Masked Causes of Failure in Series Systems: A Likelihood Framework

Alexander Towell
lex@metafunctor.com
ORCID: 0000-0001-6443-9897

February 2026

Abstract

We develop a general likelihood framework for estimating component reliability from series system data when the component cause of failure is masked. The framework applies to any parametric specification of component hazard functions, including covariate-dependent hazards. Three sufficient conditions on the masking mechanism—that the candidate set contains the true cause, that masking probabilities are symmetric across candidates, and that masking probabilities are independent of the component lifetime parameters—allow the unknown masking distribution to be eliminated from the likelihood. We present the resulting likelihood contributions for exact failures with masked cause, right-censored, left-censored, and interval-censored observations. The framework serves as a foundation for distribution-specific inference, and we provide a summary of instantiations for five common lifetime distribution families.

1 Introduction

Estimating the reliability of individual components within a series system is a fundamental problem in reliability engineering [Agustin, 2011]. A series system fails when any one of its components fails, so the system lifetime is determined by the weakest component. In many practical settings, only the system-level failure time is observable—the specific component that caused the failure may be unknown or only partially identified. This *masking* of the failure cause arises naturally in industrial diagnostics, field warranty data, and accelerated life testing, where post-failure inspection is infeasible, costly, or imprecise.

A common diagnostic outcome is a *candidate set*: a subset of components that plausibly contains the failed component. When the candidate set is a proper subset of all components but not a singleton, the failure cause is partially masked. When the candidate set is the full component set, the cause is fully masked. When it is a singleton, the cause is exactly identified.

The purpose of this paper is to provide a self-contained reference for the likelihood framework for masked failure data in series systems. We present the likelihood under three sufficient conditions (C1–C2–C3) on the masking mechanism that allow the unknown distribution of candidate sets to be eliminated from the likelihood function. The resulting likelihood is expressed entirely in terms of component reliability and hazard functions, enabling maximum likelihood estimation for any parametric specification of component hazard functions.

The framework is deliberately general: we derive the general likelihood structure in terms of component hazard functions, without specializing to any particular distributional form. This work grew out of an earlier master's project [Towell, 2023b] that developed the likelihood model for

Weibull series systems with simulation studies; the present paper extracts and generalizes the core likelihood framework. Distribution-specific treatments—including derivations of score equations, Fisher information, and simulation studies—are deferred to companion papers that cite the present work. In Section 7, we provide hazard function specifications for five common families (Exponential, Weibull, Pareto, Log-normal, and Gamma), enabling practitioners to apply the framework directly.

1.1 Related Work

The C1–C2–C3 conditions and the basic masked-data likelihood have a substantial history in the reliability literature. Miyakawa [1984] introduced the conditions for analyzing incomplete competing risks data. Usher and Hodgson [1988] formulated the MLE problem for masked series system data under these conditions, and Usher et al. [1993] derived exact maximum likelihood estimates for exponential components. Guess et al. [1991] established that the conditions hold in many practical diagnostic scenarios and developed component reliability estimation under partial failure information. Sarhan [2001] extended reliability estimation to broader settings with masked system life data. Most of this prior work specialized to the Exponential or Weibull families.

The present paper does not claim the C1–C2–C3 conditions or the basic likelihood derivation as novel contributions. Rather, its role is to serve as a *foundational reference* that provides a unified treatment of the framework in a single self-contained document. The value lies in the unified presentation, the extension of the likelihood to all four censoring types (exact, right, left, and interval), the identifiability analysis, and the five-family instantiation table (Section 7). Companion papers and software packages can cite this work for the general theory while focusing on distribution-specific derivations and simulation studies.

Non-parametric competing risks methods—including the Kaplan–Meier estimator, the Nelson–Aalen cumulative hazard estimator, and the Cox proportional hazards partial likelihood—require observing *which* event type occurred in order to separate cause-specific hazard contributions. Masking eliminates precisely this information: the candidate set does not identify the cause, so partial likelihood and non-parametric estimators cannot be applied directly. The parametric structure of $h_j(t; \boldsymbol{\theta}_j)$ is what makes the problem tractable under masking—it provides enough structure for the likelihood to separate component hazards even when the data cannot. Our framework is therefore parametric by necessity (for identifiability under masking), not merely by convenience.

The remainder of this paper is organized as follows. Section 2 establishes the series system model and derives the system reliability, density, and hazard functions. Section 3 derives the distribution of the component cause of failure. Section 4 defines the observational model, including the masked data notation and a taxonomy of observation types. Section 5 presents the three conditions and derives the core likelihood contribution—the central result of this paper. Section 6 discusses maximum likelihood estimation in the general framework. Section 7 provides hazard function specifications for common parametric families. Sections 8 and 9 discuss extensions, relaxations, and concluding remarks. Table 1 summarizes the principal notation used throughout the paper.

2 Series System Model

Consider a system composed of m components arranged in a series configuration. Each component and system has two possible states: functioning or failed. We assume throughout that all component lifetime distributions are absolutely continuous with respect to Lebesgue measure (i.e., each T_{ij} has a density). We observe n independent systems (which need not be identically distributed, since systems may differ by covariates). The lifetime of the ith system is denoted by the random variable

Table 1: Summary of notation.

Symbol	Description
m	Number of components in the series system
T_i	System lifetime (random variable) for the ith system
T_{ij}	Lifetime of component j in system i
K_i	Component cause of system failure
$h_j(t;\boldsymbol{\theta}_j)$	Hazard function of component j
$R_j(t;\boldsymbol{\theta}_j)$	Reliability function of component j
$f_j(t;\boldsymbol{\theta}_j)$	Density function of component j
$\boldsymbol{\theta}$	Full parameter vector $(\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_m)$
$D_i=(s_i,\omega_i,c_i,\boldsymbol{x}_i)$	Observation tuple for system i
s_i	Observed time (t_i, τ_i, or interval (a_i,b_i))
$\omega_i\in\{E,R,L,I\}$	Observation type label
$c_i\subseteq\{1,\ldots,m\}$	Candidate set (possible failure causes)
β_i	Masking probability for observation i
$\mathcal{E},\mathcal{R},\mathcal{L},\mathcal{I}$	Index sets by observation type
$\ell(\boldsymbol{\theta})$	Log-likelihood function

T_i and the lifetime of its jth component by T_{ij}. We assume that the component lifetimes within a single system are statistically independent but not necessarily identically distributed.

Remark 1 (Component independence)**.** Component independence is a standard assumption in the series system and competing risks literature, but it rules out common-cause failures, load-sharing, and environmental coupling. The entire framework developed below—system reliability as a product of component reliabilities (Theorem 1), system hazard as a sum of component hazards (Theorem 3), and the cause-of-failure distribution (Theorem 6)—depends critically on this assumption. Extensions to dependent competing risks exist but require copula models or other dependence structures and face fundamental identifiability challenges: Tsiatis [1975] showed that, without independence, marginal component distributions are not identifiable from system lifetime data alone. Such extensions are beyond the scope of this paper.

A series system fails when any component fails, so the system lifetime is

$$T_i=\min\{T_{i1},T_{i2},\ldots,T_{im}\}. \tag{1}$$

The reliability function of the ith system is $R_{T_i}(t)=\Pr\{T_i>t\}$, the probability that the system survives beyond time t. The probability density function (pdf) of T_i is $f_{T_i}(t)=-\frac{d}{dt}R_{T_i}(t)$, and the hazard function is

$$h_{T_i}(t)=\frac{f_{T_i}(t)}{R_{T_i}(t)}, \tag{2}$$

representing the instantaneous failure rate at time t given survival to time t.

Each component's lifetime distribution is specified by its hazard function $h_j(t;\boldsymbol{\theta}_j,\boldsymbol{x}_i)$, where $\boldsymbol{\theta}_j$ is a finite-dimensional parameter vector and $\boldsymbol{x}_i$ is an optional covariate vector for the ith system.

The cumulative hazard, reliability, and density follow:

$$H_j(t;\boldsymbol{\theta}_j,\boldsymbol{x}_i) = \int_0^t h_j(u;\boldsymbol{\theta}_j,\boldsymbol{x}_i)\,du, \tag{3}$$
$$R_j(t;\boldsymbol{\theta}_j,\boldsymbol{x}_i) = \exp\bigl\{-H_j(t;\boldsymbol{\theta}_j,\boldsymbol{x}_i)\bigr\}, \tag{4}$$
$$f_j(t;\boldsymbol{\theta}_j,\boldsymbol{x}_i) = h_j(t;\boldsymbol{\theta}_j,\boldsymbol{x}_i)\,R_j(t;\boldsymbol{\theta}_j,\boldsymbol{x}_i). \tag{5}$$

This specification subsumes standard parametric families and accommodates non-standard hazard shapes (bathtub curves, piecewise-constant rates) and covariate-dependent hazards such as proportional hazards models. When covariates are absent, we suppress $\boldsymbol{x}_i$ and write $h_j(t;\boldsymbol{\theta}_j)$.

