathrm{H}(\mathrm{K}) = -\sum_{k \in \{1,\ldots,m\}} p_{\mathrm{K}}(k) \log_2 p_{\mathrm{K}}(k) \,. \tag{4.1}$$

This is known as the *entropy* of the random variable K, which quantifies the amount of "uncertainty" about what value K realizes.

The greater the entropy, the greater the number of questions the diagnostician needs on average to determine the component cause of failure. Each component may have a different distribution of lifetimes, and thus knowing when a system failure occurs may *reduce* the uncertainty about the component cause of failure,

$$\mathrm{H}(\mathrm{K} \mid \mathrm{S}) \leq \mathrm{H}(\mathrm{K}) \,. \tag{4.2}$$

If K and S are *statistically independent*, then

$$\mathrm{H}\left(\mathrm{K} \mid \mathrm{S}\right) = \mathrm{H}\left(\mathrm{K}\right) \,, \tag{4.3}$$

that is, observing the system failure time does not reduce the uncertainty.

The joint entropy of K and S is given by

$$\mathrm{H}\left(\mathrm{K}, \mathrm{S}\right) = \mathrm{H}\left(\mathrm{K} \mid \mathrm{S}\right) + \mathrm{H}\left(\mathrm{S}\right) \,. \tag{4.4}$$

The *information* conveyed by a system failure at time t is the *reduction in uncertainty* about the component failure. The information is given by

$$\mathrm{I}(\mathrm{K} \mid \mathrm{S}) = \mathrm{H}(\mathrm{K}) - \mathrm{H}(\mathrm{K} \mid \mathrm{S}) \,. \tag{4.5}$$

If K and S are independent, then $\mathrm{I}(\mathrm{K} \mid \mathrm{S}) = 0$, i.e., there is no reduction in uncertainty about K given that a system failure has been observed at a particular time.

The probabilistic model of the component responsible for a system failure at a given time described in **??** has an *entropy* given by

$$\mathrm{H}\big(\mathrm{K} \,|\, \mathrm{S}\big) \,. \tag{4.6}$$

Given that the system failure time has a support $(0, \infty)$ and the mean time to system failure is $\mathbb{E}[\mathrm{S}] = \frac{1}{\lambda}$, the *maximum entropy distribution* for the system lifetime is exponentially distributed with a failure rate λ,

$$\mathrm{S} \sim \mathrm{EXP}(\lambda) \,. \tag{4.7}$$

$$\frac{\lambda_j}{\sum_{p=1}^{m} \lambda_p} \tag{4.8}$$

where $\lambda_1 + \cdots + \lambda_p = \lambda$.

Consider a series system of m components. The entropy is minimized if $m-1$ components have infinite expect Any degenerate distribution T

Chapter 5

Likelihood and Fisher information

In what follows, we frequently prefer to work with vectors rather than matrices.

Definition 5.1. *Let* vec: $\mathbb{R}^{m \times q} \mapsto \mathbb{R}^{m \cdot q}$ *denote the linear transformation of a matrix of m rows and q columns into a column vector of $m \cdot q$ elements, where consecutive matrix rows are placed side by side.*

Notation. A matrix parameter index $\mathbf{\Theta}$ has a vector representation denoted by

$$\boldsymbol{\theta} \equiv \operatorname{vec}(\mathbf{\Theta}) \,, \tag{5.1}$$

thus the true matix parameter index $\mathbf{\Theta}^\star$ has a vector representation denoted by $\boldsymbol{\theta}^\star$. We use matrix and vector representations interchangeably.

The subsequent material makes a few assumptions, denoted *regularity conditions*, given by the following.

Assumption 5.1. *The following regularity conditions hold for the $\mathcal{F}_{\mathbf{\Theta}}$-family of the series system:*

1. *$\boldsymbol{\theta}^\star$ is interior to the parameter support Ω.*
2. *The log-likelihood ℓ is continuous, thrice differentiable, and bounded.*
3. *The true parameter value $\boldsymbol{\theta}^\star$ is identified, i.e.,*

$$\boldsymbol{\theta}^\star = \operatorname*{arg\,max}_{\boldsymbol{\theta} \in \Omega} \mathbb{E}_{\boldsymbol{\theta}^\star}\Big(\ln f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}\left(\mathbf{C}, \mathrm{S} \,|\, w, \boldsymbol{\theta}\right)\Big) \,. \tag{5.2}$$

By theorem 3.8, the probability density that $\mathbf{M}(w,n) = \mathbf{m_n}$ is given by

$$f_{\mathbf{M}(w,n)}(\mathbf{m_n} \,|\, w, \boldsymbol{\theta}^\star) = \prod_{i=1}^{n} f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}(\mathbf{c}_i, t_i \,|\, w, \boldsymbol{\theta}^\star) \,. \tag{3.14 revisited}$$

If we fix $\mathbf{m_n}$ and allow the parameter $\boldsymbol{\theta}$ to change, we have the likelihood function.

Definition 5.2. *The likelihood function represents the* likelihood *of parameter index $\boldsymbol{\theta}$ conditioned on $\mathbf{M}(w,n) = \mathbf{m_n}$ is given by*

$$\mathcal{L}(\boldsymbol{\theta} \,|\, w, \mathbf{m_n}) = \prod_{i=1}^{n} f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}(\mathbf{c}_i, t_i \,|\, w, \boldsymbol{\theta}) \,. \tag{5.3}$$

The likelihood is a measure of the extent to which the given sample provides support for particular values of a parameter in $\mathcal{F}_{\boldsymbol{\Theta}}$.

The surface of the likelihood function captures all the information there is about $\boldsymbol{\theta}^\star$ in a sample. Any statistic that is equivalent to the likelihood function is a *sufficient* statistic. The likelihood function is a minimally sufficient statistic.

The log-likelihood plays a more central role than the likelihood for reasons that will be made clear later on. The log-likelihood function is given by the following theorem.

Definition 5.3. *The log-likelihood with respect to $\boldsymbol{\theta}$ given a particular $\mathbf{M}(w, n) = \mathbf{m_n}$ is given by*

$$\ell(\boldsymbol{\theta} \mid w, \mathbf{m_n}) = \sum_{i=1}^{n} \ln f_{\mathbf{C},\mathrm{S} \mid \mathrm{W}}(\mathbf{c}_i, t_i \mid w, \boldsymbol{\theta})\,. \tag{5.4}$$

In subsequent material, we need the gradient operator as given by the following definition.

Definition 5.4 (Gradient operator)**.** *The gradient of a function* $\mathrm{g}\colon \mathbb{R}^q \mapsto \mathbb{R}$ *with respect to column vector* $\mathbf{x} \in \mathbb{R}^q$ *is given by*

$$\nabla_{\mathbf{x}}\, \mathrm{g}\,(\mathbf{x}) = \left(\tfrac{\partial\, \mathrm{g}(\mathbf{x})}{\partial x_1}, \ldots, \tfrac{\partial\, \mathrm{g}(\mathbf{x})}{\partial x_q}\right)^{\intercal}. \tag{5.5}$$

The score function is given by the following definition.

Definition 5.5 (Score function)**.** *The gradient of the log-likelihood function with respect to $\boldsymbol{\theta}$ is the Fisher's score function is denoted by* $\mathrm{s}\colon \mathbb{R}^{m \cdot q} \mapsto \mathbb{R}^{m \cdot q}$*. That is,*

$$\mathrm{s}\left(\boldsymbol{\theta} \mid w, \mathbf{m_n}\right) = \nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta} \mid w, \mathbf{m_n})\,. \tag{5.6}$$

The variance-covariance is given by the following definition.

Definition 5.6 (Variance-covariance)**.** *A random vector $\mathbf{X} \in \mathbb{R}^p$ has a variance-covariance given by*

$$\mathrm{Var}(\mathbf{X}) = \mathbb{E}\left[\left(\mathbf{X} - \mathbb{E}[\mathbf{X}]\right)\left(\mathbf{X} - \mathbb{E}[\mathbf{X}]\right)^{\intercal}\right], \tag{5.7}$$

which is a p-by-p semi-positive definite matrix.

The score function when given a particular realization $\mathbf{m_n}$ is a statistic. When we consider it as a function of a random sample, it is a random vector. When evaluated at the true parameter index $\boldsymbol{\theta}^\star$, it has a variance-covariance given by the following definition.

Definition 5.7 (Fisher information matrix)**.** *The score as a function of a random sample $\mathbf{M}(w, n)$ has a variance-covariance known as the Fisher information matrix, a $(m \cdot q)$-by-$(m \cdot q)$ matrix denoted by*

$$\boldsymbol{\mathcal{I}_n}\left(\boldsymbol{\theta}^\star \mid w\right) = \mathrm{Var}_{\boldsymbol{\theta}^\star}\!\left[\mathrm{s}\left(\boldsymbol{\theta}^\star \mid w, \mathbf{M}(w, n)\right)\right], \tag{5.8}$$

where the variance is taken with respect to the joint probability density function of $\mathbf{C}, \mathrm{S} \mid \mathrm{W} = w$ indexed by $\boldsymbol{\theta}^\star$.

The Hessian is given by the following definition.

Definition 5.8 (Hessian operator)**.** *The Hessian of a function* $\mathrm{g}\colon \mathbb{R}^q \mapsto \mathbb{R}$ *with respect to* $\mathbf{x} \in \mathbb{R}^q$ *is a q-by-q matrix where the (j, k)-th element is given by*

$$\mathcal{H}\left(\mathrm{g}(\mathbf{x})\right)_{jk} = \frac{\partial^2\, \mathrm{g}(\mathbf{x})}{\partial x_j \partial x_k}\,. \tag{5.9}$$

Under the regularity conditions, the information matrix may be computed from the Hessian and is given by the following postulate.

Postulate 5.1. *The information matrix is given by the expectation of the negative of the Hessian of the log-likelihood function ℓ evaluated at $\boldsymbol{\theta}^\star$. That is,*

$$\boldsymbol{\mathcal{I}_n}\left(\boldsymbol{\theta}^\star \mid w\right) = -\,\mathbb{E}_{\boldsymbol{\theta}^\star}\left[\mathcal{H}\left(\ell\left(\boldsymbol{\theta} \mid w, \mathbf{M}(w,n)\right)\right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star}\right], \tag{5.10}$$

where the expectation is taken with respect to $\boldsymbol{\theta}^\star$ and the Hessian is taken with respect to $\boldsymbol{\theta}$ and then evaluated at $\boldsymbol{\theta}^\star$.

For clarity, the (i,j)-th element of the information matrix is given by the following definition.

Definition 5.9. *The (i,j)-th element of $\boldsymbol{\mathcal{I}_n}(\,\cdot \mid w)$ is given by*

$$\begin{aligned}\boldsymbol{\mathcal{I}_n}\left(\boldsymbol{\theta}^\star \mid w\right)_{i\,j} = -\sum_{\mathbf{c}\in\boldsymbol{\mathcal{C}_w}}\int_0^\infty \left[\tfrac{\partial^2}{\partial\theta_i\partial\theta_j}\ln f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}\left(\mathbf{c}_i, \mathrm{t_i} \mid \mathrm{w}, \boldsymbol{\theta}\right)\right]\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} \times \\ f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}\left(\mathbf{c}_i, \mathrm{t_i} \mid \mathrm{w}, \boldsymbol{\theta}^\star\right)\mathrm{d}t\,,\end{aligned} \tag{5.11}$$

where $\boldsymbol{\mathcal{C}_w} \times (0,\infty)$ is the sample space.

Notation. We denote the Fisher information matrix of $\boldsymbol{\theta}^\star$ given $\mathbf{M}(w,1)$ by

$$\boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^\star \mid w\right) \equiv \boldsymbol{\mathcal{I}_1}\left(\boldsymbol{\theta}^\star \mid w\right). \tag{5.12}$$

The information matrix is a key quantity in large sample theory; it quantifies the *expected* information (as a reduction of uncertainty) $\mathbf{M}(w,n)$ carries about $\boldsymbol{\theta}^\star$. For sufficiently large samples, the log-likelihood has a Taylor series approximation given by

$$\ell\left(\boldsymbol{\theta} \mid w\right) \approx -\frac{1}{2}\left(\boldsymbol{\theta}-\boldsymbol{\theta}^\star\right)^\intercal \boldsymbol{\mathcal{I}_n}\left(\boldsymbol{\theta}^\star \mid w\right)\left(\boldsymbol{\theta}-\boldsymbol{\theta}^\star\right), \tag{5.13}$$

which is maximized at $\boldsymbol{\theta}^\star$. Intuitively, the more "peaked" the log-likelihood is expected to be at $\boldsymbol{\theta}^\star$, the more information a random sample is expected to carry about $\boldsymbol{\theta}^\star$.

The information about $\boldsymbol{\theta}^\star$ of independent samples is *additive* as given by the following postulate.

Postulate 5.2. *Algebraically, the Fisher information about the true parameter index $\boldsymbol{\theta}^\star$ of independent samples is additive. That is,*

$$\boldsymbol{\mathcal{I}_{n_1+n_2}}\left(\boldsymbol{\theta}^\star \mid \cdot\,\right) = \boldsymbol{\mathcal{I}_{n_1}}(\boldsymbol{\theta}^\star \mid \cdot\,) + \boldsymbol{\mathcal{I}_{n_2}}(\boldsymbol{\theta}^\star \mid \cdot\,)\,. \tag{5.14}$$

By postulate 5.2, the Fisher information of a random sample of n masked system failure times drawn from $\mathbf{M}(w,n)$ is given by

$$\boldsymbol{\mathcal{I}_n}\left(\boldsymbol{\theta}^\star \mid w\right) = n\,\boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^\star \mid w\right). \tag{5.15}$$

A *masked system failure time* in which the cardinality of the candidate set is distributed according to W, as described by **??**, has a Fisher information matrix given by the following theorem.

Theorem 5.1. *The* Fisher information *matrix of true parameter index* $\boldsymbol{\theta}^{\star}$ *with respect to a realization of jointly distributed random system lifetime* $\mathrm{S} = t$ *and candidate set* $\mathbf{C} = \mathbf{c}$ *is given by*

$$\boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^{\star}\right) = \sum_{w=1}^{m-1} p_{\mathrm{W}}(w) \boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^{\star} \,|\, w\right) , \tag{5.16}$$

where $p_{\mathrm{W}}(w)$ *is the marginal distribution of* $\mathrm{W} = |\mathbf{C}|$.

Proof. Suppose we have a random sample of n masked system failure times. By **??**, the expected number of sample points that have w candidates is given by

$$n \cdot p_{\mathrm{W}}(w) . \tag{a}$$

By the additive property of *Fisher information* given by postulate 5.2, the information matrix is given by

$$\boldsymbol{\mathcal{I}_n}\left(\boldsymbol{\theta}^{\star}\right) = \sum_{w=1}^{m-1} n \cdot p_{\mathrm{W}}(w) \boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^{\star} \,|\, w\right) . \tag{b}$$

Setting n to 1 results in

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}^{\star}) = \sum_{w=1}^{m-1} p_{\mathrm{W}}(w) \boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^{\star} \,|\, w\right) . \tag{c}$$

□

Remark. A model is a simplification of reality. The model denoted by $\mathbf{M_n}$ is an approximation of some (unknown) underlying generative model, e.g., W and S may not truly be independent. Since the *expected* information matrix is an expectation with respect to the simplified model, the *expected* information is also a simplified approximation. The *observed* information matrix, given by

$$\boldsymbol{\mathcal{J}_n}(\hat{\boldsymbol{\theta}}_{\mathbf{n}} \,|\, \mathbf{m_n}) = -\,\mathcal{H}\left(\ell\left(\boldsymbol{\theta} \,|\, \mathbf{m_n}\right)\right)\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathbf{n}}} , \tag{5.17}$$

is a statistic of a real sample and therefore makes less of a commitment to the simplified model. Generally, the *observed* information matrix is preferable but since we know (by prescription) the underlying generative model,

$$\boldsymbol{\mathcal{J}_n}\left(\hat{\boldsymbol{\theta}}_{\mathbf{n}} \,|\, \mathbf{M_n}\right) \xrightarrow{P} \boldsymbol{\mathcal{I}_n}(\boldsymbol{\theta}^{\star}) , \tag{5.18}$$

and so we proceed with the *expected* information matrix.

