Interval reliability inference for multi-component systems

KRPELIK, DANIEL

Interval reliability inference for multi-component systems

Daniel Krpelik

A Thesis presented for the degree of
Doctor of Philosophy

Department of Mathematical Sciences
Durham University
United Kingdom

8. May, 2024

Interval reliability inference for multi-component systems

Daniel Krpelik

Submitted for the degree of Doctor of Philosophy

8. May, 2024

Abstract: This thesis is a collection of investigations on applications of imprecise probability theory to system reliablity engineering with emphasis on using survival signatures for modelling complex systems. Survival signatures provide efficient representation of system structure and facilitate several reliability assessments by separating the computationally expensive combinatorial part from the subsequent evaluations submitted to only polynomial complexity. This proves useful for situations which also account for the statistical inference on system component lifetime distributions where Bayesian methods require repeated numerical propagation for the samples from the posterior distribution. Similarly, statistical methods involving imprecise probabilistic models composed of sets of precise probability distributions also benefit from the simplification by the signature representation. We will argue the pragmatic benefits of using statistical models based on imprecise probability models in reliability engineering from the perspective of inferential validity and provision of objective guarantees for the statistical procedures. Imprecise probability methods generally require solving an optimization problem to obtain bounds on the assessments of interest, but monotone system structures simplify them without much additional complexity. This simplification extends to survival signature models, therefore many reliability assessments with imprecise (interval) component lifetime models tend to be tractable as will be demonstrated on several examples.

Statement of Copyright

Acknowledgements

I would like to thank my supervisors professor Frank P.A. Coolen and Dr Louis Aslett for their guidance during my doctoral studies. Their expertise and insights allowed me to gain overview and understanding of the field of reliability engineering and statistics to much deeper extent then I would imagine was possible. Thanks to their patience, I was allowed to explore the vast amount of topics related to problems of system reliability engineering and, eventually, compile a few of them into a thesis.

I would also thank to professor Massimiliano Vasile and the project UTOPIAE (Uncertainty Treatment of Optimization In Aerospace Engineering) of which he was the primary investigator. The project, which was a collaboration of experts from field of mathematics, statistics, engineering and software development, allowed to put the explored topics into a wider perspective of practical engineering and aids in my further pursuit of the field in following research projects.

Table of contents

Chapter 1

Introduction

Devices are of little use to us if they do not function properly, but whether they will function or not is subjected to uncertainty. Reliability theory studies the failure laws, It constructs mathematical models and reasons with the chance of that a device is functioning. Once we have obtained such models, we can take the reliability aspects into account during the design and decision processes.

Systems are collections of interconnected devices, their components, which together provide a common functionality, or a multitude of them. Reliability of a system, its capacity so fulfil its purpose, is also studied in reliability theory. A system can be treated as a singular device, but that would make assessment about its reliability and its improvement practically difficult since the survival analysis requires multiple observations of device's failure in order to draw conclusions. Instead, the model of a system can be split into two parts. The first aims to describe properties and failure laws of each of the devices which compose it, and the second describe their interconnections and their role in providing system's functionalities. Hence, reliability is assessed for each of the components individually and then integrated in order to obtain the reliability of the overall system. This approach allows us to inspect the effect of particular components and propose changes which would be beneficial to

overall system performance.

Reliability itself, as a quantity, is defined as the probability of realization of a desired event. Assessing reliability can therefore be considered as a predictive inference problem. Since system reliability is assessed by integrating reliability of its components, with respect to their role in providing system functionalities, we must consider how to construct predictive inferences on transformed random variables. Given the component probability laws, assessment of system reliability becomes a numerical problem of performing uncertainty propagation.

Assessing reliability of large real-world systems is still a challenging task due to the necessity of summation over the set of all possible combinations of component states in order to apply the law of total probability. Several simplifications are possible for systems with special structure function, like fault trees, or systems with concise structure representation, like binary decision diagrams. In practice, many systems are contain multiple component of the same type which follow identical failure laws. For this special case, survival signature methodology for describing system structure was developed to facilitate reliability analyses.

In a realistic case, we must also consider that the inference on component reliability must be drawn through statistical procedures on samples with finite, often small, size. This introduces additional uncertainty which need to be accounted and has been addressed. The most widely used statistical method uses point estimates which can be easily propagated into system reliability, but these occlude the sampling uncertainties unless some sensitivity analysis is performed, e.g. via bootstrapping. Another class of methods, which incorporate uncertainty of the estimates, are based on Bayesian statistics, which can construct predictive distributions for component states which can easily be numerically propagated into predictive distribution about the system state via Monte Carlo methods. Nevertheless, they rely on the choice of precise prior distributions which may influence the inferences with small samples and are hence referred to as subjective methods.

Sensitivity analysis for Bayesian methods require us to consider sets of prior distributions and their influence on the system reliability. Considering sets of probability laws generate and extension of precise probability theory into *imprecise probability theory*, which provides richer uncertainty modelling framework and leads to more robust statistical methods. Bayesian sensitivity analysis, considering multiple prior distributions, is directly integrated with the statistical inference into generalized Bayesian inference. Imprecise probabilities also provide framework for alternative statistical inferences through direct, possibly set-valued, inversion of the sampling models. Nevertheless, computations with imprecise probability models are not tractable in the general case due to the intrinsic optimization task which is required for constructing bounds on event probability assessments and expected values.

The objective of this thesis is to explore possibilities of using imprecise probability theory for system reliability assessments. It originated from research work for UTOPIAE, H2020, an EU project which explored topics of uncertainty modelling - including imprecise probability theory - and optimization in problems of engineering design.

Several chapters are adaptations of published work of the author:

- Section 4.2 on phased mission analysis is based on a SECESA2018 conference contribution [1],
- Section 4.3 decomposition approach to computing survival signatures is based on an ESREL2019 conference contribution [2],
- Section 5.1 on masked system inference is based on a SMPS2018 conference contribution [3],
- Section 5.4 robust markov analysis is based on a IDT2019 coference contribution [4],
- Section 5.5 on markov chain monte carlo method for sampling set of probability distributions is based on a UQOP2021 conference contribution [5],
- and parts of the reliability theory sections and imprecise probability sections

are based on the respective chapters contributed to a book on uncertainty optimization [6].

Further original contributions include:

- investigation geometrical properties of survival signatures in Section 2.3.4,
- computation of system remaining useful life in Section 4.1,
- theoretical analysis of imprecise Monte Carlo method for a set of uniform distributions in Section 5.6.
- adaptation of inference with coarse observations to system reliability inference in Section 5.3

The rest of the thesis is organized as follows: The basic definitions and mathematical relations used in reliability theory are given in Section 2.1. The section includes definition of reliability and basic notions used to describe its time evolution. Subsequent parts focus on describing and computing reliability of interconnected systems. Several system models are described in Section 2.2. System/survival signatures as an alternative efficient description of the system structure are described in Section 2.3 together with an overview of their applications in Section 2.3.5 and methods for their computation in Section 2.3.6. A brief introducion to uncertainty is given in Section 3.1 followed by an overview of topics from probability theory in Section 3.2. Section 3.3 provides an overview of mathematical theories of imprecise probability models and Section 3.4 gives a brief description of statistical method and their connection to imprecise probabilities. Some examples on applications of system signatures are given in Section 4.1 for computing the system remaining life and in Section 4.2 for assessing reliability of phased missions, which is adapted from [1]. A specific method for computation of survival signatures using decomposition approach adapted from [2] is described in Section 4.3. Section 5.1 studies possibilities of system identification under interval uncertainties adapted from [3]. Investigation in probability models with interval parameters in the context of phased missions are in Section 5.2. Some examples of statistical inferences for system reliability solved

using imprecise probability theory are given in Section 5.3 applied for interval valued observations and in Section 5.4, which is adapted from [4], for assessment of system reliability from system failure observations using markov models. The remaining sections investigate possibilities of numerical methods for imprecise probability inference using Markov chain Monte Carlo in general in Section 5.5, which was presented in [5], and on a simple model to investigate its theoretical properties in Section 5.6.