The overall parameter vector is

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_m), \tag{6}$$

belonging to a parameter space $\boldsymbol{\Omega}$.

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

$$R_{T_i}(t;\boldsymbol{\theta}) = \prod_{j=1}^m R_j(t;\boldsymbol{\theta}_j). \tag{7}$$

Proof. Since the system fails when any component fails, $\{T_i > t\} = \{T_{i1} > t,\ldots,T_{im} > t\}$. By the independence of component lifetimes,

$$R_{T_i}(t;\boldsymbol{\theta}) = \Pr\{T_{i1} > t\}\cdots\Pr\{T_{im} > t\} = \prod_{j=1}^m R_j(t;\boldsymbol{\theta}_j). \qquad \square$$

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

$$f_{T_i}(t;\boldsymbol{\theta}) = \sum_{j=1}^m f_j(t;\boldsymbol{\theta}_j) \prod_{\substack{k=1\\k\neq j}}^m R_k(t;\boldsymbol{\theta}_k). \tag{8}$$

Proof. Differentiating the system reliability function,

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

By the product rule applied recursively,

$$-\frac{d}{dt}\prod_{j=1}^m R_j(t;\boldsymbol{\theta}_j) = \sum_{j=1}^m \Bigl(-\frac{d}{dt}R_j(t;\boldsymbol{\theta}_j)\Bigr)\prod_{\substack{k=1\\k\neq j}}^m R_k(t;\boldsymbol{\theta}_k) = \sum_{j=1}^m f_j(t;\boldsymbol{\theta}_j)\prod_{\substack{k=1\\k\neq j}}^m R_k(t;\boldsymbol{\theta}_k). \qquad \square$$

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

$$h_{T_i}(t;\boldsymbol{\theta}) = \sum_{j=1}^m h_j(t;\boldsymbol{\theta}_j). \tag{9}$$

Proof. By definition, $h_{T_i}(t;\boldsymbol{\theta}) = f_{T_i}(t;\boldsymbol{\theta})/R_{T_i}(t;\boldsymbol{\theta})$. Substituting from Theorems 2 and 1,

$$h_{T_i}(t;\boldsymbol{\theta}) = \frac{\sum_{j=1}^m f_j(t;\boldsymbol{\theta}_j)\prod_{\substack{k=1\\k\neq j}}^m R_k(t;\boldsymbol{\theta}_k)}{\prod_{j=1}^m R_j(t;\boldsymbol{\theta}_j)} = \sum_{j=1}^m \frac{f_j(t;\boldsymbol{\theta}_j)}{R_j(t;\boldsymbol{\theta}_j)} = \sum_{j=1}^m h_j(t;\boldsymbol{\theta}_j). \qquad \square$$

Remark 2. Combining the hazard and reliability representations, the system pdf admits the convenient form

$$f_{T_i}(t;\boldsymbol{\theta}) = h_{T_i}(t;\boldsymbol{\theta})\, R_{T_i}(t;\boldsymbol{\theta}) = \left\{\sum_{j=1}^{m} h_j(t;\boldsymbol{\theta}_j)\right\}\left\{\prod_{j=1}^{m} R_j(t;\boldsymbol{\theta}_j)\right\}, \tag{10}$$

which we use extensively in subsequent derivations.

3 Component Cause of Failure

Whenever a series system fails, precisely one component is the cause (almost surely, since the component lifetime distributions are assumed to be absolutely continuous with respect to Lebesgue measure). We denote the component cause of failure of the ith system by the discrete random variable K_i, with support $\{1,\ldots,m\}$. The event $K_i = j$ means that component j had the shortest lifetime among all components of the ith system.

Theorem 4 (Joint distribution of K_i and T_i)**.** *The joint pdf of the component cause of failure K_i and the system lifetime T_i is*

$$f_{K_i,T_i}(j,t;\boldsymbol{\theta}) = h_j(t;\boldsymbol{\theta}_j)\prod_{l=1}^{m} R_l(t;\boldsymbol{\theta}_l). \tag{11}$$

Proof. The event $\{K_i = j, T_i = t\}$ requires that component j fails at time t while all other components survive past time t. By the independence of component lifetimes,

$$f_{K_i,T_i}(j,t;\boldsymbol{\theta}) = f_j(t;\boldsymbol{\theta}_j)\prod_{\substack{l=1\\ l\neq j}}^{m} R_l(t;\boldsymbol{\theta}_l).$$

Since $f_j(t;\boldsymbol{\theta}_j) = h_j(t;\boldsymbol{\theta}_j)R_j(t;\boldsymbol{\theta}_j)$, substituting gives

$$f_{K_i,T_i}(j,t;\boldsymbol{\theta}) = h_j(t;\boldsymbol{\theta}_j)R_j(t;\boldsymbol{\theta}_j)\prod_{\substack{l=1\\ l\neq j}}^{m} R_l(t;\boldsymbol{\theta}_l) = h_j(t;\boldsymbol{\theta}_j)\prod_{l=1}^{m} R_l(t;\boldsymbol{\theta}_l). \qquad \square$$

Theorem 5 (Marginal probability of cause)**.** *The probability that the jth component is the cause of failure is*

$$\Pr\{K_i = j\} = \mathbb{E}_{\boldsymbol{\theta}}\left[\frac{h_j(T_i;\boldsymbol{\theta}_j)}{\sum_{l=1}^{m} h_l(T_i;\boldsymbol{\theta}_l)}\right]. \tag{12}$$

Proof. Marginalizing the joint pdf over the system lifetime,

$$\Pr\{K_i = j\} = \int_0^\infty f_{K_i,T_i}(j,t;\boldsymbol{\theta})\,dt = \int_0^\infty h_j(t;\boldsymbol{\theta}_j)\,R_{T_i}(t;\boldsymbol{\theta})\,dt.$$

Since $R_{T_i}(t;\boldsymbol{\theta}) = f_{T_i}(t;\boldsymbol{\theta})/h_{T_i}(t;\boldsymbol{\theta})$, we can rewrite this as

$$\Pr\{K_i = j\} = \int_0^\infty \frac{h_j(t;\boldsymbol{\theta}_j)}{h_{T_i}(t;\boldsymbol{\theta})}\, f_{T_i}(t;\boldsymbol{\theta})\,dt = \mathbb{E}_{\boldsymbol{\theta}}\left[\frac{h_j(T_i;\boldsymbol{\theta}_j)}{\sum_{l=1}^{m} h_l(T_i;\boldsymbol{\theta}_l)}\right]. \qquad \square$$

Theorem 6 (Conditional probability of cause given failure time)**.** *Given that the system fails at time t_i, the probability that the jth component is the cause is*

$$\Pr\{K_i = j \mid T_i = t_i\} = \frac{h_j(t_i; \boldsymbol{\theta}_j)}{\sum_{l=1}^m h_l(t_i; \boldsymbol{\theta}_l)}. \tag{13}$$

Proof. By the definition of conditional probability,

$$\Pr\{K_i = j \mid T_i = t_i\} = \frac{f_{K_i,T_i}(j, t_i; \boldsymbol{\theta})}{f_{T_i}(t_i; \boldsymbol{\theta})} = \frac{h_j(t_i; \boldsymbol{\theta}_j) \prod_{l=1}^m R_l(t_i; \boldsymbol{\theta}_l)}{h_{T_i}(t_i; \boldsymbol{\theta}) \prod_{l=1}^m R_l(t_i; \boldsymbol{\theta}_l)} = \frac{h_j(t_i; \boldsymbol{\theta}_j)}{\sum_{l=1}^m h_l(t_i; \boldsymbol{\theta}_l)}. \qquad \square$$

Theorem 6 is a conditional version of Theorem 5; it is also the operationally important result, since the likelihood (Section 5) depends on the conditional cause probability given the observed failure time rather than on the marginal cause probability of Theorem 5. The conditional probability of cause is determined entirely by the ratio of the component's hazard to the total system hazard at the observed failure time. In a well-designed series system, the designer aims to balance failure rates so that no single component dominates, resulting in roughly balanced cause-of-failure probabilities across components, though this ideal is not always achieved in practice.

4 Observational Model

In practice, we do not observe the component lifetimes directly. Instead, we observe system-level data that may be subject to two forms of masking: censoring of the failure time and masking of the failure cause.

4.1 Observation Types

We consider four types of observations that may arise in reliability studies of series systems [see Klein and Moeschberger, 2005, for background on censored data in survival analysis]:

1. **Exact failure with candidate set.** The system is observed to fail at time t_i, and a candidate set $c_i \subseteq \{1, \ldots, m\}$ is observed that is indicative of the component cause of failure. When $|c_i| = 1$, the cause is exactly identified; when $|c_i| > 1$, the cause is masked.