Chapter 6

Sampling distribution of parameter estimators

By definitions 2.10 and 2.11, the random system lifetime S has a probability density function that is a member of the $\mathcal{F}_{\boldsymbol{\Theta}}$-family with a true parameter index $\boldsymbol{\theta}^\star$. Let some estimator of $\boldsymbol{\theta}^\star$ be a function of the information in a random sample of n masked system failure times. The estimator is a function of a random sample, thus it has *sampling distribution*, a random vector denoted by $\mathbf{Y_n}$. That is,

$$\mathbf{Y_n} = \boldsymbol{\psi}\left(\mathbf{M_n}\right) . \tag{6.1}$$

A particular realization of the estimator is denoted by $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$, i.e., $\mathbf{Y_n} = \bar{\boldsymbol{\theta}}_{\mathbf{n}}$. All else being equal, we prefer estimators which vary only *slightly* from sample to sample with a *central tendency* around $\boldsymbol{\theta}^\star$. That is, we prefer unbiased estimators in which each component has small variance.

The sampling distribution of $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$ has a variance-covariance given by $\operatorname{Var}\left(\mathbf{Y_n}\right)$ (see eq. (5.7)) and a bias given by the following definition.

Definition 6.1. *The bias of a point estimator* $\mathbf{Y_n} = \bar{\boldsymbol{\theta}}_{\mathbf{n}}$ *is given by*

$$\operatorname{bias}\left(\mathbf{Y_n}\right) = \mathbb{E}\left[\mathbf{Y_n}\right] - \boldsymbol{\theta}^\star . \tag{6.2}$$

We are often faced with a trade-off between bias and variance. A measure of estimator accuracy that is a function of both the bias and the variance is the mean squared error as given by the following definition.

Definition 6.2 (MSE)**.** *The mean squared error of the sampling distribution of* $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$ *is given by*

$$\operatorname{MSE}\left(\mathbf{Y_n}\right) = \mathbb{E}\left[\left(\mathbf{Y_n} - \boldsymbol{\theta}^\star\right)^\intercal \left(\mathbf{Y_n} - \boldsymbol{\theta}^\star\right)\right] \tag{6.3}$$

An equivalent way to compute the mean squared error is given by the following postulate.

Postulate 6.1. *The mean squared error of the sampling distribution of* $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$ *as given by definition 6.2 is equivalent to*

$$\operatorname{MSE}\left(\mathbf{Y_n}\right) = \operatorname{tr}\left(\operatorname{Var}\left(\mathbf{Y_n}\right)\right) + \operatorname{bias}^2\left(\mathbf{Y_n}\right) , \tag{6.4}$$

where $\operatorname{tr}\left(\boldsymbol{A}\right)$ *computes the sum of the diagonal elements of square matrix* $\boldsymbol{A}$*.*

Fisher information reduces the uncertainty about $\boldsymbol{\theta}^\star$. The *minimum-variance unbiased estimator* (UMVUE) has a lower-bound given by the inverse of the *Fisher information matrix*, denoted the

Cramér-Rao lower-bound. The minimum variance obtainable from a random sample of n masked system failure times drawn from $\mathbf{M_n}$ is given by

$$\mathbf{CRLB_n} = \frac{1}{n}\boldsymbol{\mathcal{I}}^{-1}\left(\boldsymbol{\theta}^\star\right) . \tag{6.5}$$

By eq. (6.4), if $\bar{\boldsymbol{\theta}}_\mathbf{n}$ is an unbiased estimator of $\boldsymbol{\theta}^\star$, then the mean squared error is given by

$$\operatorname{MSE}\left(\mathbf{Y_n}\right) = \operatorname{tr}\left(\operatorname{Var}\left(\mathbf{Y_n}\right)\right) . \tag{6.6}$$

By eqs. (6.5) and (6.6), the mean squared error of any unbiased point estimator of $\boldsymbol{\theta}^\star$ in which the only *a priori* information is given by a random sample $\mathbf{M_n}$ has a lower-bound given by the trace of $\mathbf{CRLB_n}$,

$$\operatorname{MSE}\left(\mathrm{Y}\right) \geq \operatorname{tr}\left(\mathbf{CRLB_n}\right) . \tag{6.7}$$

6.1 Maximum likelihood estimator given w candidates

Suppose the only information about the true parameter index $\boldsymbol{\theta}^\star$ is given by a sample of n masked system failure times in which only those system failures with w candidates are observed. The maximum likelihood estimator based on this conditional sample is given in the following definition.

Definition 6.3. *A parameter index $\hat{\boldsymbol{\theta}}(n \mid w)$ that maximizes the* likelihood *of observing $\mathbf{M}(w, n) = \mathbf{m_n}$ (see definition 3.5) is a maximum likelihood estimator of $\boldsymbol{\theta}^\star$. That is,*

$$\hat{\boldsymbol{\theta}}(n \mid w) = \underset{\boldsymbol{\theta} \in \boldsymbol{\Omega}}{\arg\max}\, \mathcal{L}(\boldsymbol{\theta} \mid w, \mathbf{m_n}) \,, \tag{6.8}$$

where $\boldsymbol{\Omega}$ is the feasible parameter space of the parametric family.

As further justification of assumption 2.6, the maximum likelihood estimator maximizes the likelihood of generating the given $\mathbf{m_n}$ such that for each system failure time, a component in the corresponding candidate set is the cause.

The accuracy of the maximum likelihood estimator is a function of how indicative candidate sets are. For instance, we assume that the candidates are generated from the joint probability density of S and $\mathbf{C}$. If this is not a very accurate model (e.g., the inspector who constructed the candidate sets in the sample preferred to check a subset of the components for failure, and tended to neglect the others), the maximum likelihood estimator will be inconsistent since the model is not consistent with reality. Rather, the maximum likelihood estimator will generate an estimate that makes the provided sample the most likely to be generated by the assumed model, i.e., the component lifetimes will not be accurately modeled since they must be modeled in such a way as to make the biased sample more likely.

The logarithm is a monotonically increasing function, thus the parameter value $\hat{\boldsymbol{\theta}}(n \mid w)$ that maximizes $\mathcal{L}$ also maximizes the log-likelihood ℓ (see definition 5.3), i.e.,

$$\hat{\boldsymbol{\theta}}(n \mid w) = \underset{\boldsymbol{\theta} \in \boldsymbol{\Omega}}{\arg\max}\, \ell(\boldsymbol{\theta} \mid w, \mathbf{m_n}) \,. \tag{6.9}$$

According to Bickel [1], if the parameter support $\boldsymbol{\Omega}$ of the parametric family is open, the log-likelihood ℓ is differentiable with respect to $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}(n \mid w)$ exists, then $\hat{\boldsymbol{\theta}}(n \mid w)$ must be a stationary point of ℓ as given by

$$\mathrm{s}\left(\hat{\boldsymbol{\theta}}(n \mid w) \mid w, \mathbf{m_n}\right) = \mathbf{0} \,, \tag{6.10}$$

where s is the score function given by eq. (5.6). In cases where there are no closed-form solutions to eq. (6.10), iterative methods may be used as described in appendix B.

An estimator of $\boldsymbol{\theta}^\star$ given a sample $\mathbf{M}(w, n) = \mathbf{m_n}$ is the maximum likelihood estimator $\hat{\boldsymbol{\theta}}(n \mid w)$ described in definition 6.3. Since $\hat{\boldsymbol{\theta}}(n \mid w)$ is a function of the random sample $\mathbf{M}(w, n)$, it has a *sampling distribution.*

Definition 6.4. *The* sampling distribution *of $\hat{\boldsymbol{\theta}}(n \mid w)$ is a random vector denoted by $\mathbf{Y}(n \mid w) \in \mathrm{R}^{m \cdot q}$.*

The generative model for the maximum likelihood estimator, `GenerateMLE`$(w \mid \cdot\,)$, is described by algorithm 2. This generative model depends on `GenerateMSFT`, the generative model for the *masked system failure time* as described by algorithm 1. Note that in algorithm 2, line 7 may be approximated with `FindMLE`, a function that numerically solves the stationary points of the maximum likelihood equations as described by algorithm 5.

Algorithm 2: Generative model of the maximum likelihood estimator conditioned on w candidates

Result: a realization of a maximum likelihood estimate from the sampling distribution of $\hat{\boldsymbol{\theta}}(n \mid w)$.

Input:

$\boldsymbol{\theta}^\star$, the true parameter index.

w, the cardinality of the candidate set.

n, the number of masked system failure times.

Output:

$\hat{\boldsymbol{\theta}}(n \mid w)$, a realization of $\mathbf{Y}(n \mid w)$.

1 **Model** `GenerateMLE`$(n, w, \boldsymbol{\theta}^\star)$
2     $\mathbf{m_n} \leftarrow \emptyset$
3   **for** $i \leftarrow 1$ **to** n **do**
4         $\mathbf{M} \leftarrow$ `GenerateMSFT`$(w, \boldsymbol{\theta}^\star)$
5         $\mathbf{m_n} \leftarrow \mathbf{m_n} \cup \{\mathbf{M}\}$
6   **end**
7     $\hat{\boldsymbol{\theta}}(n \mid w) \leftarrow \arg\max_{\boldsymbol{\theta} \in \Omega} \ell\left(\boldsymbol{\theta} \mid w, \mathbf{m_n}\right)$
8   **return** $\hat{\boldsymbol{\theta}}(n \mid w)$

The asymptotic sampling distribution of $\hat{\boldsymbol{\theta}}(n \mid w)$ is a consistent estimator of $\boldsymbol{\theta}^\star$.

Postulate 6.2. *As $n \to \infty$, the sampling distribution of $\hat{\boldsymbol{\theta}}(n \mid w)$ converges in probability to $\boldsymbol{\theta}^\star$, written*

$$\mathbf{Y}(n \mid w) \xrightarrow{P} \boldsymbol{\theta}^\star\,, \tag{6.11}$$

since

$$\lim_{n \to \infty} \Pr\left[\mathrm{MSE}\left(\mathbf{Y}(n \mid w)\right) < \epsilon\right] = 1 \tag{6.12}$$

for every $\epsilon > 0$.

By eqs. (5.7) and (6.11), the variance-covariance of the asymptotic sampling distribution of $\hat{\boldsymbol{\theta}}(n \mid w)$ is given by

$$\mathrm{Var}\big[\mathbf{Y}(n \mid w)\big] = \mathbb{E}\left[\left(\mathbf{Y}(n \mid w) - \boldsymbol{\theta}^\star\right)\left(\mathbf{Y}(n \mid w) - \boldsymbol{\theta}^\star\right)^\intercal\right]. \tag{6.13}$$

Postulate 6.3. *The variance-covariance of the asymptotic sampling distribution of* $\hat{\boldsymbol{\theta}}(n \mid w)$ *obtains the* Cramér-Rao lower-bound *for point estimators that are strictly a function of* n *masked system failure times in which candidate sets are of cardinality* w. *We denote this variance-covariance by*

$$\boldsymbol{\Sigma_n}(\boldsymbol{\theta}^\star \mid w) \equiv \frac{1}{n} \boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta}^\star \mid w) \; . \tag{6.14}$$

The asymptotic sampling distribution of $\hat{\boldsymbol{\theta}}(n \mid w)$ is normally distributed.

Postulate 6.4. *As* $n \to \infty$, *the sampling distribution of* $\hat{\boldsymbol{\theta}}(n \mid w)$ *converges in distribution to a multivariate normal distribution with a mean* $\boldsymbol{\theta}^\star$ *and a variance-covariance* $\boldsymbol{\Sigma_n}(\boldsymbol{\theta}^\star \mid w)$, *written*

$$\mathbf{Y}(n \mid w) \xrightarrow{d} \operatorname{MVN}\Big(\boldsymbol{\theta}^\star, \boldsymbol{\Sigma_n}(\boldsymbol{\theta}^\star \mid w)\Big) \; . \tag{6.15}$$

The maximum likelihood estimator $\hat{\boldsymbol{\theta}}(n \mid w)$ is an asymptotically *efficient* estimator since it obtains the Cramér-Rao lower-bound as given by **??**. Thus, for a sufficiently large sample of masked system failure times, the sampling distribution of $\hat{\boldsymbol{\theta}}(n \mid w)$ varies only *slightly* from sample to sample with a *central tendency* around $\boldsymbol{\theta}^\star$.

By eqs. (6.5), (6.6) and (6.11) and **??**, the asymptotic sampling distribution of $\hat{\boldsymbol{\theta}}(n \mid w)$ has the minimum mean squared error of any unbiased estimator,

$$\operatorname{MSE}\big(\mathbf{Y}(n \mid w)\big) = \frac{1}{n} \operatorname{tr}\Big(\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta}^\star \mid w)\Big) = \operatorname{tr}\left(\mathbf{CRLB_{n\,w}}\right) \; . \tag{6.16}$$

Remark. Look up Slverty's theorem to justify the next paragraph.

To estimate the sampling distribution of $\hat{\boldsymbol{\theta}}(n \mid w)$, we may assume the sample size is sufficiently large such that the asymptotic distribution becomes a reasonable approximation. Since $\mathbf{Y}(n \mid w) \xrightarrow{P} \boldsymbol{\theta}^\star$ and $\mathbf{Y}(n \mid w) \xrightarrow{d} \operatorname{MVN}\Big(\boldsymbol{\theta}^\star, \boldsymbol{\Sigma_n}\left(\boldsymbol{\theta}^\star \mid w\right)\Big)$, it follows that

$$\mathbf{Y}(n \mid w) \xrightarrow{d} \operatorname{MVN}\left(\hat{\boldsymbol{\theta}}(n \mid w), \boldsymbol{\Sigma_n}\left(\hat{\boldsymbol{\theta}}(n \mid w) \mid w\right)\right) \; . \tag{6.17}$$

Thus, we can approximate $\boldsymbol{\theta}^\star$ and $\boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^\star \mid w\right)$ and obtain the following result. The proof of this theorem is beyond the scope of this paper.

Theorem 6.1. *For sufficiently large sample size* n, *the sampling distribution of* $\hat{\boldsymbol{\theta}}(n \mid w)$ *is approximately normally distributed with a mean* $\hat{\boldsymbol{\theta}}(n \mid w)$ *and a variance-covariance matrix* $\boldsymbol{\Sigma_n}\left(\hat{\boldsymbol{\theta}}(n \mid w) \mid w\right)$, *written*

$$\mathbf{Y}(n \mid w) \overset{approx.}{\sim} \operatorname{MVN}\left(\hat{\boldsymbol{\theta}}(n \mid w), \boldsymbol{\Sigma_n}\left(\hat{\boldsymbol{\theta}}(n \mid w) \mid w\right)\right) \; . \tag{6.18}$$

6.2 Maximum likelihood estimator

Consider an i.i.d. sample of r asymptotically unbiased estimates of $\boldsymbol{\theta}^\star$ denoted by $\bar{\boldsymbol{\theta}}^{(i)}$ which have sampling distributions with variance-covariances given respectively by $\boldsymbol{\Sigma}^{(i)}$ for $i = 1, \ldots, r$. The maximum likelihood estimator of $\boldsymbol{\theta}^\star$ given these point estimates is the inverse-variance weighted mean and is given by

$$\hat{\boldsymbol{\theta}} = \left(\sum_{i=1}^{r} \boldsymbol{A_i}\right)^{-1} \left(\sum_{i=1}^{r} \boldsymbol{A_i} \bar{\boldsymbol{\theta}}^{(i)}\right) \; , \tag{6.19}$$

where $\boldsymbol{A_i}$ is the inverse of $\mathbf{\Sigma}^{(i)}$.