Chapter 2

Reliability

2.1 Mathematical theory of reliability

Reliability theory is a field at the interface of mathematics and engineering. The primary interest is to ensure that a system (a device, policy, treatment, etc.) will behave as desired. Historically, since the behaviour cannot be predicted with certainty, a lot of interest was allocated into assessing the *validity* of the logical statement *the system will work*. The natural choice of the *validity* measure seems to be probability, since probability theory offers a consistent reasoning apparatus in which one can, deductively, from a set of basic assessments, derive probabilities of related statements similarly as in the familiar system of binary logic.

Probability theory is commonly used for assessing statements about both the frequency of events, and the likeliness of occurrence of a specific outcome in the next conducted trial. The first interpretation focuses on describing the sampling process, the anticipated relative ratio of occurrences of any particular trait of outcome, the aleatory uncertainty. If the predictive model is *well-calibrated*, the relative frequency tends to be *close* to the numerical value assigned to it by the model, where the closeness is understood in a limit of infinitely many trials, as in the law of large

numbers. The second interpretation is also often termed *epistemic uncertainty*, since for any particular separate observation, an assigned numerical value of probability does not correspond to any observable quantity. The outcome of the next observation is a precise value and we will perceive it as such, once the observation is realised. Probability theory, here, serves as a tool to describe our state of knowledge, perception of *likeliness* of occurrence of an event, and allows us to reason about particular attributes of the future observation, including what actions we might take in order to improve the *chance* that the future observation will have desired properties. Like, "Is the system more likely to function if we use component A instead of component B?". In the epistemic interpretation, the quality of a reasoning procedure manifests as an observable quantity through relative frequency of *correct* decisions in a series of repeated applications and the quality of the analytic methods rather than of the constructed models.

Both interpretations are relevant for applications of reliability theory. Aleatory interpretation plays a role, e.g., for planning processes in which we assume that components will need to be replaced over time and we need to schedule the maintenance and inspection policies ([7, 8]), or for the statistical quality control. With good enough models, we can assess the long time costs associated with operating the systems and also optimise the policies addressing their manufacture, maintenance, and the logistical issues associated with the replacements. In such scenarios, failures are anticipated and sometimes, we may discover that using a lower quality component may be beneficial from an economical perspective, leading to overall lower costs of the operation.

On the other hand, with some systems, usually the one-of-a-kind ones, we simply cannot afford for them to fail. Some examples are nuclear power plants, air planes, residential buildings, or space missions. In these cases, we need to *ensure* that either the system works perfectly, or that we can detect an upcoming failure in time to mitigate its consequences. The issue is addressed by the so-called risk analysis which

focuses on enlisting possible undesirable events and construct measures aimed at preventing them. Reliability analysis is used here for quantitative assessment of whether the preventive measures are adequate, but the underlying interpretation of probability is purely epistemic.

In this chapter, we will describe the basic question of reliability theory - predicting the functionality of the system. We will revise the basics of survival analysis and system modelling with further focus on the methodology of survival signatures for system modelling and reliability assessment. In this chapter, we restrict ourselves to description using probability theory, which has been standard in this field, although the nature of the uncertainties involved in some of the applications is better captured by imprecise probability models which will be described later in Section 3.3. The actual meaning behind the assigned numerical probabilities varies among applications and it is, therefore, up to the particular analysts to translate it for their situation. For simplicity, we may, here-on, assume the frequency interpretation of probability measures, thus the probability of system functioning will mean that with increasing amount of instances of the same system, reliability is the limit fraction of the functional ones. Nevertheless, if any quantity in a model is uncertain in an epistemic sense, the overall model inherits this interpretation and can further only be used to describe our degree of *belief* in the occurrence of an event.

2.1.1 Structural reliability

The core of the reliability theory is a specific problem of probability theory - evaluating the probability of an event [9–11]. Advanced problems, survival analysis, multi-state and multi-component systems, then use a more complicated structure of the probability space. The constructed uncertainty models can then be applied to solve decision making problems of maintenance planning or reliability allocation.

Our general aim is to construct a system which will function as desired. Once put into operation, the system will occupy a specific state $x \in \Omega_X$. Suppose that in the

set Ω_X we may further distinguish states which we label as being desirable, $\Omega_M \subset \Omega_X$, to represent what we actually *mean* by if a system functions. This might represent that the stresses on a bridge are smaller than its resistance so it will not collapse or that two planes pass at safe distance and will not crash, etc. But since the system is subjected to interaction with the real world, hence inherits its intrinsic uncertainties, our knowledge about the actual state will also be uncertain. Say we model it by a random variable X obtaining values in Ω_X. Now, instead of precisely determining whether it is functioning (true or false), we must employ more sophisticated method to measure the *validity* of statements. Since we have decided to model uncertainties by the means of probability theory, we will measure the **reliability** of a system by the probability that the event $\{X \in \Omega_M\}$ occurs. The greater the probability becomes, the greater confidence we have that the system will actually work once deployed, or, in the frequency interpretation, the larger fraction of deployed systems will be functional.

From the high level perspective, and for simplicity, we will consider the state of a system as a binary random variable $X \in \{0, 1\}$, with $X = 1$ representing that the system functions, that $X \in \Omega_M$, and $X = 0$ otherwise, i.e. $X \in \Omega_M^c$. There is also a possibility to refine our model to include states of partial failure, or even several degrees of degradation, but we will omit that for it would shift our concerns away from the basic reliability formulation towards general performance prediction.

2.1.2 Survival analysis

A common property of real devices is their deterioration, their reliability will gradually decrease in time. But devices are usually required to function over the whole time periods, w.l.o.g. say the interval $[0, T]$. In order to take the time evolution into account, instead of a single random variable X, we need to investigate the whole stochastic process $X(t)$, representing the state of the system at time t, and reformulate the device mission event as $\{\forall t \in [0, T_M] : X(t) \in \Omega_M\}$.

Much interest in reliability theory lies in modelling this deterioration process [9]. An intuitive way is to consider that a device depletes some intrinsic resources (or equivalently that it is accumulating a "wear out") and define the failure state as such with these resources depleted. Let us consider a non-decreasing function $H(t)$, which models cumulative depletion of some inner resource, and H^0 the amount of this resource available to the device. Then we consider the device functional at time t, if it has not yet depleted its inner resources, i.e. if $H(t) < H^0$. Equivalently, we can interpret this as the device not having reached a critical amount of "wear out", e.g. like accumulation of sediments, or overall mass loss due to abrasion.

In order to provide an assessment of reliability, we need to model $H(t)$ and H^0, which will generally both be uncertain. If these models are available (e.g. based on physics models), we can, again, employ probability theory to assess probability of the event of interest. These models may be available for specific problems (crack formation, abrasion, see [12] for more). If they are not available for the investigated system, reliability theory aims to provide ways of constructing them by the methods of statistical inference. Some examples are shown in Section 3.4 and more can be found in statistics textbooks, e.g. [13].

Let us consider a common scenario for a new device put into operation. If we assume that the device is functioning at time $t = 0$ (we try to assure this by post-production testing but it is possible to generalize the methods for cases with so-called *hidden failures*), we can model the **time to failure** (TTF), the time when the device depletes its inner resources - the time it fails. As a random variable, TTF is non-negative and can be described by its cumulative distribution function (CDF) $F(t)$ or, more commonly in the reliability theory context, its **survival function** $R(t)$.

$$F_{\mathrm{TTF}}(t) = P(\mathrm{TTF} < t), \quad R_{\mathrm{TTF}}(t) = 1 - F_{\mathrm{TTF}}(t). \tag{2.1}$$

The advantage of modelling the TTF lies in the straightforward specification of the probability that a device will be operational over the whole mission time T_M,

$$P(\text{device fulfils its mission}) = P(\text{TTF} > T_M) = R_{\text{TTF}}(T_M). \tag{2.2}$$

Therefore, we investigate how the system reliability changes in time. The state is modelled as a stochastic process $X(t)$ for some $t \geq 0$. The simplest case is when the X is a non-increasing processes (once a system fails it will not recover its functionality). Then we model just the time to the component's failure, TTF, and define the stochastic process as $X(t) = \{\text{TTF} > t\}$.