2. **Right-censored.** The system is still functioning at observation time τ_i; we know only that $T_i > \tau_i$.

3. **Left-censored.** The system has already failed before observation time τ_i; we know only that $T_i \leq \tau_i$. A candidate set c_i may also be observed if a diagnostic is performed at inspection time τ_i to partially identify the failure cause.

4. **Interval-censored.** The system failure occurred in the interval (a_i, b_i); we know only that $a_i < T_i \leq b_i$. A candidate set c_i may accompany the observation when a diagnostic is performed at one of the inspection times.

4.2 Masked Data Notation

Each observation is represented as a tuple $D_i = (s_i, \omega_i, c_i, \boldsymbol{x}_i)$, where:

- s_i encodes the observed time information: a failure time t_i, a censoring time τ_i, or an interval (a_i, b_i);

- $\omega_i \in \{E, R, L, I\}$ is a label indicating the observation type: exact failure (E), right-censored (R), left-censored (L), or interval-censored (I);
- $c_i \subseteq \{1, \ldots, m\}$ is the candidate set, relevant whenever a failure is known to have occurred (exact, left-censored, or interval-censored observations). When no cause information is available, $c_i = \{1, \ldots, m\}$. For right-censored observations (no failure observed), $c_i = \emptyset$ (no component has failed); and
- $\boldsymbol{x}_i$ is the covariate vector for the ith system (possibly empty when no covariates are recorded).

The complete data set is $D = \{D_1, \ldots, D_n\}$, assumed to be independent draws from the observational model.

4.3 Likelihood Contributions by Observation Type

Each observation type contributes differently to the likelihood function $L(\boldsymbol{\theta}) = \prod_{i=1}^{n} L_i(\boldsymbol{\theta})$. The likelihood contributions depend on three conditions developed in Section 5; Table 2 previews the final results for reference.

Table 2: Likelihood contributions by observation type under Conditions C1–C2–C3 (Section 5). All rows involving a failure (exact, left-censored, interval-censored) carry a candidate set $c_i \subseteq \{1, \ldots, m\}$ and require C1–C2–C3. The right-censored row is a standard survival analysis result.

Observation type	**Likelihood contribution $L_i(\boldsymbol{\theta})$**
Exact failure + candidate set	$\prod_{l=1}^{m} R_l(t_i; \boldsymbol{\theta}_l) \cdot \sum_{j \in c_i} h_j(t_i; \boldsymbol{\theta}_j)$
Right-censored	$\prod_{j=1}^{m} R_j(\tau_i; \boldsymbol{\theta}_j)$
Left-censored + candidate set	$\int_0^{\tau_i} \sum_{j \in c_i} h_j(t; \boldsymbol{\theta}_j) \prod_{l=1}^{m} R_l(t; \boldsymbol{\theta}_l)\, dt$
Interval-censored + candidate set	$\int_{a_i}^{b_i} \sum_{j \in c_i} h_j(t; \boldsymbol{\theta}_j) \prod_{l=1}^{m} R_l(t; \boldsymbol{\theta}_l)\, dt$

When no cause information is available ($c_i = \{1, \ldots, m\}$), the left-censored and interval-censored contributions reduce to the familiar forms $1 - \prod_{j=1}^{m} R_j(\tau_i; \boldsymbol{\theta}_j)$ and $\prod_{j=1}^{m} R_j(a_i; \boldsymbol{\theta}_j) - \prod_{j=1}^{m} R_j(b_i; \boldsymbol{\theta}_j)$, respectively, since $\sum_{j=1}^{m} h_j(t; \boldsymbol{\theta}_j) \prod_{l=1}^{m} R_l(t; \boldsymbol{\theta}_l) = f_{T_i}(t; \boldsymbol{\theta})$ by Equation (10).

The right-censored contribution follows directly from the system reliability function (Theorem 1). The remaining three rows—all involving a confirmed failure—require the three conditions developed in Section 5.

4.4 Dependency Structure

Figure 1 depicts the dependency structure of the data generating process. Observed quantities $(S_i, \omega_i, \mathcal{C}_i, \boldsymbol{x}_i)$ are shown as shaded nodes; latent quantities (component lifetimes $T_{i1}, \ldots, T_{im}$, the system lifetime T_i, and the component cause of failure K_i) are shown as open nodes. Solid arrows denote structural dependencies; dashed arrows denote influences that the C1–C2–C3 conditions allow us to ignore in the likelihood. The covariate vector $\boldsymbol{x}_i$ influences both the component lifetimes (through the hazard functions $h_j(t; \boldsymbol{\theta}_j, \boldsymbol{x}_i)$) and potentially the candidate set $\mathcal{C}_i$ (e.g., when the operating environment affects diagnostic quality). The conditions in Section 5 allow us to construct a likelihood that does not require modeling the distribution of $\mathcal{C}_i$.

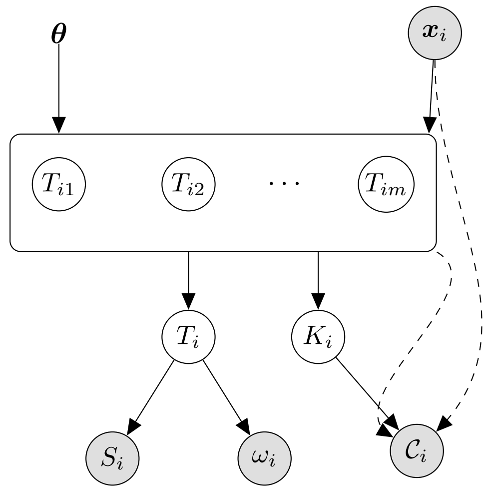

Figure 1: Dependency model for the data generating process. Shaded nodes are observed; open nodes are latent. Solid arrows are structural dependencies; dashed arrows indicate influences that are eliminated from the likelihood under C1–C2–C3.

4.5 Example: Masked Data

Table 3 shows an example of masked data for a 3-component series system with a right-censoring time $\tau = 5$. Systems 5 and 6 are right-censored (their failures were not observed before time τ). System 2 has a singleton candidate set, so its cause of failure is exactly identified. The remaining failed systems have candidate sets of size 2, representing partial masking.

Table 3: Example of right-censored lifetime data with masked component cause of failure for a 3-component series system ($\tau = 5$).

System	s_i	ω_i	c_i
1	1.1	E	$\{1, 2\}$
2	1.3	E	$\{2\}$
3	2.6	E	$\{2, 3\}$
4	3.7	E	$\{1, 3\}$
5	5.0	R	$\emptyset$
6	5.0	R	$\emptyset$

5 The C1–C2–C3 Likelihood

We now derive the likelihood contribution for an observation where the system failure time is known but the component cause of failure is masked by a candidate set. The derivation below recapitulates and extends the classical argument [Miyakawa, 1984, Usher and Hodgson, 1988] in a unified, general form.

5.1 Joint Distribution of T_i, K_i, and $\mathcal{C}_i$

Our goal is to estimate $\boldsymbol{\theta}$ from observed data $D_i = (s_i, \omega_i, c_i, \boldsymbol{x}_i)$. When the system failure is observed ($\omega_i = E$), we observe the system failure time t_i and a candidate set c_i. The joint pdf of the system lifetime T_i and the candidate set $\mathcal{C}_i$ is

$$f_{T_i,\mathcal{C}_i}(t_i, c_i; \boldsymbol{\theta}) = f_{T_i}(t_i; \boldsymbol{\theta}) \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i \mid T_i = t_i\}. \tag{14}$$

While we assume the system lifetime pdf $f_{T_i}(t_i; \boldsymbol{\theta})$ is known (up to parameters), the conditional distribution $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i \mid T_i = t_i\}$ is generally unknown—it depends on the diagnostic procedure, which we do not model.

Since $\mathcal{C}_i$ and K_i are statistically dependent, we can introduce K_i into the analysis. By Theorem 4, the joint pdf of T_i and K_i is $f_{T_i,K_i}(t_i, j; \boldsymbol{\theta}) = h_j(t_i; \boldsymbol{\theta}_j) \prod_{l=1}^{m} R_l(t_i; \boldsymbol{\theta}_l)$. The joint pdf of T_i, K_i, and $\mathcal{C}_i$ is therefore

$$f_{T_i,K_i,\mathcal{C}_i}(t_i, j, c_i; \boldsymbol{\theta}) = h_j(t_i; \boldsymbol{\theta}_j) \prod_{l=1}^{m} R_l(t_i; \boldsymbol{\theta}_l) \cdot \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i \mid T_i = t_i, K_i = j\}. \tag{15}$$

Marginalizing over K_i,

$$f_{T_i,\mathcal{C}_i}(t_i, c_i; \boldsymbol{\theta}) = \prod_{l=1}^{m} R_l(t_i; \boldsymbol{\theta}_l) \sum_{j=1}^{m} \left\{ h_j(t_i; \boldsymbol{\theta}_j) \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i \mid T_i = t_i, K_i = j\} \right\}. \tag{16}$$

The unknown conditional probability $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i \mid T_i = t_i, K_i = j\}$ prevents direct use of this expression for likelihood-based inference. We now introduce three conditions that successively simplify Equation (16) until the unknown masking distribution drops out entirely.