Suppose that the estimates are given by the maximum likelihood estimator described in **??**. The maximum likelihood estimator given these estimates has a sampling distribution given by the following definition.

Definition 6.5. *Let $\mathbf{M_n}$ be a random sample of n masked system failure times in which n_i realizations have w_i candidates for $i = 1, \ldots, r$. The maximum likelihood estimator given by*

$$\hat{\boldsymbol{\theta}}_{\mathbf{n}} = \hat{\boldsymbol{\theta}} = \left(\sum_{i=1}^{r} \boldsymbol{A_i}\right)^{-1} \left(\sum_{i=1}^{r} \boldsymbol{A_i}\bar{\boldsymbol{\theta}}^{(i)}\right), \tag{6.20}$$

has a sampling distribution given by

$$\mathbf{Y_n} = \left(\sum_{i=1}^{r} \boldsymbol{A_i}\right)^{-1} \left(\sum_{i=1}^{r} \boldsymbol{A_i}\mathbf{Y}^{(i)}(n_i \,|\, w_i)\right), \tag{6.21}$$

where $\mathbf{Y}^{(i)}(n_i \,|\, w_i)$ is the sampling distribution of $\hat{\boldsymbol{\theta}}(n_i \,|\, w_i)$ for $i = 1, \ldots, r$ and

$$\boldsymbol{A_i} = n_i \boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^\star \,|\, w_i\right) \tag{6.22}$$

is the information matrix for $\boldsymbol{\theta}^\star$ with respect to $\mathbf{M}(w_i, n_i)$.

Theorem 6.2. *The random vector $\mathbf{Y_n}$ is an asymptotically unbiased estimator of $\boldsymbol{\theta}^\star$.*

Proof. The expectation of $\mathbf{Y_n}$ is given by

$$\mathbb{E}\left[\mathbf{Y_n}\right] = \mathbb{E}\left[\left(\sum_{i=1}^{r} \boldsymbol{A_i}\right)^{-1} \left(\sum_{i=1}^{r} \boldsymbol{A_i}\mathbf{Y}^{(i)}(n_i \,|\, w_i)\right)\right] \tag{a}$$

$$= \left(\sum_{i=1}^{r} \boldsymbol{A_i}\right)^{-1} \left(\sum_{i=1}^{r} \boldsymbol{A_i}\,\mathbb{E}\left[\mathbf{Y}^{(i)}(n_i \,|\, w_i)\right]\right) \tag{b}$$

$$= \left(\sum_{i=1}^{r} \boldsymbol{A_i}\right)^{-1} \left(\sum_{i=1}^{r} \boldsymbol{A_i}\right)\boldsymbol{\theta}^\star \tag{c}$$

$$= \boldsymbol{\theta}^\star. \tag{d}$$

□

Theorem 6.3. *The variance-covariance of $\mathbf{Y_n}$ is given by*

$$\operatorname{Var}\left[\mathbf{Y_n}\right] = \left(\sum_{i=1}^{r} n_i \boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^\star \,|\, w_i\right)\right)^{-1}. \tag{6.23}$$

Proof. Let

$$\boldsymbol{B} = \left(\sum_{i=1}^{r} \boldsymbol{A_i}\right)^{-1}. \tag{a}$$

The variance-covariance of $\mathbf{Y_n}$ is given by

$$\operatorname{Var}\left[\mathbf{Y_n}\right] = \operatorname{Var}\left[\boldsymbol{B}\left(\sum_{i=1}^{r}\boldsymbol{A_i}\mathbf{Y}^{(i)}(n_i \mid w_i)\right)\right] \tag{b}$$

$$= \boldsymbol{B}\operatorname{Var}\left[\sum_{i=1}^{r}\boldsymbol{A_i}\mathbf{Y}^{(i)}(n_i \mid w_i)\right]\boldsymbol{B}^\intercal \tag{c}$$

$$= \boldsymbol{B}\left(\sum_{i=1}^{r}\operatorname{Var}\left[\boldsymbol{A_i}\mathbf{Y}^{(i)}(n_i \mid w_i)\right]\right)\boldsymbol{B}^\intercal \tag{d}$$

$$= \boldsymbol{B}\left(\sum_{i=1}^{r}\boldsymbol{A_i}\operatorname{Var}\left[\mathbf{Y}^{(i)}(n_i \mid w_i)\right]\boldsymbol{A_i}^\intercal\right)\boldsymbol{B}^\intercal . \tag{e}$$

By **??**, the variance-covariance of $\mathbf{Y}^{(i)}(n_i \mid w_i)$ is given by

$$\frac{1}{n_i}\boldsymbol{\mathcal{I}}^{-1}\left(\boldsymbol{\theta}^\star \mid w_i\right) . \tag{f}$$

By eq. (6.22), this is equivalent to $\boldsymbol{A_i^{-1}}$. Performing this substitution results in

$$\operatorname{Var}\left[\mathbf{Y_n}\right] = \boldsymbol{B}\left(\sum_{i=1}^{r}\boldsymbol{A_i}\boldsymbol{A_i^{-1}}\boldsymbol{A_i}^\intercal\right)\boldsymbol{B}^\intercal \tag{g}$$

$$= \boldsymbol{B}\left(\sum_{i=1}^{r}\boldsymbol{A_i}^\intercal\right)\boldsymbol{B}^\intercal . \tag{h}$$

The summation is equivalent to $\boldsymbol{(B^\intercal)^{-1}}$. Performing this substitution results in

$$\operatorname{Var}\left[\mathbf{Y_n}\right] = \boldsymbol{B(B^\intercal)^{-1}B^\intercal} \tag{i}$$

$$= \boldsymbol{B} \tag{j}$$

By eq. (a), $\boldsymbol{B}$ is given by

$$\left(\sum_{i=1}^{r}\boldsymbol{A_i}\right)^{-1} \tag{k}$$

and by eq. (6.22), $\boldsymbol{A_i}$ is given by

$$n_i\boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^\star \mid w_i\right) . \tag{l}$$

Performing these substitution results in

$$\operatorname{Var}\left[\mathbf{Y_n}\right] = \left(\sum_{i=1}^{r}n_i\boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^\star \mid w_i\right)\right)^{-1} . \tag{m}$$

□

The weighted estimator asymptotically achieves the Cramér-Rao lower-bound as given by eq. (6.5) thus it is the asymptotic UMVUE estimator of $\boldsymbol{\theta}^\star$ given a random sample of masked system failure times in which n_i realizations have w_i candidates for $i = 1, \ldots, r$.

A linear combination of multivariate normal distributions is a multivariate normal distribution, thus the asymptotic sampling distribution of $\hat{\boldsymbol{\theta}}_{\mathbf{n}}$ is normally distributed.

Postulate 6.5. *As $n \to \infty$, the sampling distribution of $\hat{\boldsymbol{\theta}}_{\mathbf{n}}$ converges in distribution to a multivariate normal with a mean $\boldsymbol{\theta}^\star$ and a variance-covariance given by eq. (6.23), written*

$$\mathbf{Y_n} \xrightarrow{d} \operatorname{MVN}\left(\boldsymbol{\theta}^\star, \left(\sum_{i=1}^{r} n_i \boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^\star \mid w_i\right)\right)^{-1}\right). \tag{6.24}$$

The generative model for $\mathbf{Y_n}$ is given by algorithm 4.

Algorithm 3: Generative model of maximum likelihood estimator

Input:
  $\boldsymbol{\Theta}^\star$, the true parameter value of the series system.

Output:
  $\hat{\boldsymbol{\theta}}_{\mathbf{n}}$, a realization of $\mathbf{Y_n}$.

```
Model GenerateMLE(Θ*)
    draw candidate cardinality w ~ p_W( · )
    θ̂_n ← GenerateMLE(Θ*, w)
    return θ̂_n
```

Bootstrap of sample covariance

Another estimator of the variance-covariance of $\mathbf{Y_n}$ is given by the sample covariance.

Definition 6.6. *Given a sample of r maximum likelihood estimates, $\hat{\boldsymbol{\theta}}_{\mathbf{n}}^{(\mathbf{1})}, \ldots, \hat{\boldsymbol{\theta}}_{\mathbf{n}}^{(\mathbf{r})}$, the maximum likelihood estimator of the variance-covariance of the sampling distribution of $\hat{\boldsymbol{\theta}}_{\mathbf{n}}$ is given by the sample covariance,*

$$\hat{\operatorname{Var}}\left[\mathbf{Y_n}\right] = \frac{1}{r}\sum_{i=1}^{r}\left(\hat{\boldsymbol{\theta}}_{\mathbf{n}}^{(\mathbf{i})} - \boldsymbol{\theta}^\star\right) \cdot \left(\hat{\boldsymbol{\theta}}_{\mathbf{n}}^{(\mathbf{i})} - \boldsymbol{\theta}^\star\right)^\intercal. \tag{6.25}$$

The sample covariance depends upon a random sample of maximum likelihood estimates, and thus the sample covariance is a random matrix. However, by the property that maximum likelihood estimators converge in probability to the true value,

$$\hat{\operatorname{Var}}\left[\mathbf{Y_n}\right] \xrightarrow{P} \operatorname{Var}\left[\mathbf{Y_n}\right], \tag{6.26}$$

it follows that

$$\mathbf{Y_n} \xrightarrow{d} \operatorname{MVN}\left(\boldsymbol{\theta}^\star, \hat{\operatorname{Var}}\left[\mathbf{Y_n}\right]\right). \tag{6.27}$$

If only one maximum likelihood estimate is realized, $\mathbf{Y_n} = \hat{\boldsymbol{\theta}}_{\mathbf{n}}$, then the sample covariance given by **??** cannot be computed. In the *parametric Bootstrap*, the sample covariance is approximated by using $\hat{\boldsymbol{\theta}}_{\mathbf{n}}$ as an estimate of $\boldsymbol{\theta}^\star$ in the generative model described by **??**, i.e.,

$$\hat{\boldsymbol{\theta}}_{\mathbf{n}}^{(\mathbf{1})}, \ldots, \hat{\boldsymbol{\theta}}_{\mathbf{n}}^{(\mathbf{r})} \leftarrow \texttt{GenerateMLE}\left(\hat{\boldsymbol{\theta}}_{\mathbf{n}}\right), \tag{6.28}$$

and the sample covariance of these approximate maximum likelihood estimates is an asymptotically unbiased estimator of $\mathrm{Var}\left[\mathbf{Y_n}\right]$.

Algorithm 4: Sample covariance of a bootstrapped sample of maximum likelihood estimates.

Input:
$\hat{\boldsymbol{\theta}}_{\mathbf{n}}$, an estimate of $\boldsymbol{\theta}^{\star}$.

Output:
sample covariance of a bootstrapped sample of maximum likelihood estimates.

```
Model BoostrapCovariance
    (θ̂_n)
for i = 1 to r do
    θ̂_n^(i) ← GenerateMLE(θ̂_n)
end
return sample covariance of θ̂_n^(1), ... , θ̂_n^(r)
```

Chapter 7

Sampling distribution of functions of parameters

Suppose we have a characteristic of interest that is a function of $\boldsymbol{\theta}^\star \in \mathbb{R}^{m \cdot q}$, denoted by $\mathbf{g}(\boldsymbol{\theta}^\star) \colon \mathbb{R}^{m \cdot q} \mapsto \mathbb{R}^p$. By the invariance property of maximum likelihood estimators, if $\hat{\boldsymbol{\theta}}_\mathbf{n}$ is the maximum likelihood estimator of $\boldsymbol{\theta}^\star$ then $\mathbf{g}(\hat{\boldsymbol{\theta}}_\mathbf{n})$ is the maximum likelihood estimator of $\mathbf{g}(\boldsymbol{\theta}^\star)$.

By postulate 6.4, $\mathbf{Y_n}$ is a random vector drawn from a multivariate normal distribution,

$$\mathbf{Y_n} \sim \mathrm{MVN}\left(\boldsymbol{\theta}^\star, \mathrm{Var}\left[\mathbf{Y_n}\right]\right) , \qquad \text{(6.24 revisited)}$$

Therefore, $\mathbf{g}\left(\mathbf{Y_n}\right)$ is a random vector. Under the regularity conditions (see assumption 5.1), $\mathbf{g}\left(\mathbf{Y_n}\right)$ is asymptotically normally distributed with a mean given by $\mathbf{g}(\boldsymbol{\theta}^\star)$ and a variance-covariance given by $\mathrm{Var}\left[\mathbf{g}(\mathbf{Y_n})\right]$, written

$$\mathbf{g}\left(\mathbf{Y_n}\right) \xrightarrow{d} \mathrm{MVN}\left(\mathbf{g}\left(\boldsymbol{\theta}^\star\right), \mathrm{Var}\left[\mathbf{g}(\mathbf{Y_n})\right]\right) . \tag{7.1}$$

7.1 Confidence intervals

A property of multivariate normal distributions is that the marginal of any subset of the components is given by dropping the irrelevant components from the mean vector $\boldsymbol{\theta}^\star$ and the variance-covariance $\mathrm{Var}\left[\mathbf{Y_n}\right]$.

If we are interested in the asymptotic marginal distribution of the j^{th} component, we drop all of the other components, resulting in the univariate normal distribution given by

$$\left[\mathbf{Y_n}\right]_j \sim \mathcal{N}\left(\theta_j^\star, \frac{\sigma^2}{n}\right) , \tag{7.2}$$

where $\sigma^2 = \left[\boldsymbol{\mathcal{I}}^{-1}\left(\boldsymbol{\theta}^\star\right)\right]_{jj}$.

Asymptotically, the random variable $\left[\mathbf{Y_n}\right]_j$ realizes value less than constant a with probability given by

$$\Pr\left[\left[\mathbf{Y_n}\right]_j < a\right] = \Phi\left(\frac{a - \boldsymbol{\theta}_j^\star}{\sigma_j}\right) , \tag{7.3}$$

where Φ is the cumulative distribution function of the standard normal.

The smallest interval in which $[\mathbf{Y_n}]_j$ is realized with probability p is given by

$$\boldsymbol{\theta}^{\star}_j \pm \sigma_j \, \Phi^{\text{-}1}\big(p/2\big) \;, \tag{7.4}$$

where Φ^{-1} is the inverse cumulative distribution function of the standard normal.

By the invariance property of maximum likelihood estimators, a maximum likelihood estimator of eq. (7.4) is the confidence interval given by the following definition.

Definition 7.1. *An asymptotic* $(1-\alpha)\cdot 100\%$ *confidence interval for the* j^{th} *component of the true parameter index* $\boldsymbol{\theta}^{\star}$ *is given by*

$$\hat{\theta}_j \pm \hat{\sigma}_j \;\, \Phi^{\text{-}1}\big(1-\alpha/2\big) \;, \tag{7.5}$$

where $\hat{\theta}_j$ *is the* j^{th} *component of* $\hat{\boldsymbol{\theta}}_{\mathbf{n}}$ *and* $\hat{\sigma}_j$ *is the* j^{th} *diagonal element of* $\hat{\mathrm{Var}}\,[\mathbf{Y_n}]$*.*

Note that the confidence interval given by eq. (7.5) is a realization of a random interval since $\hat{\theta}_j$ is a realization of the normal distribution given by eq. (7.2) and $\hat{\sigma}_j$ is a function $\hat{\theta}_j$, where it is expected that $(1-\alpha/2)\cdot 100\%$ of the random intervals generated contain the true parameter index $\theta^{\star}_j$.

