Bootstrapping statistics of the maximum likelihood estimator of components in a series systems from masked failure data

Alex Towell

Abstract

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

Contents

1 Introduction

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

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

1. Deriving a likelihood model that accounts for right-censoring and masked failure causes through candidate sets. This allows the available masked data to be used for estimation.

2. Validating the accuracy, precision, and robustness of the maximum likelihood estimator through an extensive simulation study under different sample sizes, masking probabilities, and censoring levels.

3. Demonstrating that bootstrapping provides well-calibrated confidence intervals for the MLEs even with small samples.

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

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

2 Series System Model

We consider a system composed of m components arranged in a series configuration. Each component and system has only two possible states, functioning or failed. The lifetime of the j^{th} component in the i^{th} system is denoted by the random variable T_{ij}. We assume the component lifetimes in a system are statistically independent and non-identically distributed. Here, lifetime is defined as the elapsed time from when the new, functioning component is put into operation until it fails for the first time.

A series system fails when any component fails. Thus, the lifetime of the i^{th} system, T_i, is given by the component with the shortest lifetime,

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

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

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

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

Definition 1. *The survival function, $R_T(t)$, of a random lifetime T is the probability that it realizes a value larger than some specified duration of time t,*

$$R_T(t) = \Pr\{T > t\} \tag{2.1}$$

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

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

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

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

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

Definition 2. *The hazard function $h_T(t)$ for a continuous random variable T is the instantaneous failure rate at time t, which is given by*

$$\begin{aligned} h_T(t) &= \lim_{\Delta t \to 0} \frac{\Pr\{t < T < t + \Delta t\}}{\Delta t} \frac{1}{\Pr\{T > t\}} \\ &= \frac{f_T(t)}{R_T(t)}. \end{aligned} \tag{2.2}$$

If the i^{th} system has the parameter vector $\boldsymbol{\theta}$, we denote its pdf by $f_{T_i}(t; \boldsymbol{\theta})$ and likewise for other distribution functions, e.g., its relability function is denoted by $R_{T_i}(t; \boldsymbol{\theta})$. We denote the pdf of the j^{th} component by $f_j(t; \boldsymbol{\theta_j})$ and its reliability function by $R_j(t; \boldsymbol{\theta_j})$. If it is clear from the context which random variable a distribution function is for, we may drop the subscripts, e.g., $F(t)$ instead of $F_T(t)$. Finally, as an abuse of notation, we often write a function as $f(t)$ when we really mean that f is a function of variable t.

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

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

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

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

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

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

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

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

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

Proof. The reliability function is defined as

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

which may be rewritten as

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

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

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

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

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

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

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

□

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

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

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

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

Proof. By definition, the pdf may be written as

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

By the product rule, this may be rewritten as

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

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

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

which simplifies to

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

□

Theorem 2 shows the pdf of the system lifetime as a function of the pdfs and reliabilities of its components. We continue with the hazard function of the system lifetime, defined in the next theorem.

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

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

Proof. The i^{th} series system lifetime has a hazard function defined as

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

Plugging in expressions for these functions results in

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

which can be simplified to

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

□

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

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

and we can rearrange this to get

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

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

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

2.1 Component Cause of Failure

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

Definition 3. *The component cause of failure of a series system is denoted by the random variable K_i whose support is given by $\{1, \ldots, m\}$. For example, $K_i = j$ indicates that the component indexed by j failed first, i.e.,*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Generalizing from this completes the proof. $\square$