5.2 Condition 1: Candidate Set Contains the True Cause

Condition 1 (C1)**.** The candidate set $\mathcal{C}_i$ contains the index of the failed component:

$$\Pr_{\boldsymbol{\theta}}\{K_i \in \mathcal{C}_i\} = 1. \tag{17}$$

Condition 1 is the minimal requirement for the candidate set to carry useful information about the failure cause: the true cause must not be excluded. In practice, real diagnostics work by *narrowing down* from the full component set—eliminating candidates that pass functional checks—rather than constructing the candidate set from scratch. Because exclusion of the true cause would require the diagnostic to *affirmatively misidentify* a functioning component as the sole failure site, C1 holds whenever the diagnostic is competent in this limited sense.

Two common diagnostic architectures illustrate the point. First, automotive on-board diagnostics (OBD) fault codes are generated by the failing module itself: a voltage exceedance or communication timeout triggers the code, so the candidate set inherently includes the true cause. Second, hierarchical troubleshooting trees prune branches based on pass/fail tests at each level; the true cause remains in the surviving subtree at every step unless a test yields a false negative, which is a calibration failure rather than a structural feature of the diagnostic.

In short, C1 asks for diagnostic *competence*—not actively wrong—rather than diagnostic *precision*—exactly right. Violating C1 is a pathological scenario (the diagnostic positively excludes the failed component) that would undermine any analysis, masked or otherwise.

What C1 buys. Under C1, if $j \notin c_i$ then $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i \mid T_i = t_i, K_i = j\} = 0$. The summation in Equation (16) therefore reduces from $\{1, \ldots, m\}$ to c_i:

$$f_{T_i,\mathcal{C}_i}(t_i, c_i; \boldsymbol{\theta}) = \prod_{l=1}^{m} R_l(t_i; \boldsymbol{\theta}_l) \sum_{j \in c_i} \left\{ h_j(t_i; \boldsymbol{\theta}_j) \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i \mid T_i = t_i, K_i = j\} \right\}. \tag{18}$$

What breaks without C1. Without C1, the summation must range over all m components, and the likelihood depends on the masking probabilities for components *outside* the candidate set. This means the likelihood cannot be simplified without modeling the full masking mechanism.

5.3 Condition 2: Symmetric Masking Within the Candidate Set

Condition 2 (C2)**.** Given an observed system failure time $T_i = t_i$ and candidate set c_i, the masking probability is the same regardless of which component in c_i is the true cause:

$$\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i \mid T_i = t_i, K_i = j'\} = \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i \mid T_i = t_i, K_i = j\} \quad \text{for all } j, j' \in c_i. \tag{19}$$

Condition 2 is a requirement on the masking *mechanism*, not merely on the observed data: it must hold for all candidate sets c_i that the mechanism can produce, not just the particular sets realized in the sample. When $|c_i| = 1$ (a singleton candidate set), C2 is satisfied vacuously—the condition is non-trivial only when $|c_i| \geq 2$. In words, C2 requires that the diagnostic does not discriminate between components within the candidate set: given that a particular candidate set is reported, no member of that set is favored over another. Symmetry arises naturally whenever masking is determined by *structural grouping*—subsystem, module, or physical region—rather than by component-specific properties. Components that share a group are indistinguishable to the diagnostic precisely because the diagnostic operates at the group level.

Two additional examples reinforce this pattern. In avionics maintenance, field technicians replace *line-replaceable units* (LRUs): every component inside the unit is equally suspect because the diagnostic identified the unit, not the component. In industrial settings, SCADA monitoring systems report alarms at the subsystem level—e.g., "pump station fault"—without distinguishing which element (motor, valve, seal) triggered the alarm; the candidate set is the group, and all members are symmetric.

We acknowledge that C2 is the condition most likely to be violated in practice, since partial diagnostic information can make one candidate more plausible than another. When asymmetry is suspected, a practical mitigation is to redefine the candidate set at the finest resolution where symmetry still holds, effectively trading a smaller candidate set for a valid application of C2.

What C2 buys. Under C1 and C2, the masking probability $\Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i \mid T_i = t_i, K_i = j\}$ is constant for all $j \in c_i$ and can be factored out of the summation in Equation (18):

$$f_{T_i,\mathcal{C}_i}(t_i, c_i; \boldsymbol{\theta}) = \Pr_{\boldsymbol{\theta}}\{\mathcal{C}_i = c_i \mid T_i = t_i, K_i = j'\} \prod_{l=1}^{m} R_l(t_i; \boldsymbol{\theta}_l) \sum_{j \in c_i} h_j(t_i; \boldsymbol{\theta}_j), \tag{20}$$

where j' is any element of c_i.

What breaks without C2. Without C2, the masking probabilities remain inside the summation, coupling the hazard contributions with component-specific masking weights. The MLE then depends on the unknown masking probabilities, which must be jointly estimated or modeled.

5.4 Condition 3: Masking Independent of $\boldsymbol{\theta}$

Condition 3 (C3)**.** The masking probabilities, conditioned on the failure time and the component cause of failure, do not depend on the system parameter $\boldsymbol{\theta}$:

$$\beta_i := \Pr\{\mathcal{C}_i = c_i \mid T_i = t_i, K_i = j'\} \tag{21}$$

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

Condition 3 states that the diagnostic procedure's behavior is determined by factors external to the component lifetime parameters. The masking probability β_i may depend on the failure time (for exact failures, where the diagnostic is performed at or near the observed failure time), the diagnostician, the testing equipment, or other covariates—but not on $\boldsymbol{\theta}$. For censored observations, the diagnostic is performed at the inspection time rather than at the unknown failure time; see Remark 3 below. The justification is fundamentally causal: the diagnostic tool was designed and calibrated *before* any failures were observed, so its behavior cannot depend on the unknown $\boldsymbol{\theta}$ we are estimating. For instance, OBD voltage thresholds are hard-coded at manufacture; a vibration sensor's frequency band is set during installation. Neither adapts to the lifetime parameters of the components it monitors.

Condition 3 is closely related to the concept of *ignorability* in the missing-data framework of Little and Rubin [2002]. In our setting, the "missingness" is the loss of exact cause information through masking; C3 ensures that the masking mechanism is *ignorable* for likelihood-based inference, in the sense that the conditional distribution of the candidate set need not be modeled when maximizing the likelihood over $\boldsymbol{\theta}$ [see Little and Rubin, 2002, Ch. 6].

When covariates $\boldsymbol{x}_i$ are present, C3 requires that the masking probabilities do not depend on $\boldsymbol{\theta}$ *given* the covariates and failure time. If the same covariates influence both the failure rate and the diagnostic quality (e.g., an extreme operating environment that accelerates failures and also degrades sensor accuracy), the practitioner should verify that the masking mechanism remains ignorable after conditioning on $\boldsymbol{x}_i$.

What C3 buys. Under C1, C2, and C3, the joint pdf becomes

$$f_{T_i,\mathcal{C}_i}(t_i, c_i; \boldsymbol{\theta}) = \beta_i \prod_{l=1}^{m} R_l(t_i; \boldsymbol{\theta}_l) \sum_{j \in c_i} h_j(t_i; \boldsymbol{\theta}_j). \tag{22}$$

When we view this as a function of $\boldsymbol{\theta}$ (with the data fixed), β_i is a constant multiplier that does not affect the location of the maximum. The likelihood contribution is therefore

$$L_i(\boldsymbol{\theta}) \propto \prod_{l=1}^{m} R_l(t_i; \boldsymbol{\theta}_l) \sum_{j \in c_i} h_j(t_i; \boldsymbol{\theta}_j). \tag{23}$$

What breaks without C3. Without C3, the factor β_i depends on $\boldsymbol{\theta}$ and cannot be dropped. The practitioner would need to model the dependence of the masking mechanism on $\boldsymbol{\theta}$—a substantially harder problem that requires additional data or assumptions about the diagnostic process.