It is also possible that a system fails and regains its functionality later due to repair or replacement of its components. In these scenarios, we rather model system's **availability** $A(t)$ - its probability of functioning at any givent time: $A(t) = \mathbb{E}X(t)$

The most commonly used distribution to describe the TTF is the exponential one because of its simplicity and convenient mathematical properties. This model assumes that the failures occur at random with constant failure rate regardless its operational history. For the exponential distribution, $R(t) = \exp(-\lambda t)$, for a *failure rate* parameter λ. The exponential distribution is often used just for its mathematical convenience, although there are situations where its usage is justified, e.g. for modelling devices during the stable life phase (see Figure 2.1). Other distributions, which provide more flexible modelling opportunities, are, e.g., Weibull, Cauchy, Log-normal or Gamma distributions. More about the basic mathematical models can be found in any introductory text in reliability theory, e.g. in [9].

A useful specification of TTF can be done via the failure rate function (also sometimes called the *hazard rate*) $\lambda(t)$. Failure rate describes an immediate failure probability:

$$0 \leq \lambda(t) := \lim_{h \to 0^+} \frac{P(\text{TTF} \in [t, t+h] | \text{TTF} > t)}{h \cdot P(\text{TTF} > t)} = \frac{f(t)}{R(t)}, \tag{2.3}$$

where the second equality is valid in cases of absolutely continuous CDFs and $f(.)$

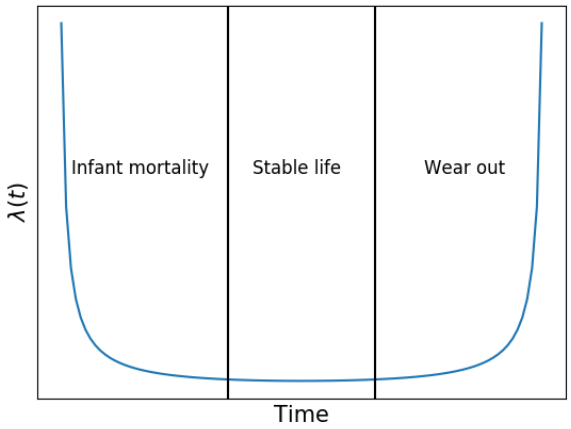


Figure 2.1: The **bathtub curve** demonstrating the evolution of the failure rate for common types of devices' failure modes. Different devices may exhibit various combinations of the depicted modes.

denotes the probability density function (PDF) of the TTF.

Using Equation 2.3, the failure rate is determined by the distribution of the TTF. The opposite is also true, as one can derive the TTF distribution from the failure rate as

$$R(t) = \exp\left(-\int_0^t \lambda(x)dx\right). \tag{2.4}$$

It can therefore be seen also as the rate of depleting the inner resources H^0, as described in this section's introduction. The failure rate also allows us to describe some qualitative properties of the failure laws. A general model of the evolution of the failure rate of a device over its lifetime is depicted in Figure 2.1. During the first period, failures are mainly caused due to the faults of the manufacturing process (the infant mortality), in the second, the device experiences random failures due to the volatile nature of its environment (stable life), and in the last, the failures tend to be caused by wearing out of parts and components (wear out phase).

These phases may be mixed during the device lifetime and often just one is used to describe device failure law. Mathematically, a failure rate is a combination of the following phases:

- the failure rate is constant (stable life) - e.g. electrical components are judged to have constant failure rates,
- the failure rate is increasing (wear out) - e.g. mechanical components are subjected to abrasion, etc.,
- the failure rate is decreasing (infant mortality) - e.g. software, as bugs are discovered and fixed during early stages.

2.2 System reliability

One of the typical attributes of the contemporary world is its complexity. Every device and service available to us provides us with a building block, an opportunity to use it for constructing a new system with its own objectives. As a consequence, many of the devices and services that we rely upon are practically composed of many interconnected smaller subsystems. We can reflect this by our mathematical models and use the tools of probability theory to deduce the probability model for the system behaviour from the probability models for the behaviour of its components and their mutual interconnection. Exploiting the system structure also leads to significant savings of resources. If we were to assess the reliability of a complex system, like a space shuttle, by standard statistical methods, we would have to design an experiment in which we test (break) multiple copies of the same system. This would clearly lead to a vast waste of resources in the system design process. Identifying the system components will, instead, allow us to carry out cheaper experiments separately for those and even utilize our past experience with them. Nevertheless, the separate experiments would not allow us to learn about dependencies among the failure modes of components (common cause failures, cascading failures) which have to be addressed separately.

The way we carry out inference about complex systems may be decomposed into three major stages:

1. Construct a model of the system consisting of components (subsystems) reflecting dependencies between states of the components and the state of the system as a whole.
2. Gather data and carry out inference about components, based on e.g. statistical methods or expert elicitation procedures.
3. Integrate the acquired models of the components behaviour with the model of their influence on the system behaviour in order to obtain a model for time to failure of the whole system.

The first stage is a domain of engineers who have to specify the system topology and carry out risk analyses to identify potential modes of failure and describe how the system operates. Component models can be obtained by statistical methods and are also often included in the component specification in the case of sub-contracting. Although, often, only partial specifications are available in this case like first and second moments of the component failure laws. In this section, we will further focus on the first and the third part of the inference process; on how to integrate the acquired information.

2.2.1 Structure function

Description of the dependency among the state of the system and the states of its components can be provided by a deterministic function. For each possible combination of components states, functioning or failed, we determine whether the system is functioning or not. The uncertainty of the system state will then arise solely due to the uncertainties about the states of its components.

We will restrict ourselves to systems with binary components since it covers many practical scenarios. This restriction can be dropped if necessary to describe any relationship among the system and its components, but would lead to more complicated mathematical models. Let us denote the (deterministic) state of the system as $x_S \in \{0, 1\}$ and a vector of states of its N components as $\mathbf{x} \in \{0, 1\}^N$. We define

the (deterministic) **structure function** as a function φ which maps states of the components onto the state of the system, thus $x_S = \varphi(\mathbf{x})$. The structure function is therefore, in our restricted case, a boolean function on N variables (an example is given in Table 2.1).

Table 2.1: Structure function of a serial-parallel system

$\mathbf{x}$	$\varphi(\mathbf{x})$	$\mathbf{x}$	$\varphi(\mathbf{x})$
0 ,0 ,0 ,0	0	0 ,0 ,0 ,1	0
1 ,0 ,0 ,0	1	1 ,0 ,0 ,1	1
0 ,1 ,0 ,0	0	0 ,1 ,0 ,1	1
1 ,1 ,0 ,0	1	1 ,1 ,0 ,1	1
0 ,0 ,1 ,0	0	0 ,0 ,1 ,1	1
1 ,0 ,1 ,0	1	1 ,0 ,1 ,1	1
0 ,1 ,1 ,0	0	0 ,1 ,1 ,1	1
1 ,1 ,1 ,0	1	1 ,1 ,1 ,1	1

If an uncertainty about the component states is present, first, we model the states of the components by a random vector $\mathbf{X}$. Note the capital letter representing random variables as usual in the probability theory literature. The state of the system will inherit the uncertainty from the states of its components and becomes a binary random variable X_S. We can now assess the system reliability by taking the expectation of $\varphi(X)$,

$$Rel = P(X_S = 1) = \mathbb{E}\varphi(\mathbf{X}) = \sum_{\mathbf{x}\in\{0,1\}^N} \varphi(\mathbf{x}) P(\mathbf{X} = \mathbf{x}). \tag{2.5}$$

The reliability of a system can also be expressed by the reliability function $h : [0,1]^N \to [0,1]$, which directly models the relation between a vector representing probabilities that each individual component functions and probability that the system functions. For example, for a series system (all components have to function to

consider the system to be functioning) with $N = 3$ components with $p_i := P(X_i = 1)$ being the reliability of component i, it holds that $P(X_S = 1) = h(p_1, p_2, p_3) = p_1 \cdot p_2 \cdot p_3$.