7.2 Hypothesis testing

The intervals discussed in section 7.1 ignore correlations between the components of $\mathbf{Y_n}$. Asymptotically, $\mathbf{Y_n}$ has a probability p of occurring within the hyper-ellipsoid (the smallest hyper-volume with probability p) centered around $\boldsymbol{\theta}^{\star}$ given by

$$\big(\mathbf{Y_n} - \boldsymbol{\theta}^{\star}\big)^{\mathsf{T}} \boldsymbol{\mathcal{I}_n}\left(\boldsymbol{\theta}^{\star}\right)\big(\mathbf{Y_n} - \boldsymbol{\theta}^{\star}\big) \;\leq\; \chi^2_{m\cdot q}(p)\,, \tag{7.6}$$

where $\chi^2_{m\cdot q}(p)$ is the p quantile of the chi-square distribution with $m \cdot q$ degrees of freedom.

By the invariance property of maximum likelihood estimators, a maximum likelihood estimator of eq. (7.6) is given by the confidence region.

Definition 7.2. *Suppose we have a maximum likelihood estimate* $\hat{\boldsymbol{\theta}}_{\mathbf{n}}$ *and hypothesize that the true parameter index is given by* $\boldsymbol{\theta}$*. At significance level* α *we fail to reject the null hypothesis* $H_0\colon \boldsymbol{\theta}^{\star} = \boldsymbol{\theta}$ *versus the alternative hypothesis* $H_a\colon \boldsymbol{\theta}^{\star} \neq \boldsymbol{\theta}$ *if*

$$\Big(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\mathbf{n}}\Big)^{\mathsf{T}} \boldsymbol{\mathcal{I}_n}\Big(\hat{\boldsymbol{\theta}}_{\mathbf{n}}\Big)\Big(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\mathbf{n}}\Big) \;\leq\; \chi^2_{m\cdot q}\left(1-\alpha\right) \;, \tag{7.7}$$

where $\chi^2_{m\cdot q}(1-\alpha)$ *is the* $(1-\alpha)$ *quantile of the chi-square distribution with* $m\cdot q$ *degrees of freedom.*

Applying eq. (7.7) to the marginal of $\big[\mathbf{Y}(n\,|\,w)\big]_j$ generates the confidence interval described in **??**.

7.3 Monte-carlo simulation

The generative model of the sampling distribution of $\mathbf{Y_n}$, as described in algorithm 4, is asymptotically equivalent to generating samples from the multivariate normal distribution given by **??**. However, generating samples from the multivariate normal is much less computationally demanding than the non-asymptotic model.

The generative model may be used to generate samples of any statistic that is a function of the true parameter index $\boldsymbol{\theta}^\star$. A sample drawn from $\mathbf{g}(\mathbf{Y}(n \mid w))$ may be generated by drawing r maximum likelihood estimates from the sampling distribution

$$\hat{\boldsymbol{\theta}}_{\mathbf{n}}^{(\mathbf{1})}, \ldots, \hat{\boldsymbol{\theta}}_{\mathbf{n}}^{(\mathbf{r})}, \tag{7.8}$$

and applying $\mathbf{g}$ to each, resulting in the sample

$$\mathbf{g}\left(\hat{\boldsymbol{\theta}}_{\mathbf{n}}^{(\mathbf{1})}\right), \ldots, \mathbf{g}\left(\hat{\boldsymbol{\theta}}_{\mathbf{n}}^{(\mathbf{r})}\right). \tag{7.9}$$

An estimate of the variance-covariance $\operatorname{Var}\left[\mathbf{g}(\mathbf{Y}_{\mathbf{n}}\right]$ is given by the sample covariance

$$\hat{\operatorname{Var}}\left[\mathbf{g}\left(\mathbf{Y}_{\mathbf{n}}\right)\right] = \frac{1}{r}\sum_{i=1}^{r}\left(\mathbf{g}\left(\hat{\boldsymbol{\theta}}_{\mathbf{n}}^{(\mathbf{i})}\right) - \mathbf{g}(\hat{\boldsymbol{\theta}}_{\mathbf{n}})\right)\left(\mathbf{g}\left(\hat{\boldsymbol{\theta}}_{\mathbf{n}}^{(\mathbf{i})}\right) - \mathbf{g}(\hat{\boldsymbol{\theta}}_{\mathbf{n}})\right)^{\intercal}. \tag{7.10}$$

Thus,

$$\mathbf{g}\left(\mathbf{Y}_{\mathbf{n}}\right) \sim \operatorname{MVN}\left(\mathbf{g}\left(\hat{\boldsymbol{\theta}}_{\mathbf{n}}\right), \hat{\operatorname{Var}}\left[\mathbf{Y}_{\mathbf{n}}\right]\right). \tag{7.11}$$

In what follows, we explore a couple of characteristics of interest about the series system.

Mean time to failure of components

The *mean time to failure* of component j is given by

$$\mathrm{g}(\boldsymbol{\theta}^\star) = \mathbb{E}_{\boldsymbol{\theta}^\star}\left[\mathrm{T}_j\right]. \tag{7.12}$$

Suppose $\mathrm{g}(\boldsymbol{\theta}^\star) \in \mathbb{R}^{m \cdot q}$, then the variance-covariance of

$$\mathcal{J}\left(\mathrm{g}(\boldsymbol{\theta}^\star)\right) \boldsymbol{\mathcal{I}_n}^{-1}\left(\boldsymbol{\theta}^\star\right) \mathcal{J}\left(\mathrm{g}(\boldsymbol{\theta}^\star)\right)^{\intercal}, \tag{7.13}$$

where $\mathcal{J}\left(\,\cdot\,\right) \in \mathbb{R}^{?}$ is the Jacobian as given by **??**.

The maximum likelihood estimator of $\boldsymbol{\theta}^\star$ is given by $\mathrm{g}(\hat{\boldsymbol{\theta}}_{\mathbf{n}})$ and is approximately normally distributed,

$$\mathrm{g}(\mathbf{Y}_{\mathbf{n}}) \xrightarrow{d} \mathcal{N}\left(\mathrm{g}\left(\hat{\boldsymbol{\theta}}_{\mathbf{n}}\right), \hat{\sigma}^2\right). \tag{7.14}$$

where $\hat{\sigma}^2$ is the sample variance.

Chapter 8

Case study: exponentially distributed component lifetimes

Consider a series system in which the component lifetimes are exponentially distributed. It is a popular choice due to its analytical tractability, e.g., it is the only continuous distribution with a constant failure rate.

8.1 System distribution

Consider component j with an exponentially distributed lifetime and a failure rate $\lambda_j^\star$, denoted by $\mathrm{T}_j \sim \mathrm{EXP}(\lambda_j^\star)$. The random variable T_j has reliability, density, and failure rate functions given respectively by

$$R_j(t \mid \lambda_j^\star) = \exp(-\lambda_j^\star \cdot t)\,, \tag{8.1}$$

$$f_j(t \mid \lambda_j^\star) = \lambda_j^\star \exp(-\lambda_j^\star \cdot t)\,, \tag{8.2}$$

$$h_j(\cdot \mid \lambda_j^\star) = \lambda_j^\star\,, \tag{8.3}$$

where $t > 0$ and $\lambda_j^\star > 0$.

A series system composed of m components with exponentially distributed lifetimes, $\mathrm{T}_j \sim \mathrm{EXP}(\lambda_j^\star)$ for $j = 1, \ldots, m$, has a random lifetime denoted by the random variable $\mathrm{S} = \min(\mathrm{T}_1, \ldots, \mathrm{T}_m)$ with a true parameter value given by

$$\boldsymbol{\lambda}^\star = \left(\lambda_1^\star, \ldots, \lambda_m^\star\right)^\intercal\,, \tag{8.4}$$

where T_j has a true parameter value indexed by element j of $\boldsymbol{\lambda}^\star$. The system's reliability is given by

$$R_{\mathrm{S}}\left(t \mid \boldsymbol{\lambda}^\star\right) = \exp\left(-\left[\sum_{j=1}^m \lambda_j^\star\right] t\right) \mathbb{1}_{\mathrm{t}>0} + \mathbb{1}_{\mathrm{t}\leq 0}\,. \tag{8.5}$$

Proof. By theorem 3.1,

$$R_{\mathrm{S}}(t \mid \boldsymbol{\lambda}^\star) = \prod_{j=1}^m R_j(t \mid \lambda_j^\star)\,. \tag{a}$$

Plugging in the component reliability functions as given by eq. (8.1) results in

$$R_{\mathrm{S}}(t \,|\, \boldsymbol{\lambda}^{\star}) = \prod_{j=1}^{m} \exp(-\lambda_j^{\star} \cdot t) = \exp\left(- \left[\sum_{j=1}^{m} \lambda_j^{\star} \right] t \right) . \tag{b}$$

□

The system's reliability function belongs to the family of exponential distributions with a failure rate that is the sum of the component failure rates.

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

$$\mathrm{S} \sim \mathrm{EXP}\left(\sum_{j=1}^{m} \lambda_j^{\star} \right) , \tag{8.6}$$

where ${\lambda_j}^{\star}$ *is the failure rate of component* j.

By **??**, the system has density and failure rate functions given respectively by

$$f_{\mathrm{S}}(t \,|\, \boldsymbol{\lambda}^{\star}) = \left(\sum_{j=1}^{m} \lambda_j^{\star} \right) \exp\left(- \left[\sum_{j=1}^{m} {\lambda_j}^{\star} \right] t \right) \mathbb{1}_{\mathrm{t}>0} , \tag{8.7}$$

$$h_{\mathrm{S}}(\cdot \,|\, \boldsymbol{\lambda}^{\star}) = \sum_{j=1}^{m} \lambda_j^{\star} . \tag{8.8}$$

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

$$p_{\mathrm{K}\,|\,\mathrm{S}}(k \,|\, t, \boldsymbol{\lambda}^{\star}) = p_{\mathrm{K}}(k \,|\, \boldsymbol{\lambda}^{\star}) = \frac{\lambda_k^{\star}}{\sum_{p=1}^{m} \lambda_p^{\star}} \mathbb{1}_{\mathrm{k} \in \{1,\ldots,\mathrm{m}\}} . \tag{8.9}$$

Proof. By theorem 3.3,

$$p_{\mathrm{K}\,|\,\mathrm{S}}(k \,|\, t, \boldsymbol{\lambda}^{\star}) = \frac{h_k(t \,|\, \boldsymbol{\lambda}_k^{\star})}{h_{\mathrm{S}}(t \,|\, \boldsymbol{\lambda}^{\star})} . \tag{a}$$

Plug in the failure rate of component k and the failure rate of the system given by eqs. (8.3) and (8.8). □

Due to the constant failure rates of the components, K and S are independent. The joint density of K and S is given by

$$f_{\mathrm{K,S}}(k, t \,|\, \boldsymbol{\lambda}^{\star}) = \lambda_k^{\star} \exp\left(- \left[\sum_{j=1}^{m} {\lambda_j}^{\star} \right] t \right) \mathbb{1}_{\mathrm{t}>0} \, \mathbb{1}_{\mathrm{k} \in \{1,\ldots,\mathrm{m}\}} \tag{8.10}$$

Proof. By definition,

$$f_{\mathrm{K,S}}(k, t \,|\, \boldsymbol{\lambda}^{\star}) = p_{\mathrm{K}\,|\,\mathrm{S}}(k \,|\, t, \boldsymbol{\lambda}^{\star}) f_{\mathrm{S}}(t \,|\, \boldsymbol{\lambda}^{\star}) . \tag{a}$$

Plug in the conditional probability and the marginal probability given by eqs. (8.7) and (8.9). □

The conditional joint density of $\mathbf{C}$ and S given $\mathrm{W} = w$ is given by

$$f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}(\mathbf{c}, t \,|\, w, \boldsymbol{\lambda}^\star) = \frac{\mathbb{1}_{\mathrm{t}>0}\, \mathbb{1}_{|\mathbf{c}|=\mathrm{w}}}{\binom{m-1}{w-1}} \left(\sum_{j \in \mathbf{c}} \lambda_j \right) \exp\left[-\left(\sum_{j=1}^{m} \lambda_j \right) t \right], \tag{8.11}$$

whose proof follows from theorem 3.6.

The conditional probability that a component in $\mathbf{C}$ is the cause of a system failure given that $\mathrm{S} = t$ and $\mathrm{W} = w$ is given by

$$p_{\mathbf{C}\,|\,\mathrm{S},\mathrm{W}}(\mathbf{c} \,|\, t, w, \boldsymbol{\lambda}^\star) = \frac{\sum_{j \in \mathbf{c}} \lambda^\star_j}{\binom{m-1}{w-1} \sum_{p=1}^{m} \lambda^\star_p}\, \mathbb{1}_{\mathrm{t}>0}\, \mathbb{1}_{|\mathbf{c}|=\mathrm{w}}\,. \tag{8.12}$$

whose proof follows from corollary 3.7.1.

Note that the marginal distribution of $\mathbf{C} \,|\, \mathrm{W} = w$ is given by

$$p_{\mathbf{C}\,|\,\mathrm{W}}\left(\mathbf{c} \,|\, w, \boldsymbol{\lambda}^\star\right) = \frac{\sum_{j \in \mathbf{c}} \lambda^\star_j}{\binom{m-1}{w-1} \sum_{p=1}^{m} \lambda^\star_p}\, \mathbb{1}_{|\mathbf{c}|=\mathrm{w}}\,, \tag{8.13}$$

Proof. The marginal is given by

$$p_{\mathbf{C}\,|\,\mathrm{W}}(\mathbf{c} \,|\, w, \boldsymbol{\lambda}^\star) = \int_{-\infty}^{\infty} f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}\left(\mathbf{c}, t \,|\, w, \boldsymbol{\lambda}^\star\right) \mathrm{d}t \tag{a}$$

$$= \int_{0}^{\infty} \frac{\mathbb{1}_{\mathrm{t}>0}\, \mathbb{1}_{|\mathbf{c}|=\mathrm{w}}}{\binom{m-1}{w-1}} \left(\sum_{j \in \mathbf{c}} \lambda_j \right) \exp\left[-\left(\sum_{j=1}^{m} \lambda_j \right) t \right] \mathrm{d}t \tag{b}$$

$$= \frac{\mathbb{1}_{|\mathbf{c}|=\mathrm{w}}}{\binom{m-1}{w-1}} \sum_{j \in \mathbf{c}} \lambda_j \int_{0}^{\infty} \exp\left[-\left(\sum_{j=1}^{m} \lambda_j \right) t \right] \mathrm{d}t\,. \tag{c}$$

Multiplying the integrand by

$$\frac{\sum_{j=1}^{m} \lambda_j}{\sum_{j=1}^{m} \lambda_j} \tag{d}$$

results in

$$p_{\mathbf{C}\,|\,\mathrm{W}}(\mathbf{c} \,|\, w, \boldsymbol{\lambda}^\star) = \frac{\mathbb{1}_{|\mathbf{c}|=\mathrm{w}}}{\binom{m-1}{w-1}} \frac{\sum_{j \in \mathbf{c}} \lambda_j}{\sum_{j=1}^{m} \lambda_j} \int_{0}^{\infty} \left(\sum_{j=1}^{m} \lambda_j \right) \exp\left[-\left(\sum_{j=1}^{m} \lambda_j \right) t \right] \mathrm{d}t \tag{e}$$

$$= \frac{\mathbb{1}_{|\mathbf{c}|=\mathrm{w}}}{\binom{m-1}{w-1}} \frac{\sum_{j \in \mathbf{c}} \lambda_j}{\sum_{j=1}^{m} \lambda_j} \int_{0}^{\infty} f_{\mathrm{S}}\left(t \,|\, \boldsymbol{\lambda}^\star\right) \mathrm{d}t \tag{f}$$

$$= \frac{\mathbb{1}_{|\mathbf{c}|=\mathrm{w}}}{\binom{m-1}{w-1}} \frac{\sum_{j \in \mathbf{c}} \lambda_j}{\sum_{j=1}^{m} \lambda_j}\,. \tag{g}$$

□

The marginal distribution of $\mathbf{C} \,|\, \mathrm{W} = w$ given by **??** is the same as the joint distribution of $\mathbf{C}, \mathrm{S} \,|\, \mathrm{W} = w$ given by eq. (8.13). Thus, $\mathbf{C}$ and S are conditionally independent given $\mathrm{W} = w$.