5.5 Real-World Example

To illustrate how the three conditions arise in practice, consider an electronic device with three components arranged in a series configuration. Components 1 and 2 are on a shared circuit board, while component 3 is separate. A diagnostic tool isolates the failure to either the shared circuit board or the individual component. A more detailed board-level inspection sometimes pinpoints the specific failed component; let $p \in (0, 1)$ denote the probability that this inspection succeeds. The conditional probabilities for candidate sets are:

$$\Pr\{\mathcal{C}_i = c_i \mid T_i = t_i, K_i = j\} = \begin{cases} p & \text{if } c_i = \{j\} \text{ and } j \in \{1, 2\}, \\ 1 - p & \text{if } c_i = \{1, 2\} \text{ and } j \in \{1, 2\}, \\ 1 & \text{if } c_i = \{3\} \text{ and } j = 3, \\ 0 & \text{otherwise.} \end{cases}$$

This diagnostic tool satisfies all three conditions:

- **C1**: The candidate set always contains the failed component—the tool correctly isolates failures to the correct subsystem (each possible candidate set contains the true cause).
- **C2**: For the candidate set $\{1, 2\}$, the masking probability is the same whether component 1 or component 2 failed (both equal $1-p$). For singleton candidate sets $\{1\}$ or $\{2\}$, Condition 2 is satisfied trivially.
- **C3**: The masking probabilities depend only on the diagnostic tool (through p) and not on the component lifetime parameters $\boldsymbol{\theta}$.

The parameter $p > 0$ is essential for identifiability: when the board-level inspection never succeeds ($p = 0$), components 1 and 2 always appear together in every candidate set and their individual parameters cannot be separated—see Section 5.8 for a detailed discussion. Any $p > 0$ ensures that some observations produce singleton candidate sets $\{1\}$ or $\{2\}$, providing the information needed to distinguish the two components.

According to Guess et al. [1991], many industrial diagnostic scenarios naturally satisfy these conditions, reinforcing the practical applicability of the framework.

5.6 Censored Observations with Candidate Sets

The C1–C2–C3 derivation in Sections 5.2–5.4 applies to any observation where a failure is known to have occurred. Equation (22) gives the joint density of the failure time and candidate set at a *specific* time t_i. For left-censored and interval-censored observations, the failure time is not known exactly, so we integrate over the admissible range.

Remark 3 (Diagnostic timing and the masking probability)**.** For an exact failure, the diagnostic may be performed at or near the observed failure time t_i, so $\beta_i = \beta(t_i)$ is evaluated at a known point and is a scalar constant in the likelihood. For censored observations, the diagnostic is performed at the inspection time—when the system's failed state is discovered—not at the unknown failure time T_i. (For interval-censored observations, this is typically the later inspection time b_i, at which the failure is first detected.) The masking probability is therefore $\beta_i = \Pr\{\mathcal{C}_i = c_i \mid \text{system failed by inspection}, K_i = j'\}$, which does not depend on the integration variable t. This is the natural model: inspectors diagnose the current state at inspection time and do not have access to the unknown failure time. Consequently, β_i factors out of the integrals in the left-censored and interval-censored likelihood contributions below, and by Condition 3 it does not depend on $\boldsymbol{\theta}$, so it may be dropped.

Theorem 7 (Left-censored likelihood contribution under C1–C2–C3)**.** *Under Conditions 1–3, if the ith system is known to have failed by time τ_i with candidate set c_i, the likelihood contribution is*

$$L_i(\boldsymbol{\theta}) \propto \int_0^{\tau_i} \sum_{j \in c_i} h_j(t; \boldsymbol{\theta}_j) \prod_{l=1}^{m} R_l(t; \boldsymbol{\theta}_l)\, dt. \tag{24}$$

Proof. By Equation (22), the joint density of the failure time and candidate set at time t is $\beta_i \sum_{j \in c_i} h_j(t; \boldsymbol{\theta}_j) \prod_{l=1}^{m} R_l(t; \boldsymbol{\theta}_l)$. Since the failure time is known only to satisfy $T_i \leq \tau_i$, we integrate over $t \in (0, \tau_i]$. As noted in Remark 3, the diagnostic is performed at inspection time, so β_i does not depend on the integration variable t and factors out of the integral. By Condition 3, β_i does not depend on $\boldsymbol{\theta}$ either, so it is a constant factor that may be dropped. □

Theorem 8 (Interval-censored likelihood contribution under C1–C2–C3)**.** *Under Conditions 1–3, if the ith system failed in the interval $(a_i, b_i]$ with candidate set c_i, the likelihood contribution is*

$$L_i(\boldsymbol{\theta}) \propto \int_{a_i}^{b_i} \sum_{j \in c_i} h_j(t; \boldsymbol{\theta}_j) \prod_{l=1}^{m} R_l(t; \boldsymbol{\theta}_l)\, dt. \tag{25}$$

Proof. The argument is identical to the proof of Theorem 7—with β_i again constant by the diagnostic-timing reasoning of Remark 3—but with the integration domain restricted to $(a_i, b_i]$. □

Remark 4. When no cause information is available ($c_i = \{1, \ldots, m\}$), these contributions reduce to the standard censored-data forms. For the left-censored case,

$$\int_0^{\tau_i} \sum_{j=1}^{m} h_j(t; \boldsymbol{\theta}_j) \prod_{l=1}^{m} R_l(t; \boldsymbol{\theta}_l)\, dt = \int_0^{\tau_i} f_{T_i}(t; \boldsymbol{\theta})\, dt = 1 - \prod_{j=1}^{m} R_j(\tau_i; \boldsymbol{\theta}_j),$$

by Equation (10). Similarly, the interval-censored contribution reduces to

$$\prod_{j=1}^{m} R_j(a_i; \boldsymbol{\theta}_j) - \prod_{j=1}^{m} R_j(b_i; \boldsymbol{\theta}_j).$$

5.7 Combined Likelihood

We now assemble the full likelihood from the individual contributions (Equation (23), Theorems 7–8, and the right-censored case from Theorem 1). Let $\mathcal{E} = \{i : \omega_i = E\}$, $\mathcal{R} = \{i : \omega_i = R\}$, $\mathcal{L} = \{i : \omega_i = L\}$, and $\mathcal{I} = \{i : \omega_i = I\}$ denote the index sets of observations that are exact failures, right-censored, left-censored, and interval-censored, respectively.

Theorem 9 (Likelihood under C1–C2–C3)**.** *Under Conditions 1–3, the likelihood for the observed data $D = \{D_1, \ldots, D_n\}$ is*

$$\begin{aligned} L(\boldsymbol{\theta}) \;\propto\; & \prod_{i \in \mathcal{E}} \left[\prod_{l=1}^{m} R_l(t_i; \boldsymbol{\theta}_l) \cdot \sum_{j \in c_i} h_j(t_i; \boldsymbol{\theta}_j) \right] \prod_{i \in \mathcal{R}} \left[\prod_{j=1}^{m} R_j(\tau_i; \boldsymbol{\theta}_j) \right] \\ & \times \prod_{i \in \mathcal{L}} \left[\int_0^{\tau_i} \sum_{j \in c_i} h_j(t; \boldsymbol{\theta}_j) \prod_{l=1}^{m} R_l(t; \boldsymbol{\theta}_l)\, dt \right] \prod_{i \in \mathcal{I}} \left[\int_{a_i}^{b_i} \sum_{j \in c_i} h_j(t; \boldsymbol{\theta}_j) \prod_{l=1}^{m} R_l(t; \boldsymbol{\theta}_l)\, dt \right]. \end{aligned} \tag{26}$$

Proof. Each factor follows from the corresponding individual result: the exact-failure contribution from Equation (23), the right-censored contribution from Theorem 1, the left-censored contribution from Theorem 7, and the interval-censored contribution from Theorem 8. Independence across systems gives the product. □

5.8 Identifiability

Before addressing identifiability within our framework, we note the classical competing risks result of Tsiatis [1975]: without the component independence assumption (Remark 1), the marginal component lifetime distributions are not identifiable from system lifetime data alone—even with complete observation of the failure cause. Our framework inherits identifiability from the independence assumption, which allows the system reliability to factor as a product of component reliabilities (Theorem 1). Given independence, the identifiability question reduces to whether the observed candidate sets provide enough information to separate the component parameters.

Definition 1 (Candidate-set separability)**.** Let $\mathcal{S}$ denote the collection of candidate sets that occur with positive probability under the masking mechanism. We say the masking mechanism is *separating* if, for every pair of distinct components $j \neq j'$, there exists a candidate set $c \in \mathcal{S}$ such that $j \in c$ and $j' \notin c$. When this condition fails for some pair (j, j')—that is, $j \in c \iff j' \in c$ for every $c \in \mathcal{S}$—we say j and j' are *diagnostically confounded.*

The following theorem shows that separability is the key condition governing identifiability under C1–C2–C3.