The structure function, as defined here, is dependent only on the current states of the components and, thus, allows us to separate static structure dependencies from temporal evolution of component states as described in Section 2.1.2). The same applies for the reliability function, which only depends on the probability that components function at a given time instance. Generalisations are possible, but the actual mathematical model is dependent on the investigated scenario.

The evaluation of a system reliability has generally exponential complexity. It would require us to sum over all the elements of the state space ($\sim 2^N$). A simplification is posible using disjunctive normal form with number of terms corresponding to number of minimal path sets in the system. The reliability function share the number of terms, therefore also the computational complexity.

2.2.2 Basic systems

Here are structure and reliability functions of some common simple systems consisting of N components. Random states of components are denoted by X_i and the probability that a component functions by p_i.

Parallel systems

At least one component needs to be functional. See Figure 2.2.

$\varphi(\mathbf{X}) = \max(\{X_i : i = 1..N\})$

$Rel = 1 - \prod_{i=1..N}(1 - p_i)$

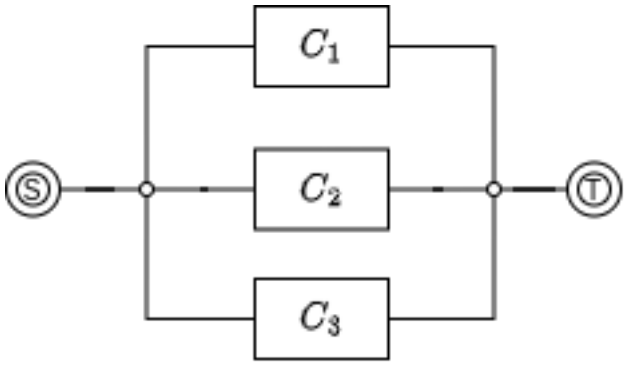


Figure 2.2: Reliability block diagram of a parallel system

Series systems

All components need to be functional. See Figure 2.3.

$\varphi(\mathbf{X}) = \min(\{X_i : i = 1..N\})$

$Rel = \prod_{i=1..N} p_i$

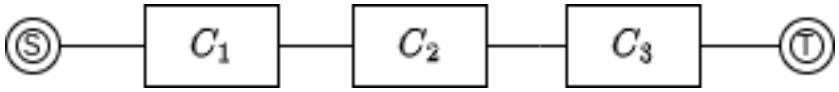


Figure 2.3: Reliability block diagram of a series system

(Voting) K-of-N systems

At least K components need to be functional. See Figure 2.4.

$$\varphi(\mathbf{X}) = \begin{cases} 1; & \sum_{i=1..N} X_i >= K \\ 0; & \sum_{i=1..N} X_i < K \end{cases}$$

$Rel = \sum_{i=K}^{N} \binom{N}{i} p^i (1-p)^{(N-i)}$

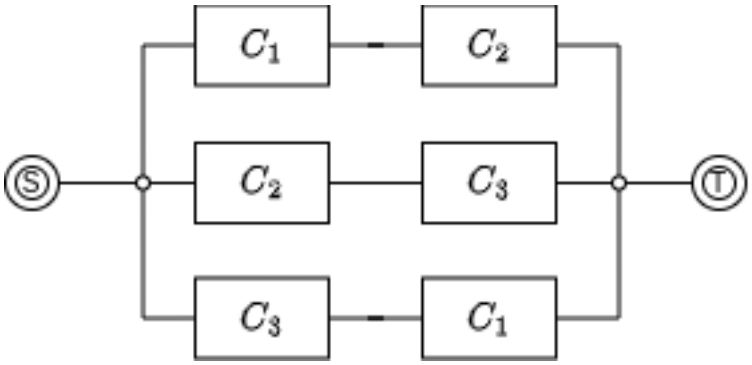


Figure 2.4: Reliability block diagram of a 2-of-3 system

Coherent systems

Coherence is a reasonable assumption for engineered systems. We assume that they function if all their components function, and that component failure cannot lead to an improvement in system function. Monotonicity provides desirable properties for numerical methods and simplifies various algorithms.

A system is coherent if

- its structure function is monotone non-decreasing,
- $\varphi(0) = 0$,
- $\varphi(1) = 1$,

where the arguments of φ is considered as a vector of component states, i.e. 0 and 1 refer to the minimal and maximal lattice element in $\{0, 1\}^N$.

2.2.3 Graphical models

The structure function can be generally described by a table, prescribing the state of the system to every possible configuration, but such a table would be impractical to construct, work with, and inspect for any system of realistic size because the number of rows grows exponentially with the number of components. There exist several alternative ways to specify the structure function. These enable us to present the structure function graphically which also allows us to analyse it qualitatively

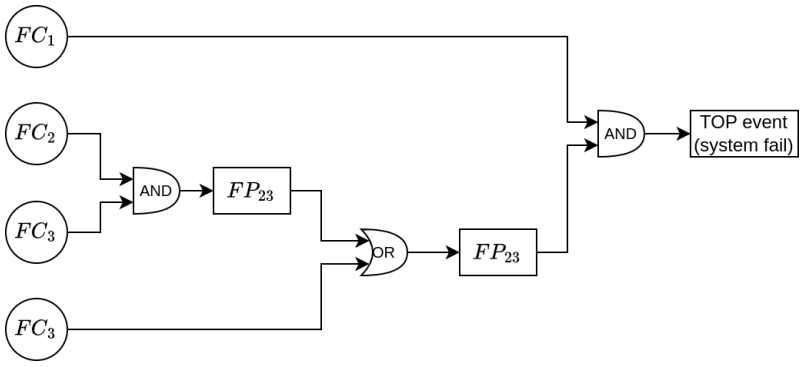


Figure 2.5: A Fault Tree of example system with structure function given in Table 2.1. "TOP" represents failure of the system, FC_i failures of respective components (basic events), FP_x are macro-events. Note that the logic is inverted in comparison with the structure function since both the basic events and the top event represent failures instead of working states.

by the tools and notions of graph theory. Graphical models provide methods to represent system structure in a visually understandable manner and facilitate their communication and validation.

2.2.3.1 Fault trees

Fault trees, or analogically success trees with inverted logic, provide a consise way of representing some multi-component systems. The requirement for using fault trees is that the system can be recursively decomposed into combinations of serial or parallel subsystems, up to the level of the indivual components. If such system representation is possible, it greatly simplifies complexity of evaluating system state and reliability [14, 15]. System reliability can be computed recursively with $O(N \log N)$ complexity.

The fault tree can be constructed using **fault tree analysis** (FTA) [9, 16]. The aim is to, recursively, describe which causes lead to an event being decomposed. We start by defining a top level event, the event of system failure, and investigate which causes trigger it. The causes do not have to be directly elicited in terms of states of singular components, but instead represent state of some subsystem. The procedure recurs to find the causes of these causes up to so-called basic events, the finest refinements of the state space. These may be individual component states. In order to specify the reliability of the system, it is necessary just to describe the

probabilities of the occurrence of these basic events. The state of the whole system is then assessed through a structure function $\varphi(\mathbf{e})$, arguments of which are vectors denoting the occurrence or states of the basic events $e_i \in \{0, 1\}$.

A fault tree represents a hierarchical boolean formula. The actual fault tree is composed of events and gates. The events are events in the sense of probability theory, subsets of the sample space, logical statements (binary). The gates are boolean functions (e.g. AND, OR, K-of-M, NOT) used to describe how the combination of events induces a macro-event higher in the tree hierarchy.

An example of a fault tree corresponding to the structure function in Table 2.1 is shown in Figure 2.5.

The fault tree methodology provide a way for conducting risk analysis of general systems where we cannot construct the structure function nor sometimes even elicit all the components and events influencing the state of the system. The reason is that we advance from the top event to arbitrary depth. The tree, which models the relations between events and the causes of these events, may be constructed from expert knowledge, or from fault logs obtained from deployed systems. The FTA can also easily consider external factors leading to failure.