8.2 Maximum likelihood estimator

According to definition 6.3, the maximum likelihood estimator of $\boldsymbol{\lambda}^\star$ is the value that maximizes the log-likelihood on a given masked system failure time sample, denoted $\mathbf{M_n}$. The likelihood function is given by

$$\mathcal{L}(\boldsymbol{\lambda} \,|\, \mathbf{M_n}) \propto \exp\left[-\left(\sum_{j=1}^{m} \lambda_j\right)\left(\sum_{i=1}^{n} t_i\right)\right]\left[\prod_{i=1}^{n}\left(\sum_{j\in\mathbf{c}_i} \lambda_j\right)\right], \tag{8.14}$$

whose proof follows from definition 5.2. Therefore, the log-likelihood function ℓ is given by

$$\ell(\boldsymbol{\lambda} \,|\, \mathbf{M_n}) = \text{const} + \sum_{i=1}^{n} \ln\left[\sum_{j\in\mathbf{c}_i} \lambda_j\right] - \left[\sum_{i=1}^{n} t_i\right]\left[\sum_{j=1}^{m} \lambda_j\right]. \tag{8.15}$$

Thus, maximum likelihood estimator of $\boldsymbol{\lambda}^\star$ is given by

$$\hat{\boldsymbol{\lambda}}_{\mathbf{n}} = \operatorname*{arg\,max}_{\boldsymbol{\lambda}>\mathbf{0}} \ell(\boldsymbol{\lambda} \,|\, \mathbf{M_n}). \tag{8.16}$$

The solution to eq. (8.16) must be a stationary point, i.e., a point at which the score function is zero. The j^{th} component of the score function is given by

$$\frac{\partial \ell}{\partial \lambda_j} = \sum_{i=1}^{n}\left(\sum_{p\in\mathbf{c}_i} \lambda_p\right)^{-1} \mathbb{1}_{\mathrm{j}\in\mathbf{c}_\mathrm{i}} - \sum_{i=1}^{n} t_i. \tag{8.17}$$

According to postulate 6.4, the sampling distribution of $\hat{\boldsymbol{\lambda}}_{\mathbf{n}}$ converges in distribution to a multivariate normal with a mean given by $\boldsymbol{\lambda}^\star$ and a variance-covariance given by

$$\frac{1}{n}\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\lambda}^\star). \tag{8.18}$$

Since $\boldsymbol{\lambda}^\star$ is not known it may be approximated by

$$\boldsymbol{\mathcal{I}}^{-1}\left(\hat{\boldsymbol{\lambda}}_{\mathbf{n}}\right). \tag{8.19}$$

The (j,k)-th element of the Fisher information matrix is given by

$$\left[\boldsymbol{\mathcal{I}}(\boldsymbol{\lambda}^\star)\right]_{jk} = \frac{\sum_{\mathbf{c}}\left[\left(\sum_{p\in\mathbf{c}} \lambda_p^\star\right)^{-1} \mathbb{1}_{\mathrm{j}\in\mathbf{c},\mathrm{k}\in\mathbf{c}}\right]}{\binom{m-1}{w-1}\sum_{p=1}^{m} \lambda_p^\star}. \tag{8.20}$$

where the summation indexed by $\mathbf{c}$ is over $\mathcal{P}(\{1,\ldots,m\})$.

Proof. By eq. (5.11), the (j,k)-th element of the information matrix is given by

$$\begin{aligned}\left[\boldsymbol{\mathcal{I}}(\boldsymbol{\lambda}^\star)\right]_{ij} = -\sum_{\mathbf{c}}\int_0^\infty \tfrac{\partial^2}{\partial\lambda_i\partial\lambda_j} \ln f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}\left(\mathbf{c}, t \,|\, w, \operatorname{mat}(\boldsymbol{\lambda})\right)\Big|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^\star} \cdot \\ f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}\left(\mathbf{c}, t \,|\, w, \operatorname{mat}(\boldsymbol{\lambda}^\star)\right) \mathrm{d}t\,,\end{aligned} \tag{a}$$

In the above equation, the second-order partial derivative of the log of the joint density is equivalent to the partial derivative with respect to the k^{th} parameter of the j^{th} component of the score, given by eq. (8.17), conditioned on a masked system failure time sample of size 1, denoted by $\mathbf{M_{n1}}$. This mixed second-order partial derivative is given by

$$\frac{\partial}{\partial \lambda_k}\frac{\partial \ell}{\partial \lambda_j} = \frac{\partial}{\partial \lambda_k}\left(\left[\sum_{p\in\mathbf{c}} \lambda_p\right]^{-1} \mathbb{1}_{\mathrm{j}\in\mathbf{c}} - t\right) \tag{b}$$

$$= -\left(\sum_{p\in\mathbf{c}} \lambda_p\right)^{-2} \mathbb{1}_{\mathrm{j}\in\mathbf{c},\mathrm{k}\in\mathbf{c}}\,. \tag{c}$$

Plugging in the joint density and substituting eq. (c) into eq. (a) results in

$$\left[\boldsymbol{\mathcal{I}}(\boldsymbol{\lambda}^\star)\right]_{ij} = \sum_{\mathbf{c}} \int_0^\infty \left(\sum_{p\in\mathbf{c}} \lambda_p\right)^{-2} \mathbb{1}_{\mathrm{p}\in\mathbf{c},\mathrm{k}\in\mathbf{c}}\Bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^\star} \frac{1}{\binom{m-1}{w-1}} \left(\sum_{p\in\mathbf{c}} \lambda_p^\star\right) \exp\left(-\left[\sum_{p=1}^m \lambda_p^\star\right] t\right) \mathrm{d}t\,. \tag{d}$$

Simplifying and rearranging results in

$$\left[\boldsymbol{\mathcal{I}}(\boldsymbol{\lambda}^\star)\right]_{ij} = \frac{1}{\binom{m-1}{w-1}} \sum_{\mathbf{c}} \left(\sum_{p\in\mathbf{c}} \lambda_p^\star\right)^{-1} \mathbb{1}_{\mathrm{p}\in\mathbf{c},\mathrm{k}\in\mathbf{c}} \int_0^\infty \exp\left(-\left[\sum_{p=1}^m \lambda_p^\star\right] t\right) \mathrm{d}t\,. \tag{e}$$

Multiplying the integral by $\frac{\sum_{p=1}^m \lambda_p^\star}{\sum_{p=1}^m \lambda_p^\star}$ results in an integral of the form

$$\int_0^\infty \frac{\sum_{p=1}^m \lambda_p^\star}{\sum_{p=1}^m \lambda_p^\star} \exp\left(-\left[\sum_{p=1}^m \lambda_p^\star\right] t\right) \mathrm{d}t\,. \tag{f}$$

Pulling the $\left(\sum_{p=1}^m \lambda_p^\star\right)^{-1}$ constant out of the integrand results in

$$\frac{1}{\sum_{p=1}^m \lambda_p^\star} \int_0^\infty \left(\sum_{p=1}^m \lambda_p^\star\right) \exp\left(-\left[\sum_{p=1}^m \lambda_p^\star\right] t\right) \mathrm{d}t = \frac{1}{\sum_{p=1}^m \lambda_p^\star} \int_0^\infty f_{\mathrm{S}}(t \mid \boldsymbol{\lambda}^\star)\,\mathrm{d}t \tag{g}$$

$$= \frac{1}{\sum_{p=1}^m \lambda_p^\star}\,. \tag{h}$$

Substituting eq. (h) into eq. (e) results in

$$\left[\boldsymbol{\mathcal{I}}(\boldsymbol{\lambda}^\star)\right]_{ij} = \frac{1}{\binom{m-1}{w-1}} \sum_{\mathbf{c}} \left\{\left(\sum_{p\in\mathbf{c}} \lambda_p^\star\right)^{-1} \frac{1}{\sum_{p=1}^m \lambda_p^\star} \mathbb{1}_{\mathrm{p}\in\mathbf{c},\mathrm{k}\in\mathbf{c}}\right\}. \tag{i}$$

$\square$

8.3 Sufficient statistics

Consider a sample of n masked system failure times denoted by $\mathbf{m_n}$. The mean system lifetime of the sample is given by

$$\bar{t} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{t_i}\,. \tag{8.21}$$

and another statistic, denoted by $\hat{\boldsymbol{\omega}}$, is a dictionary of candidate set-frequency count pairs where a candidate set $\mathbf{k} \in \boldsymbol{\mathcal{C}}$ maps to its sample frequency and is given by

$$\hat{\boldsymbol{\omega}}_{\mathbf{k}} = \sum_{i=1}^{n} \mathbb{1}_{\mathbf{k}}(\mathbf{c}_i)\,. \tag{8.22}$$

Given a sample $\mathbf{m_n}$, the likelihood $\mathcal{L}(\boldsymbol{\lambda} \,|\, \mathbf{m_n})$ described by eq. (8.14) is the same as the likelihood given by

$$\mathcal{L}(\boldsymbol{\lambda} \,|\, \bar{t}, \hat{\boldsymbol{\omega}}) = \exp\left(-n\bar{t}\sum_{j=1}^{m}\lambda_j\right)\prod_{\mathbf{c}\in\boldsymbol{\mathcal{C}_w}}\left(\sum_{j\in\mathbf{c}}\lambda_j\right)^{\hat{\boldsymbol{\omega}}_{\mathbf{c}}}, \tag{8.23}$$

where $\boldsymbol{\mathcal{C}_w}$ is the set of all candidate sets of cardinality w.

Proof. The likelihood function $\mathcal{L}$ with respect to a sample of n masked system failure times is given by

$$\mathcal{L}(\boldsymbol{\lambda} \,|\, \mathbf{m_n}) \propto \exp\left[-\left(\sum_{j=1}^{m}\lambda_j\right)\left(\sum_{i=1}^{n}t_i\right)\right]\left[\prod_{i=1}^{n}\left(\sum_{j\in\mathbf{c}_i}\lambda_j\right)\right]. \tag{8.14 revisited}$$

Substituting $\sum_{i=1}^{n}\mathrm{t_i}$ in the above equation with the equivalent expression $n\bar{t}$ results in

$$\mathcal{L}(\boldsymbol{\lambda} \,|\, \mathbf{M_n}) \propto \exp\left[-n\bar{t}\sum_{j=1}^{m}\lambda_j\right]\underbrace{\left[\prod_{i=1}^{n}\left(\sum_{j\in\mathbf{c}_i}\lambda_j\right)\right]}_{A}. \tag{a}$$

In the above equation, each unique set $\left(\lambda_{j_1} + \cdots + \lambda_{j_w}\right)$ in the product (labeled A) has multiplicity $\hat{\boldsymbol{\omega}}_{\{\lambda_{j_1}\cdots\lambda_{j_w}\}}$, thus we may substitute this product with the equivalent product

$$\prod_{\mathbf{c}\in\boldsymbol{\mathcal{C}_w}}\left(\sum_{j\in\mathbf{c}}\lambda_j\right)^{\hat{\boldsymbol{\omega}}_{\mathbf{c}}}, \tag{b}$$

where $\mathbf{c}$ is over all unique candidate sets of cardinality w. The result of this substitution is given by

$$\mathcal{L}(\boldsymbol{\lambda} \,|\, \bar{t}, \hat{\boldsymbol{\omega}}) = \exp\left(-n\bar{t}\sum_{j=1}^{m}\lambda_j\right)\prod_{\mathbf{c}\in\boldsymbol{\mathcal{C}_w}}\left(\sum_{j\in\mathbf{c}}\lambda_j\right)^{\hat{\boldsymbol{\omega}}_{\mathbf{c}}}. \tag{c}$$

□

The statistics $\bar{t}$ and $\hat{\boldsymbol{\omega}}$ are jointly sufficient for $\boldsymbol{\lambda}^\star$

Proof. By the Fisher–Neyman factorization theorem, since $\mathcal{L}$ given by eq. (8.14) can be factored as $\mathcal{L}(\bar{t}, \hat{\boldsymbol{\omega}}) = h(\bar{t}, \hat{\boldsymbol{\omega}})g(\bar{t}, \hat{\boldsymbol{\omega}})$, $\hat{\boldsymbol{\omega}}$ and $\bar{t}$ are joint sufficient statistics for $\boldsymbol{\lambda}^\star$. □

It may be helpful to consider an example given by the following

Example 1. Let a masked failure time sample of size n for a 3-out-of-3 system with exponentially distributed lifetime components be given by

$$\mathbf{m_n}4 = \Big(\big(t_1 = 1, \mathbf{c}_1 = \{1,2\}\big)\,, \big(t_2 = 1, \mathbf{c}_2 = \{1,3\}\big)\,, \big(t_3 = 2, \mathbf{c}_3 = \{1,2\}\big)\,, \big(t_4 = 2, \mathbf{c}_4 = \{1,2\}\big) \Big)\,.$$

Then, joint sufficient statics of $\boldsymbol{\lambda}^\star$ are given by

$$\bar{t} = \frac{1+1+2+2}{4} = \frac{3}{2}\,,$$
$$\hat{\boldsymbol{\omega}} = \{\hat{\boldsymbol{\omega}}_{\{1,2\}} \mapsto 3\,, \hat{\boldsymbol{\omega}}_{\{1,3\}} \mapsto 1\}\,.$$

No additional information is needed to compute the likelihood of $\boldsymbol{\lambda}$ with respect to the given $\mathbf{M_n}4$ sample.

The log-likelihood ℓ is given by

$$\ell(\boldsymbol{\theta} \,|\, \bar{t}, \hat{\boldsymbol{\omega}}, w) = -n\bar{t}\left(\sum_{j=1}^{m} \lambda_j\right) + \sum_{\mathbf{c}} \hat{\boldsymbol{\omega}}_{\mathbf{c}} \ln \left(\sum_{j \in \mathbf{c}} \lambda_j\right)\,. \tag{8.24}$$

To find the MLE $\hat{\boldsymbol{\lambda}}_n$, we find the stationary points for $\lambda_1, \ldots, \lambda_m$ given by

$$0 = -n\bar{t} + \sum_{\mathbf{c}} \hat{\boldsymbol{\omega}}_{\mathbf{c}} \left(\sum_{j \in \mathbf{c}} \lambda_j\right)^{-1} \mathbb{1}_{\mathrm{k}\in\mathbf{c}}\;; k = 1, \ldots, m\,. \tag{8.25}$$

The (j,k)-th element of the observed information matrix $\boldsymbol{\mathcal{J}}_n$ is given by

$$\left[\boldsymbol{\mathcal{J}}_n(\boldsymbol{\lambda}^\star\hat{\boldsymbol{\omega}})\right]_{jk} = -\sum_{\mathbf{c}} \hat{\boldsymbol{\omega}}_{\mathbf{c}} \left(\sum_{i \in \mathbf{c}} \lambda_i^\star\right)^{-2} \mathbb{1}_{\mathrm{j}\in\mathbf{c}}\, \mathbb{1}_{\mathrm{k}\in\mathbf{c}}\,, \tag{8.26}$$

which only depends on the statistic $\hat{\boldsymbol{\omega}}$.