Theorem 10 (Identifiability under C1–C2–C3)**.** *Let $h_j(t; \boldsymbol{\theta}_j)$ be the hazard function for component j, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m) \in \Theta \subseteq \mathbb{R}^p$ is the full parameter vector, and suppose each component family is individually identifiable: $h_j(\cdot\,; \boldsymbol{\theta}_j) = h_j(\cdot\,; \boldsymbol{\theta}'_j)$ for all $t > 0$ implies $\boldsymbol{\theta}_j = \boldsymbol{\theta}'_j$. Under C1–C2–C3:*

(a) ***Necessary condition.*** *If components j and j' are diagnostically confounded (Definition 1), then $\boldsymbol{\theta}_j$ and $\boldsymbol{\theta}_{j'}$ are not separately identifiable from the observed-data likelihood. Specifically, any reparametrization preserving $h_j(t; \boldsymbol{\theta}_j) + h_{j'}(t; \boldsymbol{\theta}_{j'})$ for all $t > 0$ yields the same likelihood value.*

(b) ***Sufficient condition.*** *If the masking mechanism is separating (Definition 1), and at least some observations are exact failures (type E), then the parameter vector $\boldsymbol{\theta}$ is identifiable for any parametric family whose hazard functions $\{h_1, \ldots, h_m\}$ are linearly independent over $(0, \infty)$.*

Proof. Part (a). Suppose $j \in c \iff j' \in c$ for every $c \in \mathcal{S}$. Then in every exact-failure likelihood contribution $R(t_i) \sum_{l \in c_i} h_l(t_i)$ (Equation 23), the hazards h_j and $h_{j'}$ appear as a sum $h_j(t_i) + h_{j'}(t_i)$ wherever they appear at all. The survival factor $R(t_i) = \exp\big(-\sum_l H_l(t_i)\big)$ likewise depends on $H_j(t_i) + H_{j'}(t_i)$ only through their sum. The same holds for right-, left-, and interval-censored contributions (Table 2), since all involve $R(t)$ and $\sum_{l \in c} h_l(t)$. Hence the likelihood $L(\boldsymbol{\theta})$ depends on $\boldsymbol{\theta}_j$ and $\boldsymbol{\theta}_{j'}$ only through the sum $h_j + h_{j'}$, and any reallocation preserving this sum leaves L unchanged.

Part (b). Suppose the masking mechanism is separating and let $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$ with $L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}')$ for all data sets. Then in particular, for a data set consisting of a single exact failure at time t with candidate set c, the log-likelihood equality gives

$$-\sum_l H_l(t) + \log \sum_{j \in c} h_j(t) = -\sum_l H'_l(t) + \log \sum_{j \in c} h'_j(t) \tag{27}$$

for all $t > 0$ and all $c \in \mathcal{S}$, where $h_j = h_j(\cdot\,; \boldsymbol{\theta}_j)$ and $h'_j = h_j(\cdot\,; \boldsymbol{\theta}'_j)$. Differentiating (27) with respect to t for two candidate sets $c_1, c_2 \in \mathcal{S}$ with $j \in c_1$, $j \notin c_2$ (which exist by separability) and subtracting yields a relation that isolates the contribution of component j. Specifically, from the survival terms we obtain $\sum_l h_l(t) = \sum_l h'_l(t)$, and from candidate sets separating j we obtain that $h_j(t) = h'_j(t)$ for all $t > 0$ whenever $h_1, \ldots, h_m$ are linearly independent. By the individual identifiability assumption, $\boldsymbol{\theta}_j = \boldsymbol{\theta}'_j$ for each j. □

Part (a) states that diagnostically confounded components are fundamentally unresolvable: the likelihood surface has a ridge along all reparametrizations of the confounded pair, producing an infinite family of equivalent maximizers rather than a unique MLE. Part (b) provides a checkable condition: the analyst needs only inspect the candidate-set structure $\mathcal{S}$ to verify separability.

Remark 5 (Linear independence of hazard functions)**.** The linear independence condition in Theorem 10(b) holds generically for all families in Table 4. The Weibull hazard $h_j(t) \propto t^{k_j - 1}$ with distinct shapes k_j yields linearly independent power functions. For the exponential family ($k_j = 1$ for all j), the hazards are constant and hence linearly *dependent*; in this case, separability of the candidate sets carries the full burden of identifiability. For families with identical functional form across components (e.g., exponential or homogeneous Weibull), identifiability requires that the candidate-set matrix $C \in \{0, 1\}^{|\mathcal{S}| \times m}$—whose rows are the indicator vectors of the candidate sets in $\mathcal{S}$—has full column rank m. This is strictly stronger than separability (which requires only that no two columns of C are identical) but is guaranteed when $\mathcal{S}$ includes a singleton $\{j\}$ for each component j.

Remediation of non-identifiability. When components j and j' are diagnostically confounded, three strategies are available:

1. **Collapse into a super-component.** Replace the always-grouped components with a single composite component whose hazard is $h_j + h_{j'}$. This reduces the parameter count and restores identifiability at the cost of losing individual component resolution.

2. **Impose equality constraints.** Assume the grouped components share identical parameter values ($\boldsymbol{\theta}_j = \boldsymbol{\theta}_{j'}$), splitting the combined hazard equally. This is appropriate when the components are physically interchangeable (e.g., identical capacitors on the same circuit board).

3. **Bayesian regularization.** Place informative priors on $\boldsymbol{\theta}_j$ and $\boldsymbol{\theta}_{j'}$ to obtain a proper posterior even when the likelihood surface is flat. The posterior concentrates as additional diagnostic information becomes available, so the prior penalty vanishes asymptotically.

In each case, the practitioner should examine the candidate-set structure of the data before fitting the model. Convergence difficulties in numerical optimization (e.g., failure to converge within a maximum number of iterations) may signal that the separability condition of Definition 1 is violated or nearly violated.

6 Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) finds the parameter values that maximize the likelihood of the observed data [Bain and Engelhardt, 1992, Casella and Berger, 2002]. A maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ satisfies

$$L(\hat{\boldsymbol{\theta}}) = \max_{\boldsymbol{\theta} \in \boldsymbol{\Omega}} L(\boldsymbol{\theta}). \tag{28}$$

For computational efficiency, we work with the log-likelihood.

Theorem 11 (Log-likelihood). *Under Conditions 1–3, the log-likelihood for the masked data model is*

$$\ell(\boldsymbol{\theta}) = \ell_E(\boldsymbol{\theta}) + \ell_R(\boldsymbol{\theta}) + \ell_L(\boldsymbol{\theta}) + \ell_I(\boldsymbol{\theta}), \tag{29}$$

where

$$\ell_E(\boldsymbol{\theta}) = \sum_{i \in \mathcal{E}} \left[\sum_{j=1}^{m} \log R_j(t_i; \boldsymbol{\theta}_j) + \log\left(\sum_{j \in c_i} h_j(t_i; \boldsymbol{\theta}_j) \right) \right], \tag{30}$$

$$\ell_R(\boldsymbol{\theta}) = \sum_{i \in \mathcal{R}} \sum_{j=1}^{m} \log R_j(\tau_i; \boldsymbol{\theta}_j), \tag{31}$$

$$\ell_L(\boldsymbol{\theta}) = \sum_{i \in \mathcal{L}} \log\left[\int_0^{\tau_i} \sum_{j \in c_i} h_j(t; \boldsymbol{\theta}_j) \prod_{l=1}^{m} R_l(t; \boldsymbol{\theta}_l)\, dt \right], \tag{32}$$

$$\ell_I(\boldsymbol{\theta}) = \sum_{i \in \mathcal{I}} \log\left[\int_{a_i}^{b_i} \sum_{j \in c_i} h_j(t; \boldsymbol{\theta}_j) \prod_{l=1}^{m} R_l(t; \boldsymbol{\theta}_l)\, dt \right]. \tag{33}$$

Proof. Taking logarithms in Theorem 9 and using $\log \prod = \sum \log$ gives the result directly. □

Remark 6 (Common special case). When only exact failures and right-censored observations are present ($\mathcal{L} = \mathcal{I} = \emptyset$), the log-likelihood reduces to $\ell(\boldsymbol{\theta}) = \ell_E(\boldsymbol{\theta}) + \ell_R(\boldsymbol{\theta})$ (Equations (30) and (31)). This is the form used in most companion papers.

6.1 Score Equations

The MLE is found by solving the score equations

$$\frac{\partial}{\partial \theta_{j,r}} \ell(\boldsymbol{\theta}) = 0, \tag{34}$$

for each parameter $\theta_{j,r}$ (the rth element of the jth component's parameter vector). In general, these equations do not admit closed-form solutions and must be solved numerically using methods such as Newton–Raphson or quasi-Newton algorithms [Nocedal and Wright, 2006, Byrd et al., 1995].