2.2.3.2 Reliability block diagrams

The influence of system components on the final system state can be specified using a network (a graph). System functionality can be defined using connectivity properties between selected nodes. This leads to possible definition of k-terminal network reliability, where the system is considered functional when the k selected nodes are connected on a subgraph composed of elements representing functional components (nodes and edges). These networks constitute a natural way for modelling systems whose function is related to various kinds of transportation and communication (railroads, computer networks, etc.).

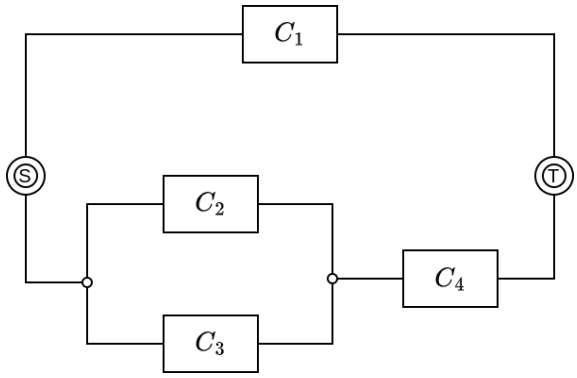


Figure 2.6: An example of RBD which leads to the structure function in Table 2.1. The system is considered functional if there exists a path between terminal nodes S,T through blocks of functional components.

Reliability block diagram (RBD) is a special case of a 2-terminal network system. Two virtual nodes are defined with rest of the nodes representing state of the system components. System is then considered functional if these two virtual nodes are connected by a path through functional components. These virtual nodes may represent initial and terminal state of a process decomposed into its individual subprocesses. Parallel branches represent alternative sub-processes and serial branches represent sequences of dependent sub-processes.

RBDs capture how the system components are connected [17] and are not limited to serial parallel systems and can be generally used to depict any structure function. An example of an RBD, equivalent to the structure function in Table 2.1, is shown in Figure 2.6. It is a 2-terminal network, where the nodes that need to be connected in order to consider the system functional are denoted “S” and “T” and do not correspond to any physical component of the system. Components of the system are represented by nodes “C1-4”, where the number indicates the column in Table 2.1 corresponding to the respective component.

2.2.3.3 Binary decision diagrams

Binary decision diagrams (BDDs) represent Boolean functions as directed acyclic graphs, where each node represents a decision variable and the edges correspond to the possible values of that variable [18]. The representation is based on Shannon’s

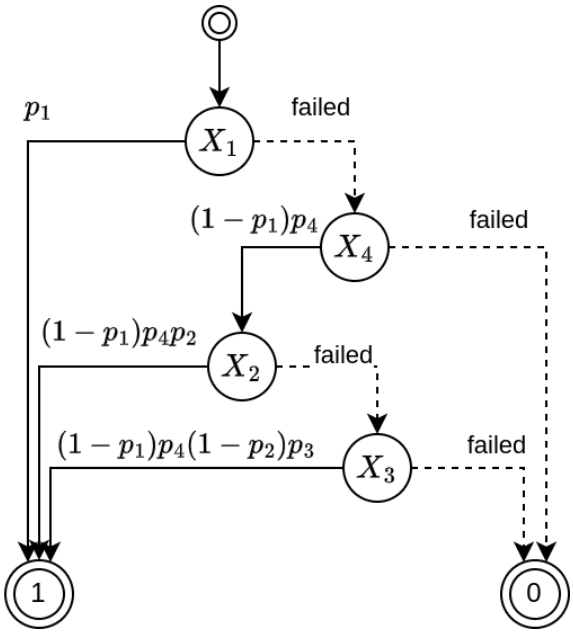


Figure 2.7: An example binary decision diagram with structure function from Table 2.1. Dashed edges represent paths where the respective components are nonfunctioning. For functioning alternative, the probability of the path set is given.

decomposition, which recursivelly decompose the structure function into branches for functional and failed alternatives of individual component states. Path through the branches represent independent events and the probability of their occurence can be reconstructed as a product of respective failure or success probabilities as depicted on Figure 2.7.

One significant advantage of BDDs in system reliability analysis is their ability to efficiently handle large-scale systems. Traditional methods for analyzing system reliability can become computationally infeasible when dealing with systems comprising numerous components or intricate failure scenarios. BDDs offer a compact representation of the system's failure space, enabling efficient traversal and manipulation of the state space [19]. BDD representation can be constructed for arbitrary structure function including fault and event trees [20, 21], networks [22], flowgraphs [23] and systems with common cause failures [24].

2.2.3.4 Bayesian networks

The structure function assumes that the dependencies between the components and the system states are precisely known, but this may not always be the case. It may be that we did not reach necessary depth when constructing a fault tree to ensure unique relationship between an event and its triggers. As an example take a railway trip. One of the trains might be delayed and you miss your connection, but you still might be lucky enough to encounter a helpful railroad clerk who will direct you to an alternative connection. Or not.

To model such situations, the structure function might be naturally generalised to include uncertainties about the dependencies among the states of the events in the fault tree simply by stating the probability of the event obtaining based on the states of the events lower in the tree hierarchy. A (graphical) tool used to depict these models is known as the **Bayesian network** (BN) [25, 26]. This graphical model visualizes assumed conditional independencies among model variables, where variables are assumed conditionally independent given state of their parent variables. This model can also be symbollicaly extended to allow modelling more complicated structures - Bayesian programs [27].

The construction may be done by FTA, but now we do not formulate the dependencies between a macro-event and its causes by boolean functions, but as conditional probabilities. This is, of course, a much more challenging task, but it also provides an advantage. It allows us to work with less detailed models since we need not to advance the tree construction up to the level in which all the relations would be deterministic and these conditional probabilities can be inferred by statistical methods. Thus the main role of the FTA would be to elicit the relevant events and the assumptions on the conditional independence. Once the BN is constructed and the stochastic models are provided, the system reliability may be assessed, deductively, according to the theory of probability.

For $\mathbf{X}$ denoting the random vector of system components' states,

$$P(X_S = 1|\mathbf{X} = \mathbf{x}) = \sum_{y_1,y_2} P(X_S = 1|X_A = y_1, X_B = y_2) \cdot \\ \cdot P(X_A = y_1, X_B = y_2|\mathbf{X} = \mathbf{x}), \tag{2.6}$$

where X_A, X_B are the only macro events such that the state of the system is conditionally independent from component states given X_A, X_B according to our structural assumptions.

The state of the system, can be assessed recursively by marginalising over $\mathbf{X}$.

$$P(X_S = 1) = \sum_{\mathbf{x}\in\{0,1\}^N} P(X_S = 1|\mathbf{X} = \mathbf{x})P(\mathbf{X} = \mathbf{x}). \tag{2.7}$$

Bayesian networks can also be used to model dependencies among the component failures, e.g. common cause failures, where some external disturbance might affect multiple components at the same time. In such a case, component reliabilities may be specified as conditional on the occurrence of this disturbing event, e.g.

$$P(X_i(t) = 1|X_D(t) = 0) = R_i(t) \\ P(X_i(t) = 1|X_D(t) = 1) = 0, \tag{2.8}$$

for some disturbing event $X_D(t)$. In this scenario, the disturbing event would surely render the component failed. In order to assess the overall system reliability, the probability of occurrence of this disturbing event also has to be specified.

An example of a graphical Bayesian network is shown in Figure 2.8. For each of its nodes, a probability table conditional on its predecessors (unconditional for the terminal events) has to be specified.