8.4 Applications

A $(1-\alpha)\cdot 100\%$-confidence interval for $\lambda_j^\star$ is given by

$$\hat{\lambda}_j \pm z_{1-\alpha/2}\sqrt{\frac{1}{n}\left[\boldsymbol{\mathcal{I}}^{-1}(\hat{\boldsymbol{\lambda}}_{\mathbf{n}})\right]_{jj}}\,. \tag{8.27}$$

The expected lifetime of the system is given by

$$\mathbb{E}_{\boldsymbol{\lambda}^\star}[\mathrm{S}] = \frac{1}{\sum_{p=1}^{m} \lambda_j^\star}\,. \tag{8.28}$$

The maximum likelihood estimator of $\mathbb{E}_{\boldsymbol{\lambda}^\star}[\mathrm{S}]$ is given by $\mathbb{E}_{\hat{\boldsymbol{\lambda}}_{\mathbf{n}}}[\mathrm{S}]$, which has an asymptotic sampling distribution given by

$$\mathrm{S}_{\mathbb{E}_{\hat{\boldsymbol{\lambda}}_{\mathbf{n}}}[\mathrm{S}]} \sim \mathcal{N}\left(\mathbb{E}_{\boldsymbol{\lambda}^\star}[\mathrm{S}], \operatorname{Var}\left[\mathbb{E}_{\hat{\boldsymbol{\lambda}}_{\mathbf{n}}}[\mathrm{S}]\right]\right), \tag{8.29}$$

where

$$\operatorname{Var}\left[\mathbb{E}_{\hat{\boldsymbol{\lambda}}_{\mathbf{n}}}[\mathrm{S}]\right] = a\,. \tag{8.30}$$

If a system failure occurs, the expected number of inspections is given by

$$\mathbb{E}_{\boldsymbol{\lambda}^\star}[\mathrm{N}] = \frac{\sum_{n=1}^{m} n \cdot \lambda^\star_{\pi(n)}}{\sum_{p=1}^{m} \lambda^\star_p}\,, \tag{8.31}$$

where $\pi(j) \leq \pi(k) \implies \lambda^\star_j \geq \lambda^\star_k$. The maximum likelihood estimator of $\mathbb{E}_{\boldsymbol{\lambda}^\star}[\mathrm{N}]$ is given by $\mathbb{E}_{\hat{\boldsymbol{\lambda}}_{\mathbf{n}}}[\mathrm{N}]$, which has an asymptotic sampling distribution given by

$$\mathrm{S}_{\mathbb{E}_{\hat{\boldsymbol{\lambda}}_{\mathbf{n}}}[\mathrm{N}]} \sim \mathcal{N}\left(\mathbb{E}_{\boldsymbol{\lambda}^\star}[\mathrm{N}], \operatorname{Var}\left[\mathbb{E}_{\hat{\boldsymbol{\lambda}}_{\mathbf{n}}}[\mathrm{N}]\right]\right), \tag{8.32}$$

where an estimator of the variance-covariance $\operatorname{Var}\left[\mathbb{E}_{\hat{\boldsymbol{\lambda}}_{\mathbf{n}}}[\mathrm{N}]\right]$ is described in section 7.3.

8.5 3-out-of-3 system

The 3-out-of-3 system's lifetime distribution functions are given by

$$R_{\mathrm{S}}(t \mid \boldsymbol{\lambda}^\star) = \exp\left[-(\lambda^\star_1 + \lambda^\star_2 + \lambda^\star_3)t\right], \tag{8.33}$$

$$f_{\mathrm{S}}(t \mid \boldsymbol{\lambda}^\star) = (\lambda^\star_1 + \lambda^\star_2 + \lambda^\star_3)\exp\left[-(\lambda^\star_1 + \lambda^\star_2 + \lambda^\star_3)t\right], \tag{8.34}$$

$$h_{\mathrm{S}}(t \mid \boldsymbol{\lambda}^\star) = \lambda^\star_1 + \lambda^\star_2 + \lambda^\star_3\,, \tag{8.35}$$

where $t > 0$, $\lambda^\star_1 > 0$, $\lambda^\star_2 > 0$, and $\lambda^\star_3 > 0$.

Sampling distributions

Two candidates If we condition on samples in which each observation consists of two candidates, $\mathrm{W} = |\mathbf{C}| = 2$, the likelihood function is given by

$$\mathcal{L}(\boldsymbol{\lambda} \mid \bar{t}, \hat{\boldsymbol{\omega}}) = (\lambda_1 + \lambda_2)^{\hat{\boldsymbol{\omega}}_{\{1,2\}}}(\lambda_1 + \lambda_3)^{\hat{\boldsymbol{\omega}}_{\{1,3\}}}(\lambda_2 + \lambda_3)^{\hat{\boldsymbol{\omega}}_{\{2,3\}}}\exp\left(-n\bar{t}\,(\lambda_1 + \lambda_2 + \lambda_3)\right) \tag{8.36}$$

and the log-likelihood function is given by

$$\begin{aligned}\ell(\boldsymbol{\lambda} \mid \bar{t}, \hat{\boldsymbol{\omega}}) = \hat{\boldsymbol{\omega}}_{\{1,2\}}\ln(\lambda_1 + \lambda_2) + \hat{\boldsymbol{\omega}}_{\{1,3\}}\ln(\lambda_1 + \lambda_3) +{}& \\ \hat{\boldsymbol{\omega}}_{\{2,3\}}\ln(\lambda_2 + \lambda_3) - n\bar{t}(\lambda_1 + \lambda_2 + \lambda_3)\,.&\end{aligned} \tag{8.37}$$

The maximum likelihood estimate $\hat{\boldsymbol{\lambda}}_{\mathbf{n}}$ is given by the solving the system of equations in which the score is zero, $\mathrm{s}(\boldsymbol{\lambda}) = \mathbf{0}$, where the score is given by

$$\mathrm{s}(\boldsymbol{\lambda}) = \begin{pmatrix} \frac{\hat{\boldsymbol{\omega}}_{\{1,2\}}}{\lambda_1+\lambda_2} + \frac{\hat{\boldsymbol{\omega}}_{\{1,3\}}}{\lambda_1+\lambda_3} \\ \frac{\hat{\boldsymbol{\omega}}_{\{1,2\}}}{\lambda_1+\lambda_2} + \frac{\hat{\boldsymbol{\omega}}_{\{2,3\}}}{\lambda_2+\lambda_3} \\ \frac{\hat{\boldsymbol{\omega}}_{\{1,3\}}}{\lambda_1+\lambda_3} + \frac{\hat{\boldsymbol{\omega}}_{\{2,3\}}}{\lambda_1+\lambda_3} \end{pmatrix} - n\bar{t}\begin{pmatrix}1\\1\\1\end{pmatrix}. \tag{8.38}$$

This has a closed form solution given by

$$\hat{\boldsymbol{\lambda}}_{\mathbf{n}} = \frac{1}{n\bar{t}} \begin{pmatrix} \hat{\boldsymbol{\omega}}_{\{1,2\}} + \hat{\boldsymbol{\omega}}_{\{1,3\}} - \hat{\boldsymbol{\omega}}_{\{2,3\}} \\ \hat{\boldsymbol{\omega}}_{\{1,2\}} - \hat{\boldsymbol{\omega}}_{\{1,3\}} + \hat{\boldsymbol{\omega}}_{\{2,3\}} \\ -\hat{\boldsymbol{\omega}}_{\{1,2\}} + \hat{\boldsymbol{\omega}}_{\{1,3\}} + \hat{\boldsymbol{\omega}}_{\{2,3\}} \end{pmatrix} . \tag{8.39}$$

The information matrix is given by

$$\boldsymbol{\mathcal{I}}\left(\boldsymbol{\lambda}^\star\right) = \frac{1}{2(\lambda_1^\star + \lambda_2^\star + \lambda_3^\star)} \begin{bmatrix} \frac{1}{\lambda_1^\star+\lambda_2^\star} + \frac{1}{\lambda_1^\star+\lambda_3^\star} & \frac{1}{\lambda_1^\star+\lambda_2^\star} & \frac{1}{\lambda_1^\star+\lambda_3^\star} \\ \frac{1}{\lambda_1^\star+\lambda_2^\star} & \frac{1}{\lambda_1^\star+\lambda_2^\star} + \frac{1}{\lambda_2^\star+\lambda_3^\star} & \frac{1}{\lambda_2^\star+\lambda_3^\star} \\ \frac{1}{\lambda_1^\star+\lambda_3^\star} & \frac{1}{\lambda_2^\star+\lambda_3^\star} & \frac{1}{\lambda_1^\star+\lambda_3^\star} + \frac{1}{\lambda_2^\star+\lambda_3^\star} \end{bmatrix} \tag{8.40}$$

To derive the variance-covariance of the sampling distribution of $\hat{\boldsymbol{\lambda}}_{\mathbf{n}}$ for a sample of n masked system failure times, denoted by $\boldsymbol{\Lambda}_{\mathbf{n}}$, we take the inverse of $n \cdot \boldsymbol{\mathcal{I}}\left(\boldsymbol{\lambda}^\star\right)$ resulting in

$$\operatorname{Var}_{\boldsymbol{\lambda}^\star}[\boldsymbol{\Lambda}_{\mathbf{n}}] = \frac{\lambda_1^\star + \lambda_2^\star + \lambda_3^\star}{n} \begin{bmatrix} \lambda_1^\star + \lambda_2^\star + \lambda_3^\star & -\lambda_3^\star & -\lambda_2^\star \\ -\lambda_3^\star & \lambda_1^\star + \lambda_2^\star + \lambda_3^\star & -\lambda_1^\star \\ -\lambda_2^\star & -\lambda_1^\star & \lambda_1^\star + \lambda_2^\star + \lambda_3^\star \end{bmatrix} . \tag{8.41}$$

By the asymptotic unbiasedness of $\hat{\boldsymbol{\lambda}}_{\mathbf{n}}$, the asymptotic mean squared error is given by the trace of the variance-covariance matrix,

$$\operatorname{MSE}\left(\boldsymbol{\Lambda}_{\mathbf{n}}\right) = \frac{3(\lambda_1^\star + \lambda_2^\star + \lambda_3^\star)^2}{n} . \tag{8.42}$$

According to theorem 6.1, for a sufficiently large sample size n,

$$\boldsymbol{\Lambda}_{\mathbf{n}} \sim \operatorname{MVN}\left(\boldsymbol{\lambda}^\star, \frac{1}{n}\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\lambda}^\star)\right) . \tag{8.43}$$

Consequently, each time we observe a particular $\boldsymbol{\Lambda}_{\mathbf{n}} = \hat{\boldsymbol{\lambda}}_{\mathbf{n}}$, we draw a random vector from the multivariate normal distribution given by eq. (8.43). Thus, the independent asymptotic $(1-\alpha)\cdot 100\%$-confidence interval for $\lambda_j^\star$ is given by

$$\hat{\lambda}_j \pm \frac{z_{1-\alpha/2}}{\sqrt{n}} \left(\hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_3\right) . \tag{8.44}$$

One candidate

If we condition on samples in which each observation consists of one candidate, $\mathrm{W} = |\mathbf{C}| = 1$, the likelihood function is given by

$$\mathcal{L}(\boldsymbol{\lambda} \,|\, \bar{t}, \hat{\boldsymbol{\omega}}) = \lambda_1^{\hat{\boldsymbol{\omega}}_{\{1\}}} \lambda_2^{\hat{\boldsymbol{\omega}}_{\{2\}}} \lambda_3^{\hat{\boldsymbol{\omega}}_{\{3\}}} \exp\left(-n\bar{t}\,(\lambda_1 + \lambda_2 + \lambda_3)\right) \tag{8.45}$$

and the log-likelihood function is given by

$$\ell(\boldsymbol{\lambda} \,|\, \bar{t}, \hat{\boldsymbol{\omega}}) = \hat{\boldsymbol{\omega}}_{\{1\}} \ln \lambda_1 + \hat{\boldsymbol{\omega}}_{\{2\}} \ln \lambda_2 + \hat{\boldsymbol{\omega}}_{\{3\}} \ln \lambda_3 - n\bar{t}\,(\lambda_1 + \lambda_2 + \lambda_3) . \tag{8.46}$$

The maximum likelihood estimator is given by

$$\hat{\boldsymbol{\lambda}}_{\mathbf{n}} = \frac{1}{n\bar{t}} \begin{pmatrix} \hat{\boldsymbol{\omega}}_{\{1\}} \\ \hat{\boldsymbol{\omega}}_{\{2\}} \\ \hat{\boldsymbol{\omega}}_{\{3\}} \end{pmatrix} , \tag{8.47}$$

and the information matrix is given by

$$\boldsymbol{\mathcal{I}}\left(\boldsymbol{\lambda}^\star\right) = \frac{1}{\lambda_1^\star + \lambda_2^\star + \lambda_3^\star}\begin{bmatrix} \frac{1}{\lambda_1^\star} & 0 & 0 \\ 0 & \frac{1}{\lambda_2^\star} & 0 \\ 0 & 0 & \frac{1}{\lambda_3^\star} \end{bmatrix}, \tag{8.48}$$

and thus the variance-covariance matrix is given by

$$\boldsymbol{\Sigma_n}\left(\boldsymbol{\lambda}^\star\right) = \frac{\lambda_1^\star + \lambda_2^\star + \lambda_3^\star}{n}\begin{bmatrix} \lambda_1^\star & 0 & 0 \\ 0 & \lambda_2^\star & 0 \\ 0 & 0 & \lambda_3^\star \end{bmatrix}. \tag{8.49}$$

The asymptotic mean squared error is given by the trace of the variance-covariance matrix,

$$\operatorname{MSE}\left(\boldsymbol{\Lambda_n}\right) = \frac{(\hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_3)^2}{n}, \tag{8.50}$$

which has three times less mean squared error than when conditioning on $|\mathbf{C}| = 2$. That is, an $\mathbf{M_n}$ sample in which each observation consists of one candidate has more information about $\boldsymbol{\lambda}^\star$ than a sample in which each observation consists of two candidates.

The independent asymptotic $(1 - \alpha) \cdot 100\%$-confidence interval for $\lambda_j^\star$ is given by

$$\hat{\lambda}_j \pm z_{1-\alpha/2}\sqrt{\frac{\hat{\lambda}_j\left(\hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_3\right)}{n}}. \tag{8.51}$$

By comparison, the confidence interval when conditioning on $|\mathbf{C}| = 2$ is

$$\sqrt{\frac{\hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_3}{\hat{\lambda}_j}} \tag{8.52}$$

as wide.

Three candidates

The degenerate case is all components are candidates, $\mathrm{W} = |\mathbf{C}| = 3$. There is no maximum likelihood estimator $\boldsymbol{\hat{\lambda}_n}$. Rather, solving the equation results in the underspecified system given by

$$\sum_{j=1}^{3} \hat{\lambda}_j - \frac{1}{\bar{t}} = 0, \tag{8.53}$$

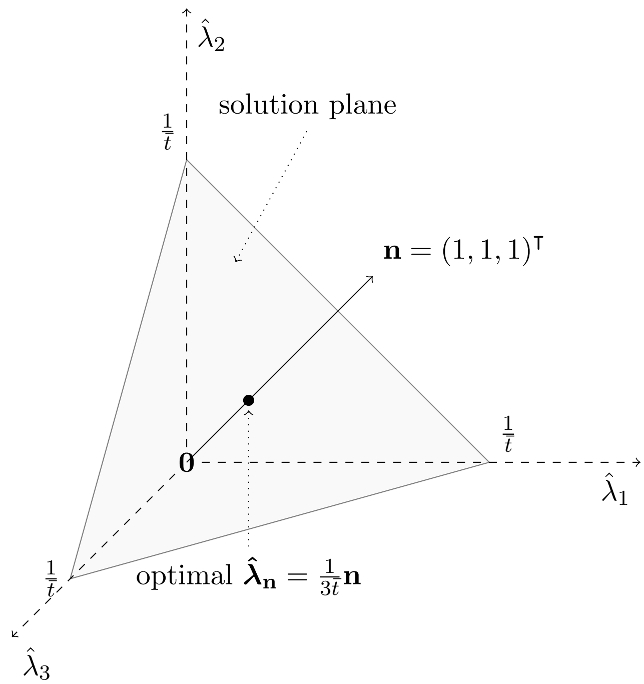
where $\hat{\lambda}_j > 0$ for $j = 1, 2, 3$.