6.2 Asymptotic Properties

Under standard regularity conditions (which must be verified for each specific distribution family)—including identifiability, smoothness of the log-likelihood, and the true parameter lying in the interior of the parameter space—the MLE is consistent, asymptotically normal, and asymptotically efficient [Casella and Berger, 2002, Lehmann and Casella, 1998]. That is, as $n \to \infty$, $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$ (the true parameter) and $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\theta}_0))$, where $\mathcal{I}(\boldsymbol{\theta}_0)$ is the per-observation Fisher information matrix. For finite samples, these asymptotic approximations may be inaccurate, and bootstrap methods (e.g., the bias-corrected and accelerated method [Efron, 1987, Efron and Tibshirani, 1994]) provide a nonparametric alternative for constructing confidence intervals.

6.3 General Recipe for Practitioners

Given a hazard function $h_j(t; \boldsymbol{\theta}_j, \boldsymbol{x}_i)$ for each component, the practitioner applies the framework as follows:

1. **Specify** the component hazard functions $h_j(t; \boldsymbol{\theta}_j, \boldsymbol{x}_i)$ and compute $R_j = \exp\bigl(-\int_0^t h_j \, du\bigr)$ (analytically if possible, numerically otherwise).

2. **Substitute** the component-specific R_j and h_j into the log-likelihood (Equation (29)).

3. **Differentiate** the log-likelihood with respect to each parameter to obtain the score equations.

4. **Solve** the score equations numerically (e.g., using L-BFGS-B [Byrd et al., 1995] or Newton–Raphson) to obtain $\hat{\boldsymbol{\theta}}$.

5. **Construct confidence intervals** using the observed Fisher information or bootstrap resampling.

Section 7 provides hazard functions for five common parametric families, enabling immediate application of this recipe.

6.4 Worked Example: Exponential Components

We illustrate the recipe using the data in Table 3 ($m = 3$ components, $n = 6$ observations) with Exponential component lifetimes. For Exponential components, $R_j(t; \lambda_j) = e^{-\lambda_j t}$ and $h_j(t; \lambda_j) = \lambda_j$, where $\lambda_j > 0$ is the failure rate of the jth component.

Since the data contain only exact failures and right-censored observations, the log-likelihood is $\ell = \ell_E + \ell_R$ (Remark 6). For Exponential components, $\log R_j(t; \lambda_j) = -\lambda_j t$ and $h_j(t; \lambda_j) = \lambda_j$, so the two contributions are:

$$\ell_E = \sum_{i \in \mathcal{E}} \left[-(\lambda_1 + \lambda_2 + \lambda_3)\, t_i + \log\left(\sum_{j \in c_i} \lambda_j \right) \right], \tag{35}$$

$$\ell_R = \sum_{i \in \mathcal{R}} \bigl[-(\lambda_1 + \lambda_2 + \lambda_3)\, \tau_i \bigr]. \tag{36}$$

Substituting the data from Table 3 (with $T_{\text{total}} = \sum_{i=1}^{n} s_i = 18.7$), the combined log-likelihood is

$$\ell = -18.7\,(\lambda_1 + \lambda_2 + \lambda_3) + \log(\lambda_1 + \lambda_2) + \log(\lambda_2) + \log(\lambda_2 + \lambda_3) + \log(\lambda_1 + \lambda_3). \tag{37}$$

The score equations $\partial \ell / \partial \lambda_j = 0$ are

$$\frac{\partial \ell}{\partial \lambda_1} = -18.7 + \frac{1}{\lambda_1 + \lambda_2} + \frac{1}{\lambda_1 + \lambda_3} = 0, \tag{38}$$

$$\frac{\partial \ell}{\partial \lambda_2} = -18.7 + \frac{1}{\lambda_1 + \lambda_2} + \frac{1}{\lambda_2} + \frac{1}{\lambda_2 + \lambda_3} = 0, \tag{39}$$

$$\frac{\partial \ell}{\partial \lambda_3} = -18.7 + \frac{1}{\lambda_2 + \lambda_3} + \frac{1}{\lambda_1 + \lambda_3} = 0. \tag{40}$$

Subtracting (40) from (38) gives $1/(\lambda_1 + \lambda_2) = 1/(\lambda_2 + \lambda_3)$, so $\hat{\lambda}_1 = \hat{\lambda}_3$. Setting $\alpha = \lambda_1 = \lambda_3$ and $\beta = \lambda_2$, the system reduces to two equations whose solution is

$$\hat{\lambda}_1 = \hat{\lambda}_3 = \frac{7 - \sqrt{17}}{74.8} \approx 0.0385, \qquad \hat{\lambda}_2 = \frac{1 + \sqrt{17}}{37.4} \approx 0.1370. \tag{41}$$

The total system failure rate is $\hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_3 = 4/18.7 \approx 0.2139$, which equals the number of observed failures divided by the total exposure time—a general property of the Exponential MLE. The masking affects only the *allocation* of hazard across components, not the total.

Component 2 has the highest estimated failure rate, consistent with its appearance in three of the four failure candidate sets (including a singleton). Components 1 and 3 have equal rates by symmetry: each appears in exactly two candidate sets, with identical structure after relabeling.

7 Common Hazard Function Specifications

Table 4 lists hazard functions for five standard parametric families, illustrating how the general framework specializes. Any hazard function $h_j(t; \boldsymbol{\theta}_j)$ satisfying the regularity conditions of Section 6.2 can be used; the named families below are common starting points.

Table 4: Common hazard function specifications. The hazard function h_j is the primary specification; the reliability function R_j and density f_j are derived via Equations (4)–(5).

Family	$h_j(t; \boldsymbol{\theta}_j)$	$R_j(t; \boldsymbol{\theta}_j)$	$f_j(t; \boldsymbol{\theta}_j)$	**Parameters**
Exponential	λ_j	$e^{-\lambda_j t}$	$\lambda_j e^{-\lambda_j t}$	$\lambda_j > 0$
Weibull	$\frac{k_j}{\lambda_j}\left(\frac{t}{\lambda_j}\right)^{k_j-1}$	$e^{-(t/\lambda_j)^{k_j}}$	$\frac{k_j}{\lambda_j}\left(\frac{t}{\lambda_j}\right)^{k_j-1} e^{-(t/\lambda_j)^{k_j}}$	$k_j, \lambda_j > 0$
Pareto	$\frac{\alpha_j}{t}$	$\left(\frac{x_{\min,j}}{t}\right)^{\alpha_j}$	$\frac{\alpha_j\, x_{\min,j}^{\alpha_j}}{t^{\alpha_j+1}}$	$\alpha_j, x_{\min,j} > 0$; $t \geq x_{\min,j}$
Log-normal	$\frac{\phi\left(\frac{\log t - \mu_j}{\sigma_j}\right)}{\sigma_j\, t\left[1 - \Phi\left(\frac{\log t - \mu_j}{\sigma_j}\right)\right]}$	$1 - \Phi\left(\frac{\log t - \mu_j}{\sigma_j}\right)$	$\frac{1}{\sigma_j t}\, \phi\left(\frac{\log t - \mu_j}{\sigma_j}\right)$	$\mu_j \in \mathbb{R}$, $\sigma_j > 0$
Gamma[1]	$\frac{f_j(t; \boldsymbol{\theta}_j)}{R_j(t; \boldsymbol{\theta}_j)}$	$1 - \frac{\gamma(\alpha_j, \beta_j t)}{\Gamma(\alpha_j)}$	$\frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)}\, t^{\alpha_j - 1} e^{-\beta_j t}$	$\alpha_j, \beta_j > 0$

In Table 4, $\Phi(\cdot)$ and $\phi(\cdot)$ denote the standard normal CDF and pdf, respectively; $\gamma(\cdot, \cdot)$ is the lower incomplete gamma function; and $\Gamma(\cdot)$ is the gamma function. For distributions with parameter-dependent support (e.g., the Pareto distribution with $t \geq x_{\min,j}$), the system support is $t \geq \max_j x_{\min,j}$, and the product formula for system reliability (Theorem 1) must be applied only on this common support.

Remark 7 (Covariate-dependent hazards)**.** The hazard function $h_j(t; \boldsymbol{\theta}_j, \boldsymbol{x}_i)$ may incorporate observation-level covariates $\boldsymbol{x}_i$. A common special case is the proportional hazards specification $h_j(t; \boldsymbol{\theta}_j, \boldsymbol{x}_i) = h_{0,j}(t; \boldsymbol{\theta}_j) \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}_j)$, where $h_{0,j}$ is a baseline hazard from any family in Table 4 and $\boldsymbol{\beta}_j$ is a component-specific regression coefficient vector. The likelihood expressions of Section 5 remain valid; only the functional form of h_j changes.