2.2.4 Phased missions

Some real systems do not operate under the same conditions and with the same functional requirements during their whole lifetime and we might be able to identify different phases of their missions. The physical system may remain the same over

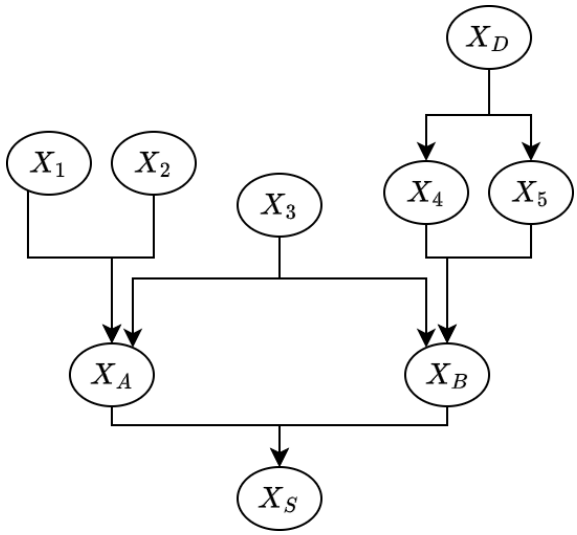


Figure 2.8: Example of a Bayesian network with two macro-events X_A, X_B and a common event X_D disturbing components $4, 5$.

these phases but the functionalities we require it to provide or the loads exerted upon the components may differ among these phases. Such scenarios are known in the literature as phased missions systems (PMS) [28]. An example might be an aircraft journey, where the aircraft must take-off, cruise along the flight path, and, finally, land again. PMSs thus do not provide an alternative way of modelling the structure function, but rather provide a whole new class of complex system scenarios to be analyzed.

The modelling is performed in two basic steps. First, we need to identify different phases and for each of those we construct a model describing what constitutes a successful operation in this phase. These models may be specified by fault tree or RBD models. Then we need to link the models of all the phases together. If the phases are specified by fault trees, this linking will result into a single extended fault tree characterizing the whole mission. Similarly with the RBDs. In both cases, the following treatment is similar to that introduced earlier, but with some specifics which need to be taken into account.

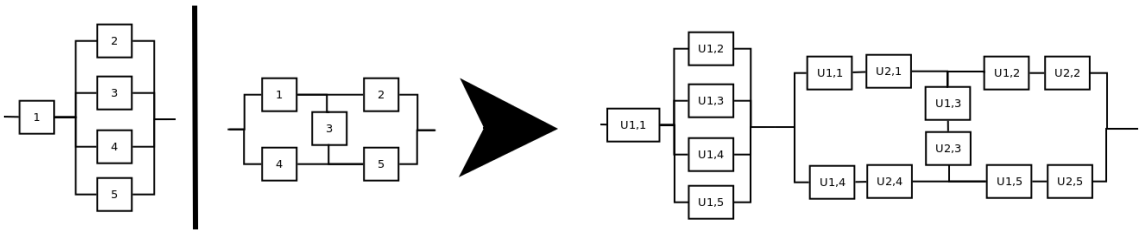


Figure 2.9: Transformation of RBDs of mission phases into a RBD of a phased mission according to the Esary's identity [28]. The events $U_{i,j}$ represent the conditional events that a component j does not fail at phase i given that it is functioning at its beginning. The "start" and "terminal" nodes are omitted.

A mission is considered successful if the system did not fail in any of its phases. From (monotone) structure function point of view, this means that for each time, which denotes the end of a mission phase, a milestone, the system must be functional at that time. The monotonicity assures that the system was functional also during the whole phase. For a mission with K phases and milestones $t_1, \ldots, t_K$, the joint mission structure function is given by:

$$\varphi_{mission}(\mathbf{X}(t_1), \ldots, \mathbf{X}(t_K)) := \prod_{i=1}^{K} \varphi_i(\mathbf{X}(t_i)), \tag{2.9}$$

where $\varphi_{mission}$ represents the structure function of the whole mission (defined as $\varphi_{mission} : \{0,1\}^{N \cdot K} \to \{0,1\}$ for an N component system) and φ_i are structure functions in the respective phases ($\varphi_i : \{0,1\}^N \to \{0,1\}$).

Phased mission models can also be used for an on-line decision making during the mission execution. Once a model of the mission is constructed, we may not only assess the probability of successful completion of a mission, but, in case we have modelled them, also the probabilities of completion of mission deviations. This may be useful in case some disturbances occur, which would endanger the mission's completion.

A method of modelling PMs by combining fault trees of individual phases is presented in [29]. Methods based on binary decision diagrams allow quick assess of risks of possible alternatives and adaptation of the mission [30, 31].

2.3 System signatures

One problem with structure functions is their high dimensionality in practical scenarios. For a system with N binary state components ($x_i \in \{0, 1\}$), the structure function requires specifying 2^N entries, which turns the following reliability analysis into a computationally expensive process.

System signatures were introduced to facilitate analyses of large heterogeneous systems [32–34]. They are used to simplify the prescription of the relation between the system state and the component states. Assuming some simplifying relations among the components interactions, we can compress the remaining relevant aspects of the structure function into a lower dimensional summary - a signature.

Two distinct types of signatures are the **system signature** and the **survival signature**.

The original **system signatures** were introduced by Samaniego [35] and celebrated successful applications in the system reliability analysis and system structure optimization [33]. On the other hand, they could only be applied to systems with components with independent identically distributed (i.i.d.) lifetimes, which is overly restrictive for many practical scenarios since most of the systems are composed of heterogeneous components. This limitation was overcome using **survival signature** (denoted as the A-vector in [36]) and its extension to systems with multiple types of component [37].

2.3.1 Signatures as conditional probabilities

The signatures utilize decompositions of the component state-space $\{0, 1\}^N$ to compute the system reliability

$$P(T_S) = \mathbb{E}\varphi(X) = \sum_{x \in \{0,1\}^N} \varphi(x) P(X = x) \tag{2.10}$$

For an arbitrary ancillary random variable L, we can decompose the formula for computing reliability, Equation 2.10, using the law of total expectation.

$$\mathbb{E}[\varphi(X)] = \mathbb{E}\left\{\mathbb{E}[\varphi(X)|L]\right\}. \tag{2.11}$$

Suppose that we can group the components of the system such that states of the components of the same group are exchangeable. We may choose the ancillary random variable L to be a vector of numbers of functioning components of each group (arbitrarily ordered) and define the system survival signature as $\Phi(l) = \mathbb{E}[\varphi(X)|L = l]$. Due to the exchangeability assumption, the conditional probability $P(X|L)$ is uniform on the subset of $\mathcal{X} = \{0,1\}^N$ for which $L = l$. This allows us to compute Φ according to the laws of classical probability as the fraction of functional states,

$$\Phi(l) = \frac{|\{x \in \mathcal{X} : L(x) = l, \varphi(x) = 1\}|}{|\{x \in \mathcal{X} : L(x) = l\}|}, \tag{2.12}$$

where the notation $L(x)$ represents the dependency of L on the actual component states.

The dynamics of the components' states will induce a stochastic process on the state of the system. Two basic assessments about the system are then usually of interest, the system availability $\mathcal{A}(t) = \mathbb{E}[\varphi(X(t))]$ (the probability that a system is operational at a given time instant, regardless of its history) and the system reliability $Rel(t) = P(\forall\tau \leq t : \varphi(X(\tau)) = 1)$, also called the survival function (the probability that the system has not failed before the time of interest). These two assessments are equivalent if the system is coherent and the component state processes are non-increasing.

The signatures we describe in this section both rely on this formula, but differ in the choice of the ancillary decomposition variable L.

2.3.2 System signature

The system signature is introduced for systems composed of components with i.i.d. times to failure. The i.i.d. requirement either restricts us to analyse systems consisting of multiple instances of the same component or to systems for which we assume that the other components are totally reliable (cannot fail).

The system signature is defined as a discrete probability vector $q_1, ..., q_N$, where q_i denotes the probability that the i-th component failure will result in the failure of the system. The expression for system reliability can be simplified into

$$P(\mathrm{TTF}_{\mathrm{sys}} > t) = \sum_{i=1}^{N} q_i P(\mathrm{TTF}_{(i:N)} > t), \tag{2.13}$$

where $\mathrm{TTF}_{(i:N)}$ denotes the ith order statistic (a random variable describing the probability distribution of ith failure time in the sample of size N).

In the i.i.d. case,

$$P(\mathrm{TTF}_{(i:N)} > t) = \sum_{r=N-i+1}^{N} \binom{N}{r} [1 - F(t)]^r [F(t)]^{N-r}, \tag{2.14}$$

where F is the common CDF for the component lifetimes.