The solution set of this underspecified system is the equation of a plane bounded in the positive region as depicted in fig. 8.1. The plane has a normal vector given by

$$\mathbf{n} = (1, 1, 1)^\intercal, \tag{8.54}$$

which is independent of the statistic $\bar{t}$. Across multiple samples, the statistic $\bar{t}$ varies but the normal vector of the plane does not. Thus, the sampling distribution of the solution set will be a density over planes bounded in the positive region with a normal vector given by eq. (8.54).

Figure 8.1: $\hat{\boldsymbol{\lambda}}_{\mathbf{n}}$ solutions given three candidates



As n goes to infinity, the probability that $\boldsymbol{\lambda}^\star$ intersects the plane goes to 1. This information may not be very useful, especially if the system lifetime S has a small expected value (the area of the bounded plane is $1/\bar{t}^2$), but if this is the only information available, what is the best course of action? An unbiased estimator that asymptotically converges to $\boldsymbol{\lambda}^\star$ is not possible, but we can minimize some other loss function. If we assume $\boldsymbol{\lambda}^\star$ takes on any supported value with equal likelihood, then the estimate that minimizes the expected Euclidean distance from the projection of $\boldsymbol{\lambda}^\star$ onto the plane of the solution set is given by

$$\tilde{\boldsymbol{\lambda}}_{\mathbf{n}} = \frac{1}{3\bar{t}}\mathbf{n}\,. \tag{8.55}$$

Rate of convergence to the asymptotic sampling distribution

In this section, we evaluate how rapidly the sampling distribution of $\hat{\boldsymbol{\theta}}(n \,|\, w)$ converges to the asymptotic distribution as given by **??**, a multivariate normal with a mean $\boldsymbol{\theta}^\star$ and a variance-covariance matrix $\frac{1}{n}\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta}^\star)$.

By eq. (6.16), the asymptotic mean squared error of the maximum likelihood estimator $\hat{\boldsymbol{\theta}}(n \,|\, w)$ is given by the trace of the variance-covariance matrix, $\text{MSE} = \frac{1}{n}\operatorname{tr}\left(\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta}^\star)\right)$. An estimator of the mean squared error from a sample of r maximum likelihood estimates is given by

$$\widehat{\text{MSE}}\left(\hat{\boldsymbol{\theta}}^{(1)}(n \,|\, w), \ldots, \hat{\boldsymbol{\theta}}^{(r)}(n \,|\, w)\right) = \frac{1}{r}\sum_{i=1}^{r}\left(\hat{\boldsymbol{\theta}}^{(i)}(n \,|\, w) - \boldsymbol{\theta}^\star\right)^\intercal\left(\hat{\boldsymbol{\theta}}^{(i)}(n \,|\, w) - \boldsymbol{\theta}^\star\right)\,. \tag{8.56}$$

By generating r masked system failure time of samples of size n, denoted by $\mathbf{M}(n,1), \ldots, \mathbf{M}(r,1)$, and finding the corresponding maximum likelihood estimates, denoted by $\hat{\boldsymbol{\theta}}^{(1)}(n \,|\, w), \ldots, \hat{\boldsymbol{\theta}}^{(r)}(n \,|\, w)$, we may estimate the mean squared error and compare the result against the asymptotic mean squared

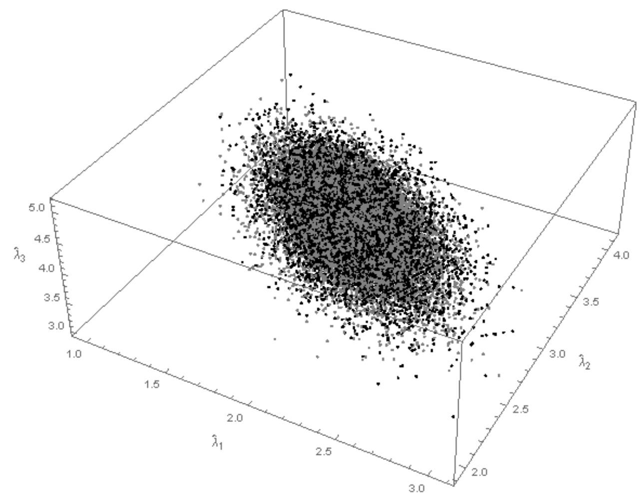


Figure 8.2: Theoretical MLE data points (gray) vs actual MLE data points (black)

error. When we perform these steps for masked system failure time samples of size $n = 1, \ldots, N$, we may plot the rate of convergence.

The absolute difference between the mean squared error of the asymptotic covariance matrix and the "true" covariance matrix is given by

$$|\,\mathrm{MSE} - \hat{\mathrm{MSE}}|\,, \tag{8.57}$$

— Let's use a profile likelihood to look at how the likelihood function changes with respect to a parameter component. Say, for instance, the other cmoponents in the parameter are nuisance parameters, and we are really just interested in the j^{th} component. —

Let $\boldsymbol{\lambda}^\star = (1,2,3)^\intercal$.

and let the masked failure sample be of size $n = 1000$. Theoretically, the MLE $\hat{\boldsymbol{\lambda}}_{\mathbf{1000}}$ is approximately normally distributed,

$$\hat{\boldsymbol{\lambda}}_{\mathbf{1000}} \sim \mathrm{MVN}\Big((2,3,4)^\intercal, \boldsymbol{\Sigma}_{\mathbf{1000}}\,(???)\Big)\,. \tag{8.58}$$

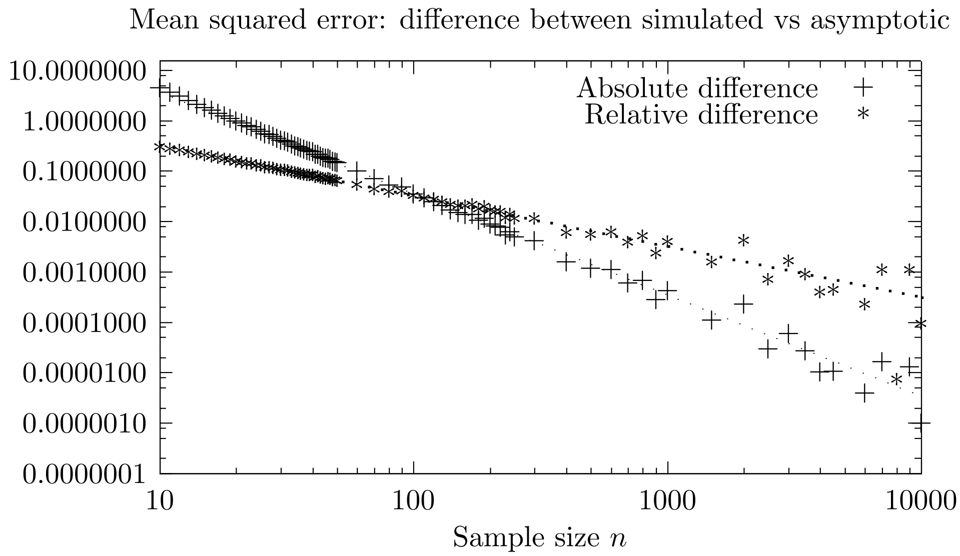
where the variance-covariance matrix is given by

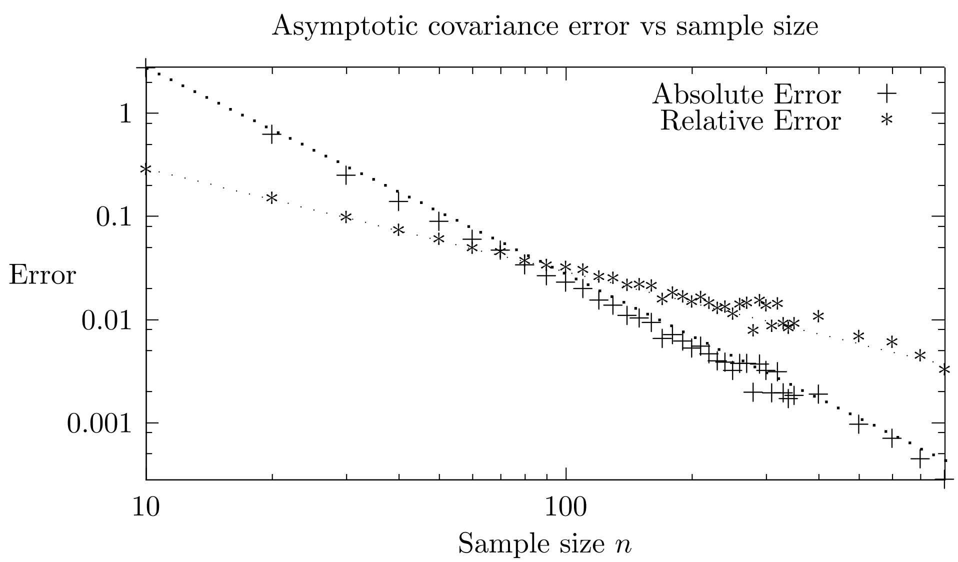
$$\boldsymbol{\Sigma}_{\mathbf{1000}} = \begin{pmatrix} 0.081 & -0.036 & -0.027 \\ -0.036 & 0.081 & -0.018 \\ -0.027 & -0.018 & 0.081 \end{pmatrix}. \tag{8.59}$$

We generated two different sets of data points, one from the theoretical distribution described in eq. (8.58) and the other from the maximum likelihood method applied to $\mathbf{M_{n}}_{1000}$ samples. Refer to fig. 8.2 for a visualization of the data points. Visually, they have a similar distribution of data points, indicating that the asymptotic distribution is a reasonable approximation of the sampling distribution of the MLE $\hat{\boldsymbol{\lambda}}_{1000}$.

A sample variance-covariance matrix of $r = 10000$ maximum likelihood estimates, $\hat{\boldsymbol{\lambda}}^{(\mathbf{1})}_{\mathbf{1000}}, \ldots, \hat{\boldsymbol{\lambda}}^{(\mathbf{10000})}_{\mathbf{1000}}$, is computed to be

$$\boldsymbol{\Sigma_n}\,1000 = \begin{pmatrix} 0.081 & -0.037 & -0.027 \\ -0.037 & 0.082 & -0.018 \\ -0.027 & -0.018 & 0.081 \end{pmatrix}, \tag{8.60}$$





which is approximately equivalent to the theoretical asymptotic distribution given by eq. (8.58).

To derive an accurate estimate of the sampling distribution of $\hat{\boldsymbol{\lambda}}_{\mathbf{n}}$ for small n, as opposed to the asymptotic sampling distribution for large n, we generate r masked system failure time samples

We define the error between two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ to be given by

$$\|\boldsymbol{A}-\boldsymbol{B}\|_F\,, \tag{8.61}$$

where $\|\boldsymbol{C}\|_F$ is the Frobenius norm of matrix $\boldsymbol{C}$,

$$\begin{aligned}\|\boldsymbol{C}\|_F &= \sqrt{\operatorname{tr}\left(\boldsymbol{C}\boldsymbol{C}^{\intercal}\right)} &(8.62)\\ &= \sqrt{\operatorname{vec}\left(\boldsymbol{A}\right)^{\intercal}\operatorname{vec}\left(\boldsymbol{A}\right)}\,. &(8.63)\end{aligned}$$

In section 8.5, we log-log plot the error between a simulated estimate of the true covariance matrix and the asymptotic covariance matrix with respect to sample size. The errors approximately follow a straight line, suggesting a power law error. We fit a power function of the form an^b to the differences with respect to sample size n and found the best fit for the relative difference to be approximately proportional to $\frac{1}{n}$ and the best fit for the absolute difference to be approximately proportional to $\frac{1}{n^2}$.

The asymptotic sampling distribution of $\hat{\boldsymbol{\theta}}(n \,|\, w)$ can be transformed into a Chi-squared distribution, which is given by

$$n(\hat{\boldsymbol{\theta}}(n \,|\, w) - \boldsymbol{\theta}^\star)^\intercal \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}^\star)(\hat{\boldsymbol{\theta}}(n \,|\, w) - \boldsymbol{\theta}^\star) \sim \chi^2_{m \cdot q}\,. \tag{8.64}$$

Therefore, for a given n, we can construct a $(1-\alpha)$ confidence region of this statistic. We expect that $(1-\alpha)$ of the $\hat{\boldsymbol{\theta}}(n \,|\, w)$ point estimates to be inside of this region.

Bibliography

[1] Peter Bickel and Kjell Doksum. *Mathematical Statistics*, volume 1, chapter 2, page 117. Prentice Hall, 2 edition, 2000.

Appendix A

Alternative proofs

A.1 Equation 3.5

Equation (3.5) on page 13 asserts that

$$f_{\mathrm{K,S}}(k, t \,|\, \boldsymbol{\Theta}^\star) = f_k(t \,|\, \boldsymbol{\Theta}_k^\star) \prod_{\substack{j=1 \\ j\neq k}}^{m} R_j(t \,|\, \boldsymbol{\Theta}_j^\star) \,.$$

Proof. First, note that $f_{\mathrm{K,S}}(k, t \,|\, \boldsymbol{\Theta}^\star)$ is a proper density since

$$\int_0^\infty \sum_{j=1}^{m} f_{\mathrm{K,S}}(j, t \,|\, \boldsymbol{\Theta}^\star) \,\mathrm{d}t = \int_0^\infty f_{\mathrm{S}}(t \,|\, \boldsymbol{\Theta}^\star) \,\mathrm{d}t = 1 \,. \tag{a}$$

Consider a 3-out-of-3 system. By assumption 2.2, T_1, T_2, and T_3 are mutually independent. Thus, the joint density that $\mathrm{T}_1 = t_1$, $\mathrm{T}_2 = t_2$, and $\mathrm{T}_3 = t_3$ is given by

$$f_{\mathrm{T}_1,\mathrm{T}_2,\mathrm{T}_3}(t_1, t_2, t_3 \,|\, \boldsymbol{\Theta}^\star) = \prod_{p=1}^{3} f_p(t_p \,|\, \boldsymbol{\Theta}_p^\star) \,. \tag{b}$$

Component 1 causes a system failure during the interval $(0, t)$, $t > 0$, if component 1 fails during the given interval and components 2 and 3 survive longer than component 1. That is, $0 < \mathrm{T}_1 < t$, $\mathrm{T}_1 < \mathrm{T}_2$, and $\mathrm{T}_1 < \mathrm{T}_3$.

Let $\Pr(t)$ denote the probability that component 1 causes a system failure during the interval $(0, t)$,

$$\Pr(t) = \Pr\left[0 < \mathrm{T}_1 < t \cap \mathrm{T}_1 < \mathrm{T}_2 \cap \mathrm{T}_1 < \mathrm{T}_3\right] \,. \tag{c}$$

This probability is given by

$$\Pr(t) = \int_{t_1=0}^{t_1=t} \int_{t_2=t_1}^{t_2=\infty} \int_{t_3=t_1}^{t_3=\infty} \prod_{p=1}^{3} f_p(t_p \,|\, \boldsymbol{\Theta}_p^\star) \,\mathrm{d}t_3 \,\mathrm{d}t_2 \,\mathrm{d}t_1 \,. \tag{d}$$

Performing the integration over t_3 results in

$$\Pr(t) = \int_{t_1=0}^{t_1=t} \int_{t_2=t_1}^{t_2=\infty} \prod_{p=1}^{2} f_p(t_p \,|\, \boldsymbol{\Theta}_p^\star) R_3(t_1 \,|\, \boldsymbol{\Theta}_3^\star) \,\mathrm{d}t_2 \,\mathrm{d}t_1 \,. \tag{e}$$

Performing the integration over t_2 results in

$$\Pr(t) = \int_{t_1=0}^{t_1=t} f_1(t_1 \,|\, \mathbf{\Theta}_1^\star) \prod_{p=2}^{3} R_p(t_1 \,|\, \mathbf{\Theta}_p^\star)\, \mathrm{d}t_1 \,. \tag{f}$$

Taking the derivative of $\Pr(t)$ with respect to t produces the density of probability near t. By the Second Fundamental Theorem of Calculus,

$$\frac{\mathrm{d}\Pr(t)}{\mathrm{d}t} = f_1(t \,|\, \mathbf{\Theta}_1^\star) \prod_{p=2}^{3} R_p(t \,|\, \mathbf{\Theta}_p^\star)\,. \tag{g}$$

The probability density $\frac{\mathrm{d}\Pr(t)}{\mathrm{d}t}$ is equivalent to the joint density that $\mathrm{K} = 1$ and $\mathrm{S} = t$ as given by theorem 3.4, i.e., $\frac{\mathrm{d}P}{\mathrm{d}t} = f_{\mathrm{K,S}}(1, t \,|\, \mathbf{\Theta}^\star)$. Generalizing from this completes the proof. □

A.2 Equation ??