Remark 8 (Nested models within a family)**.** Several families in Table 4 admit natural hierarchies of nested sub-models. For instance, the Weibull family contains a common-shape reduction ($m+1$ parameters) in which the system lifetime is itself Weibull, and a further exponential specialization (m parameters) with fully analytical inference. These nestings enable formal model selection via likelihood ratio tests, AIC, or BIC. A detailed treatment—including simulation studies quantifying when a reduced model is appropriate—is given in Towell [2025e].

8 Discussion

8.1 What the Framework Enables

The hazard-function-based likelihood framework developed in this paper provides a foundation for a family of distribution-specific companion papers. Each companion paper can focus on a specific parametric family—for example, Weibull [Towell, 2025f] or Exponential [Towell, 2025d]—deriving closed-form score equations and Fisher information matrices, conducting simulation studies to assess

[1] The Gamma hazard function has no elementary closed form; it is expressed here via its definition $h_j = f_j / R_j$.

MLE performance under varying masking and censoring scenarios, and developing specialized software packages, all while citing the present work for the general theory. The layered software stack described in Section 8.4—from component specification (`dfr.dist`) through series composition (`dfr.dist.series`) to masked-data likelihood (`dfr.lik.series.md`)—demonstrates this modularity: new component distributions inherit the full estimation pipeline without distribution-specific likelihood code.

8.2 Relaxation of Conditions

The three conditions (C1–C2–C3) provide a clean mathematical framework, but practitioners may encounter situations where one or more conditions are violated. We briefly sketch what happens in each case:

- **Relaxing C1** (candidate set may not contain the true cause): The summation in the likelihood cannot be restricted to c_i; the full component set must be considered, along with a model for the probability that the true cause is excluded from the candidate set. This introduces additional nuisance parameters.
- **Relaxing C2** (asymmetric masking): The masking probabilities $\Pr\{\mathcal{C}_i = c_i \mid T_i = t_i, K_i = j\}$ remain inside the summation and couple with the component hazards. The MLE depends on the relative masking probabilities, which must be estimated or modeled.
- **Relaxing C3** (masking depends on $\boldsymbol{\theta}$): The factor β_i cannot be dropped from the likelihood, and the MLE must account for the dependence of the masking mechanism on the lifetime parameters. This leads to a more complex joint estimation problem.

Detailed treatment of these relaxations is beyond the scope of this paper and is deferred to future work.

8.3 Computational Considerations

Several practical issues arise when applying the framework:

- **Local optima.** The log-likelihood surface may be multimodal, particularly under heavy masking or censoring. Multiple random starting points for the numerical optimizer are recommended.
- **Convergence.** Failure to converge within a reasonable number of iterations may indicate identifiability issues for the given data set. Such cases should be flagged rather than silently discarded.
- **Boundary constraints.** Many lifetime distributions have positivity constraints on parameters. Constrained optimization methods such as L-BFGS-B [Byrd et al., 1995] or reparameterization (e.g., optimizing over log-scale parameters) can enforce these constraints.

8.4 Software Ecosystem

The framework developed in this paper is implemented as a layered software stack in the R statistical computing environment [Towell, 2025a,b,c]. At the base, the `dfr.dist` package [Towell, 2025a] provides a hazard-function-first abstraction for individual component lifetime distributions: each distribution is specified by its hazard function, from which the cumulative hazard, reliability, density, and random sampling follow (cf. Equations (3)–(5)). The `dfr.dist.series` package [Towell, 2025b] composes components into series system distributions using hazard additivity (Theorem 3), with a parameter layout that maps the global vector $\boldsymbol{\theta}$ to component-specific subvectors $\boldsymbol{\theta}_j$. The `dfr.lik.series.md` package [Towell, 2025c] implements the C1–C2–C3 log-likelihood (Theorem 11) for all four observation types (exact, right-censored, left-censored, and interval-censored), taking arbitrary component hazard closures as input. Together these packages enable practitioners to apply the recipe of Section 6.3 with any component distribution in the `dfr.dist` ecosystem, without writing distribution-specific likelihood code.

For Weibull component lifetimes specifically, the `wei.series.md.c1.c2.c3` package [Towell, 2023c] provides closed-form score vectors and Fisher information, along with bootstrap confidence interval construction [Towell, 2023a].

9 Conclusion

We have developed a general likelihood framework, expressed in terms of component hazard functions, for estimating component reliability from masked series system data. The framework rests on three conditions (C1–C2–C3) that capture natural properties of diagnostic procedures and allow the unknown masking distribution to be eliminated from the likelihood.

The key results are:

- The joint distribution of the system lifetime, component cause of failure, and candidate set (Section 5.1);
- The derivation showing how each condition progressively simplifies the likelihood (Sections 5.2 through 5.4);
- The combined likelihood contribution under C1–C2–C3 (Theorem 9);
- The general log-likelihood and a recipe for applying the framework to any parametric hazard specification (Section 6);
- Hazard function specifications for five common families (Table 4).

This framework provides a rigorous foundation for distribution-specific companion papers that can focus on deriving score equations, conducting simulation studies, and developing specialized inference tools for particular lifetime distribution families.

References

Marcus Agustin. Systems in series. In *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Ltd, 2011. ISBN 9780470400531. doi: https://doi.org/10.1002/9780470400531.eorms0866. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470400531.eorms0866.

L.J. Bain and M. Engelhardt. *Introduction to Probability and Mathematical Statistics*. Duxbury Press, second edition, 1992. ISBN 9780534380205.

Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

George Casella and Roger L Berger. *Statistical Inference*. Duxbury Advanced Series, 2002.

B. Efron. Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397):171–185, 1987.

Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

Frank M. Guess, Thom J. Hodgson, and John S. Usher. Estimating system and component reliabilities under partial information on cause of failure. *Journal of Statistical Planning and Inference*, 29:75–85, sep 1991. doi: 10.1016/0378-3758(92)90123-a.

John P Klein and Melvin L Moeschberger. *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2005.

Erich L. Lehmann and George Casella. *Theory of Point Estimation*. Springer Science & Business Media, 1998.

Roderick J. A. Little and Donald B. Rubin. *Statistical Analysis with Missing Data*. Wiley-Interscience, Hoboken, NJ, second edition, 2002.

Masayuki Miyakawa. Analysis of incomplete data in competing risks model. *IEEE Transactions on Reliability*, R-33(4):293–296, 1984.

J. Nocedal and S. Wright. *Numerical Optimization*. Springer, 2006.

Ammar M. Sarhan. Reliability estimations of components from masked system life data. *Reliability Engineering & System Safety*, 74(1):107–113, October 2001. doi: 10.1016/s0951-8320(01)00072-2.

Alex Towell. *Algebraic Maximum Likelihood Estimators*, 2023a. URL https://queelius.github.io/algebraic.mle/. R package version 0.9.0. Available: https://github.com/queelius/algebraic.mle/.

Alexander Towell. Reliability estimation in series systems: Maximum likelihood techniques for right-censored and masked failure data. Master's project, Southern Illinois University Edwardsville, 2023b. URL https://zenodo.org/records/18615871.

Alexander Towell. *wei.series.md.c1.c2.c3: Estimating Reliability of Weibull Components in Series from Masked Data*, 2023c. URL https://zenodo.org/records/18615922. R package version 0.9.0.

Alexander Towell. *dfr.dist: Distribution-Free Reliability Distributions*, 2025a. URL https://github.com/queelius/dfr.dist. R package. Available: https://github.com/queelius/dfr.dist.

Alexander Towell. *dfr.dist.series: Series System Distributions from Dynamic Failure Rate Components*, 2025b. URL https://github.com/queelius/dfr.dist.series. R package version 0.1.0.

Alexander Towell. *dfr.lik.series.md: Masked-Cause Likelihood Models for Series Systems*, 2025c. URL https://github.com/queelius/dfr.lik.series.md. R package version 0.1.0.

Alexander Towell. Exponential series systems with masked failure causes: Closed-form MLE and simulation studies. In preparation, 2025d.

Alexander Towell. Model selection for reliability estimation in series systems with masked failure causes. In preparation, 2025e.

Alexander Towell. Maximum likelihood estimation for weibull series systems with masked failure causes. In preparation, 2025f.

Anastasios Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72(1):20–22, 1975.

J.S. Usher and T.J. Hodgson. Maximum likelihood analysis of component reliability using masked system life-test data. *IEEE Transactions on Reliability*, 37(5):550–555, 1988. doi: 10.1109/24.9880.

J.S. Usher, D.K.J. Lin, and F.M. Guess. Exact maximum likelihood estimation using masked system data. *IEEE Transactions on Reliability*, 42(4):631–635, 1993. doi: 10.1109/24.273596.