Samaniego has shown that the system signature may serve as a way of comparing systems. He provides theorems about how different stochastic orderings of system signatures implies stochastic orderings of system TTF [33]. This enables us to define system optimization problems as problems of finding systems with optimal signatures (although there is not exactly a one-to-one relation between systems and signatures). He also introduces random mixtures of systems, which enable us to split the system optimization problem into two steps. An optimization problem in the continuum space of mixtures rather than directly the original discrete one and a subsequent problem of finding a system with the most similar signature to that obtained in the previous step.

2.3.3 Survival signature

An extension to system signatures may be made for systems consisting of multiple types of components. In this scenario, we assume that the TTFs of components of the same type are exchangeable (i.i.d. implies exchangeability). This allows us to model more scenarios than the system signature (e.g. network system with both switch-boards and transmission ducts). The scenario in which the TTF of each of the components is different is also included as an extreme case.

Let us assume that we have a system with K distinct component types, denote G_j the set of components of type j and M_j the number of components of type j in the system. We can introduce a natural decomposition of $\{0,1\}^N$ into $D_{\mathbf{l}}$, where $\mathbf{l} \in \otimes_{i=1}^{K}\{0,1,...,M_i\}$ is a multi-index $\mathbf{l} = (l_1,...,l_K)$. The subset $D_{\mathbf{l}} := \{\mathbf{x} \in \Omega_X : \sum \mathbb{I}(x_i \in G_j) = l_j, \forall j\}$ represents the subset of the component state-space $\Omega_X = \{0,1\}^N$. This corresponds to a decomposition into disjoint sets $D_{\mathbf{l}}$ where for each component type j exactly l_j components are functioning. Denoting the event that the system is functional as S, the probability $P(S|D_i)$ may be viewed, due to the exchangeability assumption, as the probability of success in a Bernoulli trial (the number of favourable events divided by the number of all the possible events) and may be derived from the structure function as

$$\Phi(\mathbf{l}) := P(S|\mathbf{X} \in D_{\mathbf{l}}) = \frac{|\{\mathbf{x} \in D_{\mathbf{l}} : \varphi(\mathbf{x}) = 1\}|}{|D_{\mathbf{l}}|} = \left[\prod_{i=0}^{K} \binom{M_i}{l_i}^{-1}\right] \sum_{\mathbf{x} \in D_{\mathbf{l}}} \varphi(\mathbf{x}). \quad (2.15)$$

In the case that the component states are also stochastically independent, the probability that the joint component states belong to a certain partition is

$$P(\mathbf{X} \in D_{\mathbf{l}}) = \prod_{i=0}^{K} \binom{M_i}{l_i} [P(X_i = 1)]^{l_i} [P(X_i = 0)]^{M_i - l_i}. \quad (2.16)$$

Without the independence assumption, another formula needs to be used; perhaps even a numerical simulation.

The survival function of the system is therefore separated into a time dependent (component reliability) and a time independent (system structure survival signature) factors, and

$$P(TTF_{sys} > t) = \sum_{\mathbf{l}=\mathbf{0}}^{(M_1,\ldots,M_K)} P(S|\mathbf{X}(t) \in D_{\mathbf{l}}) P_t(\mathbf{X}(t) \in D_{\mathbf{l}}),$$

where the summation is over all the values of multi-index $\mathbf{l} < \mathbf{M}$ and P_t denotes the time-dependent component state process.

If the TTF distribution of the components is independent on all the other components, the relation simplifies into

$$\begin{aligned} &P(TTF_{sys} > t) = \\ &= \sum_{\mathbf{l}=\mathbf{0}}^{(M_1,\ldots,M_K)} \Phi(\mathbf{l}) \prod_{k=1}^{K} \left[\binom{M_k}{l_k} [P_t(X_i(t) = 1)]^{l_k} [P_t(X_i(t) = 0)]^{M_k - l_k} \right], \\ &= \sum_{\mathbf{l}=\mathbf{0}}^{(M_1,\ldots,M_K)} \Phi(\mathbf{l}) \prod_{k=1}^{K} \left[\binom{M_k}{l_k} [1 - F_k(t)]^{l_k} [F_k(x)]^{M_k - l_k} \right]. \end{aligned} \tag{2.17}$$

The survival signature also allows us to incorporate uncertainties about the system structure. [38] shows how structure function can be treated as a (possibly interval valued) conditionial probability and how this can be integrated into the survival signature methodology. Nevertheless, uncertainty and imprecision in the structure function would require application of specialized methods for computing the signatures from the system specification which would not depend on cut or path set representation.

2.3.4 Geometry of signatures

The survival and system signatures are closely related. For systems with a single type of components, the N-dimensional Samaniego's signature is given as a difference of the $(N + 1)$-dimensional survival signature vector [36, 37].

$$q_i = \Phi(N+2-i) - \Phi(N+1-i); i \in \{1, ..., N\}. \tag{2.18}$$

An attempt was made to define Samaniego's signatures for a multicomponent system with M_k components for types $k \in \{1..K\}$ [37]. This system signature is defined as a $\sum_{k=1..K} M_k$ vector representing probability that the system will fail upon j-th failure of k-th component type:

$$q_k(j_k) = P(T_S = T^k_{j_k:m_k}), \tag{2.19}$$

where T_S is time of system failure and $T^k_{j_k:m_k}$ the j-th order statistics for iid components of type k.

As discussed in [37], such system signature fails to achieve separation of system specification and component failure distributions since these need to be taken into account during evaluation of the signature.

The issues occur because the signature in Equation 2.19 is defined as a marginal probability, therefore needs to be marginalized during computation. Nevertheless, the perspective using system signature as a "differential" (Equation 2.18) of the survival signature still applies and is explored in [39, 40] referred to as the "third direct partial logic derivative (DPLD)", which is denoted $\frac{\partial\Phi(l_1,...,l_K)\Downarrow}{\partial l_k\downarrow}$ in the papers.

Denote $q^{\mathbf{l}+}_{\mathbf{l}-} = \Phi(\mathbf{l}+) - \Phi(\mathbf{l}-)$ the difference of two consecutive survival signature entries such that $l+ > l-$. Then this is a discrete exterior derivative of the survival signature, which acts as a potential in the space of combinations of number of working components by respective types.

For any path p representing progression of the vector $\mathbf{l}$ due to failing components from $\mathbf{l_1} := (M_1, ..., M_K)$ to $\mathbf{l_0} := (0, .., 0)$,

$$\sum_{i=1}^{\sum_{k=1..K}(M_k+1)-1} q^{\mathbf{p_i}}_{\mathbf{p_{(i+1)}}} = \Phi(\mathbf{l_1}) - \Phi(\mathbf{l_0}) = 1. \tag{2.20}$$