By **??**, S and W are independent and the marginal distribution of W is independent of $\boldsymbol{\theta}^\star$, thus the likelihood of observing a particular realization, $\mathbf{M_n} = \mathbf{m_n}$, is given by

$$\begin{aligned} f_{\mathbf{M_n}}(\mathbf{m_n} \,|\, \boldsymbol{\theta}^\star) &= \prod_{i=1}^{n} f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}\big(\mathbf{c}_i, \mathrm{t_i} \,|\, |\mathbf{c}_i|, \boldsymbol{\theta}^\star\big)\, p_{\mathrm{W}}\big(|\mathbf{c}_i|\big) \\ &= \left[\prod_{i=1}^{n} f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}\big(\mathbf{c}_i, \mathrm{t_i} \,|\, |\mathbf{c}_\mathrm{i}|, \boldsymbol{\theta}^\star\big)\right] \left[\prod_{i=1}^{n} p_{\mathrm{W}}\big(|\mathbf{c}_i|\big)\right] \,. \end{aligned} \tag{A.1}$$

If we fix $\mathbf{m_n}$ and allow the parameter $\boldsymbol{\theta}$ to change, we have the likelihood function

$$\mathcal{L}\left(\boldsymbol{\theta} \,|\, \mathbf{m_n}\right) = c \prod_{i=1}^{n} f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}\big(\mathbf{c}_i, \mathrm{t_i} \,|\, |\mathbf{c}_\mathrm{i}|, \boldsymbol{\theta}^\star\big) \,, \tag{A.2}$$

where the product over $p_{\mathrm{W}}(\,\cdot\,)$ is constant with respect to $\boldsymbol{\theta}$ and has been relabeled as c. The log-likelihood with respect to $\boldsymbol{\theta}$ then is given by

$$\begin{aligned} \ell(\boldsymbol{\theta} \,|\, \mathbf{m_n}) &= \ln c + \sum_{i=1}^{n} \ln f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}(\mathbf{c}_i, \mathrm{t_i} \,|\, |\mathbf{c}_\mathrm{i}|, \boldsymbol{\theta}) \\ &= c' + \sum_{w=1}^{m-1} \left[\sum_{i \in \mathbb{A}(w)} \ln f_{\mathbf{C},\mathrm{S}\,|\,\mathrm{W}}(\mathbf{c}_i, \mathrm{t_i} \,|\, \mathrm{w}, \boldsymbol{\theta})\right] \,, \end{aligned} \tag{A.3}$$

where $\mathbb{A}(w) = \{i \in \{1, \ldots, n\} \colon |\mathbf{c}_i| = w\}$. Let $\mathbf{m_{n_w}} = \{(\mathbf{c}_i, \mathrm{t_i}) \colon \mathrm{i} \in \mathbb{A}(\mathrm{w})\}$. The part in brackets is the log-likelihood $\ell(\boldsymbol{\theta} \,|\, \mathbf{m_{n_w}})$. Performing this substitution yields

$$\ell(\boldsymbol{\theta} \,|\, \mathbf{m_n}) = \ln c + \sum_{w=1}^{m-1} \ell(\boldsymbol{\theta} \,|\, \mathbf{m_{n_w}})\,. \tag{A.4}$$

The score function is given by

$$\mathrm{s}\left(\boldsymbol{\theta} \,|\, \mathbf{m_n}\right) = \sum_{w=1}^{m-1} \mathrm{s}\left(\boldsymbol{\theta} \,|\, w, \mathbf{m_{n_w}}\right) \,, \tag{A.5}$$

The information matrix may be computed by taking the negative of the expectation of the Jacobian of the score over the true joint distribution of S, $\mathbf{C}$, and W, giving

$$\boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^{\star}\right) = -\,\mathbb{E}_{\boldsymbol{\theta}^{\star}}\left[\mathcal{J}\left(\mathrm{s}\left(\boldsymbol{\theta}\,|\,\mathbf{M_1}\right)\right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\star}}\right]. \tag{A.6}$$

Substituting eq. (A.5) with its definition gives

$$\boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^{\star}\right) = -\,\mathbb{E}_{\boldsymbol{\theta}^{\star}}\left[\mathcal{J}\left(\sum_{w=1}^{m-1}\mathrm{s}\left(\boldsymbol{\theta}\,|\,w,\mathbf{M}(w,1)\right)\right)\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\star}}\right]. \tag{A.7}$$

The Jacobian is a linear operator so it can be moved inside the summation, giving

$$\boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^{\star}\right) = -\,\mathbb{E}_{\boldsymbol{\theta}^{\star}}\left[\sum_{w=1}^{m-1}\mathcal{J}\left(\mathrm{s}\left(\boldsymbol{\theta}\,|\,w,\mathbf{M}(w,1)\right)\right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\star}}\right]. \tag{A.8}$$

The Jacobian of s is the Hessian of ℓ. Performing this substitution gives

$$\boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^{\star}\right) = -\,\mathbb{E}_{\boldsymbol{\theta}^{\star}}\left[\sum_{w=1}^{m-1}\mathcal{H}\left(\ell\left(\boldsymbol{\theta}\,|\,w,\mathbf{M}(w,1)\right)\right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\star}}\right]. \tag{A.9}$$

The expectation is a linear operator so it may be moved inside the summation, giving

$$\boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^{\star}\right) = -\sum_{w=1}^{m-1}\mathbb{E}_{\boldsymbol{\theta}^{\star}}\left[\mathcal{H}\left(\ell\left(\boldsymbol{\theta}\,|\,w,\mathbf{M}(w,1)\right)\right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\star}}\right]. \tag{A.10}$$

The expectation sums over the probability mass function $\mathrm{W} \sim p_{\mathrm{W}}(\,\cdot\,)$, giving

$$\boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^{\star}\right) = -\sum_{w=1}^{m-1}p_{\mathrm{W}}(w)\,\mathbb{E}_{\boldsymbol{\theta}^{\star}}\left[\mathcal{H}\left(\ell\left(\boldsymbol{\theta}\,|\,w,\mathbf{M}(w,1)\right)\right)\right], \tag{A.11}$$

where the expectation is now over the true marginal joint distribution of $\mathbf{C}$ and S given $\mathrm{W} = w$. The expectation of the Hessian of ℓ is equivalent to the negative of $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}^{\star}\,|\,w,\mathbf{M}(w,1))$. Performing this substitution gives

$$\boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^{\star}\right) = \sum_{w=1}^{m-1}p_{\mathrm{W}}(w)\,\boldsymbol{\mathcal{I}}\left(\boldsymbol{\theta}^{\star}\,|\,w\right). \tag{A.12}$$

$\square$

Appendix B

Numerical solutions to the MLE

The function $\ell(\boldsymbol{\theta} \mid w, \mathbf{m_n})$ is the log-likelihood with respect to $\boldsymbol{\theta} \in \Omega$ where $\Omega \subset \mathbb{R}^{m \cdot q}$. This function has a surface in an $(m \cdot q + 1)$ dimensional space, where a particular point on this surface represents the log-likelihood of observing $\mathbf{m_n}$ with respect to $\boldsymbol{\theta}$. By definition 6.3, $\hat{\boldsymbol{\theta}}(n \mid w)$ is the point on this surface that is at a maximum,

$$\hat{\boldsymbol{\theta}}(n \mid w) = \underset{\boldsymbol{\theta} \in \Omega}{\arg\max}\, \ell(\boldsymbol{\theta} \mid w, \mathbf{m_n})\,. \tag{6.9 revisited}$$

Generally there is no closed-form solution that solves $\hat{\boldsymbol{\theta}}(n \mid w)$, in which case iterative search methods may be used to numerically approximate a solution.

The general version of iterative search that numerically approximates a solution to eq. (6.10), subject to the constraint $\boldsymbol{\theta}^{\star} \in \Omega$, is shown in algorithm 5. Since iterative search is a local search method, it may fail to converge to a global maximum.

Algorithm 5: Iterative maximum likelihood search

Result: an approximate solution to the stationary points of the maximum likelihood equation

Input:

$\boldsymbol{\Theta}^{\star}$, the true parameter index.

Output:

$\hat{\boldsymbol{\theta}}(n \mid w)$, an approximation of the maximum likelihood estimate.

Model `FindMLE`($\mathbf{m_n}$)

  $\hat{\boldsymbol{\theta}}^{(0)}(n \mid w) \leftarrow$ an initial starting point in Ω

  $i \leftarrow 0$

  while stopping criteria *not satisfied* **do**

    $\hat{\boldsymbol{\theta}}^{(i+1)}(n \mid w) \leftarrow \texttt{Project}\left(\hat{\boldsymbol{\theta}}^{(i)}(n \mid w) + \alpha^{(i)}\mathbf{d^{(i)}}, \Omega\right)$

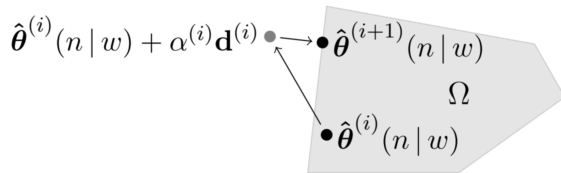
    $i \leftarrow i + 1$

  end

  return $\hat{\boldsymbol{\theta}}^{(i)}(n \mid w)$

Assume parameter space Ω is convex. Then, the function `Project`$(\boldsymbol{\theta}, \Omega)$ in algorithm 5 projects any point $\boldsymbol{\theta}$ to the nearest point in the parameter space Ω as depicted by fig. B.1. Thus, the search method is restricted to searching over the feasible parameter space.

Figure B.1: Projection onto convex set Ω.



In algorithm 5, $\mathbf{d}^{(i)}$ is a unit vector denoting a *promising* direction in which to search for a better solution than $\hat{\boldsymbol{\theta}}^{(i)}(n \mid w)$. The Fisher scoring algorithm is a slight variation of Newton-Raphson[1] in which

$$\mathbf{d}^{(i)} = \boldsymbol{\mathcal{I}}^{-1}\left(\hat{\boldsymbol{\theta}}^{(i)}(n \mid w)\right) \mathrm{s}\left(\hat{\boldsymbol{\theta}}^{(i)}(n \mid w)\right) . \tag{B.1}$$

The scalar $\alpha^{(i)} > 0$ is the distance to move from point $\hat{\boldsymbol{\theta}}^{(i)}(n \mid w)$ to generate the next point, $\hat{\boldsymbol{\theta}}^{(i+1)}(n \mid w)$. Most naturally, the solution to

$$\alpha^{(i)} = \arg\max_{\alpha > 0} \ell(\hat{\boldsymbol{\theta}}^{(i)}(n \mid w) + \alpha \cdot \mathbf{d}^{(i)}) \tag{B.2}$$

is desired, e.g., using *Golden Section* line search to find an approximate solution. Finally, the iterations are repeated until some *stopping criteria* is satisfied. Most naturally, this is given by

$$\| \mathrm{s}(\hat{\boldsymbol{\theta}}^{(i)}(n \mid w)) \| < \epsilon \, , \tag{B.3}$$

signifying a stationary point has been reached.

[1] If the *observed* information matrix is used instead of the *expected* information matrix, the Fisher scoring algortihm is equivalent to Newton-Raphson.

Appendix C

Generic joint distribution

We assume the distribution of W does not carry information about the true parameter index $\boldsymbol{\theta}^\star$. By the axioms of probability, the joint distribution of $\mathbf{C}$, S, and W is given by

$$
\begin{aligned}
&f_{\mathbf{C},\mathrm{S},\mathrm{W}}(\mathbf{c}, t, w \,|\, \boldsymbol{\theta}^\star) = \\
&\qquad p_{\mathbf{C}\,|\,\mathrm{S},\mathrm{W}}(\mathbf{c} \,|\, t, w, \boldsymbol{\theta}^\star) p_{\mathrm{W}\,|\,\mathrm{S}}(w \,|\, t) f_{\mathrm{S}}(t \,|\, \boldsymbol{\theta}^\star)\,.
\end{aligned}
\tag{C.1}
$$

Since W carries no information about the true parameter index $\boldsymbol{\theta}^\star$, the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_{\mathbf{n}}$ is independent of the distribution of W.

Proof.

$$
\begin{aligned}
\hat{\boldsymbol{\theta}}_{\mathbf{n}} &= \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta} \,|\, \mathbf{m}_{\mathbf{n}}) \\
&= \arg\max_{\boldsymbol{\theta}} \Bigg[\sum_{i=1}^{n} \ln p_{\mathbf{C}\,|\,\mathrm{S},\mathrm{W}} \left(\mathbf{c}_i \,|\, t_i, |\mathbf{c}_i|, \boldsymbol{\theta} \right) \\
&\qquad + \sum_{i=1}^{n} \ln f_{\mathrm{S}} \left(t_i \,|\, \boldsymbol{\theta} \right) + \underbrace{\sum_{i=1}^{n} \ln p_{\mathrm{W}\,|\,\mathrm{S}} \left(|\mathbf{c}_i| \,|\, t_i \right)}_{\text{constant}} \Bigg]\,,
\end{aligned}
\tag{a}
$$

where the last term is constant with respect to $\boldsymbol{\theta}$ and may be dropped from the maximum likelihood equation without changing the point that maximizes the likelihood. □

The sampling distribution of $\hat{\boldsymbol{\theta}}_{\mathbf{n}}$ is a function of $\mathbf{M}_{\mathbf{n}}$ (as opposed to a particular realization), and since $\mathbf{M}_{\mathbf{n}}$ is a function of W (the distribution of the cardinality of candidate sets), the sampling distribution of $\hat{\boldsymbol{\theta}}_{\mathbf{n}}$ is also a function of W.

Given a statistical model, the asymptotic sampling distribution of the maximum likelihood estimator is nearly automatic. The most taxing (and uncertain) part is the *model selection*. Since we may not have much confidence in any particular model of the joint distribution of W and S, the *expected* information matrix is sketchy. In general, the *observed* information matrix is preferred. The sampling distribution of $\hat{\boldsymbol{\theta}}_{\mathbf{n}}$ converges in distribution to a multivariate normal with a mean given by $\boldsymbol{\theta}^\star$ and a variance-covariance given by the inverse of the *observed* information matrix, written

$$
\mathbf{Y}_{\mathbf{n}} \xrightarrow{d} \mathrm{MVN}\left(\boldsymbol{\theta}^\star, \boldsymbol{\mathcal{J}}_{\boldsymbol{n}}^{-1} \left(\boldsymbol{\theta}^\star \,|\, \mathbf{M}_{\mathbf{n}} \right) \right)\,. \tag{C.2}
$$

If an accurate distribution of $\mathrm{W} \,|\, \mathrm{S} = \mathrm{t}$ could be constructed such that it is dependent on $\boldsymbol{\theta}^\star$, then the distribution of W in a sample carries extra information about $\boldsymbol{\theta}^\star$. A plausible model of W would depend upon the entropy of $\mathrm{K} \,|\, \mathrm{S} = \mathrm{t}$.