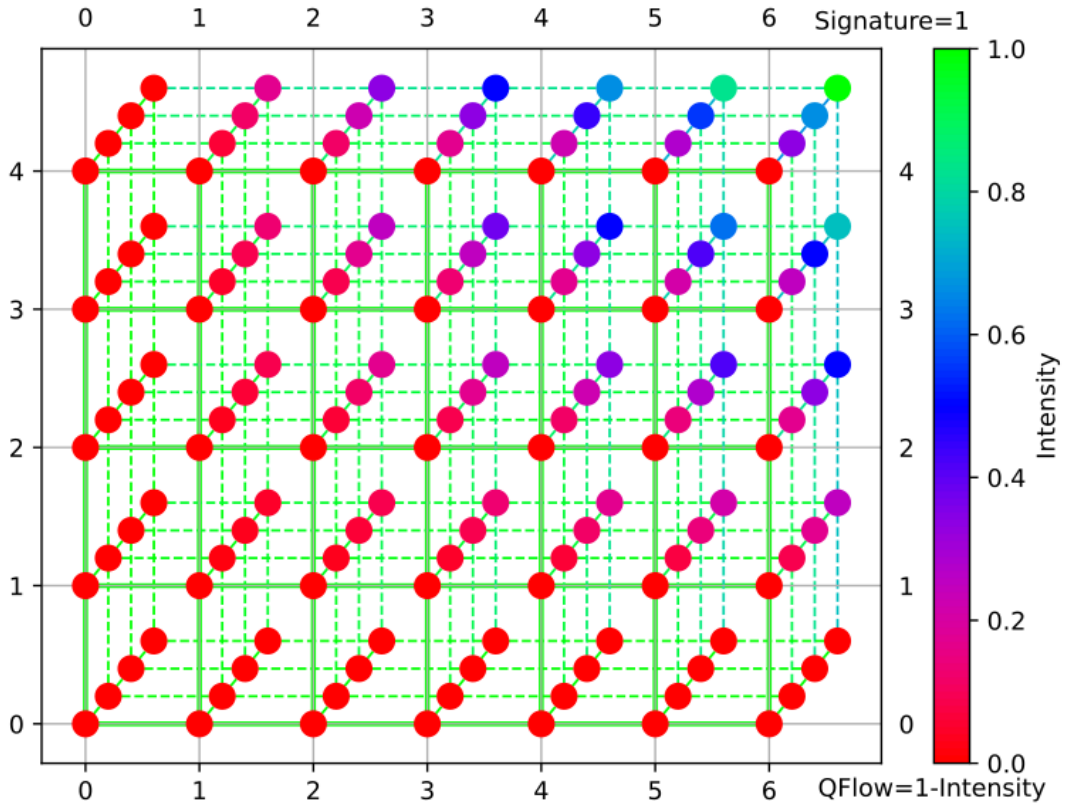


Figure 2.10: Example of the tensor structure of the survival signature. Axes x and y denote number of functional components of the first and the second type, and the tilted axis represent number of functional components of the third type. Node colors represent the survival signature value (green=1, red=0). Edge colors represent the survival derivative (red=1, green=0).

Proposition 2.3.1 (Is signature.)**.** *With the failure progression p fixed,* $q_i := q^{\mathbf{P_i}}_{\mathbf{P_{(i+1)}}}$ *is the system signature in the sense of Samaniego.*

Corollary 2.3.1 (Is independent.)**.** *Probability* $q^{\mathbf{l}+}_{\mathbf{l}-}$ *that system failure occurs during the transition* $\mathbf{l}+ \to \mathbf{l}-$ *is path independent.*

Proposition 2.3.1 can be used to define a sampling technique for computing bounds on individual signature entries. Although these bounds can be uninformative for significant signature entries, they allow to rule out insignificant entries for computation techniques, which evaluate signature entries individually. Similar elimination was used in [41], which proposes using heuristic methods suitable for network systems. The proposed bounds are applicable for general systems.

Consider randomly sampling the failure progression and a failure position based on the progression. For a specific signature entry $\mathbf{l}$ and progressions p passing through the entry, the probability that the failure position d will be strictly lower than $\mathbf{l}$ is exactly the value of the survival signature $\Phi(\mathbf{l})$. Any other progression will enter the quadrant $< \mathbf{l}$ through some entry $\mathbf{m} < \mathbf{l}$ and the probability of failure occuring at a position strictly lower than $\mathbf{l}$ is $\Phi(\mathbf{m}) \leq \Phi(\mathbf{l})$.

$$\begin{aligned} P(d < \mathbf{l}) &= P(d < \mathbf{l} | p \cap \mathbf{l} = \mathbf{l}) + P(d < \mathbf{l} | p \cap \mathbf{l} = \emptyset), \\ &\leq \Phi(\mathbf{l})(P(p \cap \mathbf{l} = \mathbf{l}) + P(p \cap \mathbf{l} = \emptyset)), \\ &\leq \Phi(\mathbf{l}). \end{aligned} \tag{2.21}$$

Similarly for the quadrant $> \mathbf{l}$ with intersection point $\mathbf{m} > \mathbf{l}$ and $\Phi(\mathbf{m}) \geq \Phi(\mathbf{l})$.

$$\begin{aligned} P(d > \mathbf{l}) &= P(d > \mathbf{l} | p \cap \mathbf{l} = \mathbf{l}) + P(d > \mathbf{l} | p \cap \mathbf{l} = \emptyset), \\ &\leq (1 - \Phi(\mathbf{l}))(P(p \cap \mathbf{l} = \mathbf{l}) + P(p \cap \mathbf{l} = \emptyset)), \\ &\leq (1 - \Phi(\mathbf{l})). \end{aligned} \tag{2.22}$$

Therefore, irrespective of the failure progression sampling,

$$P(d < \mathbf{l}) \leq \Phi(\mathbf{l}) \leq 1 - P(d > \mathbf{l}). \tag{2.23}$$

2.3.5 Applications of system signatures

Survival signatures significantly reduce time-complexity of system reliability analyses, especially when repeated computation of system reliability is required like in system survival function reconstruction, importance analysis, or redundancy allocation.

A comprehensive overview of practical applications of survival signatures is given in [32].

Several algorithms for system survival and availability function reconstruction were presented in [34]. These simulation algorithms separate sampling the progression of component states ($[X_1(t), ..., X_n(t)]$), resp. progression of number of functioning components ($[L_1(t), ..., L_k(t)]$), and transform them into a sample of progression of the system state ($X_S(t)$). The presented algorithms construct the samples of system state process by:

1. $X_S(t) \in \{0, 1\}$, such that the system failure can occur upon component failure with probability given by the conditional survival signature ($p_{\text{fail}} = 1 - \Phi_i/\Phi_{i-1}$),
2. $X_s(t) \in [0, 1]$, where the system state is given by the value of the survival signature for component states at time t and understood as a continuous "production level" which is averaged over the Monte Carlo samples.

Survival signatures also allow for seamless transformation of predictive distributions of component lifetimes to predictive distribution of the system state, including cases with imprecise probabilistic predictive distributions. [42] demonstrated integration of non-parametric predictive inference of component life-times. Bayesian inference for component life-times was investigated in [43], together with assessment of optimal redundancy allocation. Inferences based on robust Bayesian inference were integrated in system predictive distributions in [44]. Reliability analysis of two-component networks with imprecise component failure laws was demonstrated in [45].

Systems with common cause and cascading failures can also benefit from survival signature representation. [46] investigates common cause failures using nonparametric predictive inference for predicting number of failing components. Common cause failure models based on alpha-factor models and importance measures for individual common cause groups were studied in [47]. Modelling cascading and common cause failures with survival signatures is presented in [48].

Survival signatures can also be used for performing importance analyses, especially identification of critical components based on system service time. Since the survival

signature clusters components of common type, importance analysis can be natively performed for component types. An importance measure based on logical differential calculus are investigated in [39, 49], which provide component type importance measures conditional on number of functioning components. Importance on the individual component level requires construction of multiple signatures distinguishing the component state. Using survival signature for time-dependent relative component importance was investigated in [50]. This approach is extended further investigated for marginal and extended to joint importance in [51], which also presents importance measures based solely on the single, original system survival signature.

Survival signatures have also been successfully applied for phased mission reliability assessment [52]. Application of survival signatures to phased missions with imprecise component failure times were investigated in [1].

Mean residual life can be computed from the survival signature [53]. Such result can be used for constructing optimal preventive maintenance policies, as was done in [54].

Signatures can also be applied to model lifetime of stress-strength systems under repeated stress cycles [55, 56].

Possible extensions of survival signatures for multi-state systems are investigated in [57–60].

2.3.6 Computation of system signatures

Computation of the signatures is a difficult task with comparable complexity to assessing reliability from a structure function specification. The source of the complexity originates from the necessity to enumerate and evaluate the structure function for each variation of component states, number of which grows exponentially. The advantage is that, for many practical applications, it suffices to compute the signature in advance and use it as a proxy system structure model for repeated

computations in reliability analyses.

According to the definitions of respective signatures, we can compute their entries by enumeration of respective cut or path sets and proper normalization. This method is implemented in an R package ReliabilityTheory [61]. This can also be done by exhaustive traversal over all 2^N component states during which the traversed state is determined to be either a cut or a path set. If the studied system is coherent, specifically when its structure function is non-decreasing, an early stopping can be empoyed to limit the computational complexity as was done in a C++/Python package Signature Calculator [62] for reliability block diagrams.

Integrating system signatures from signatures of two series or parallel subsystems was presented in [63] for system signatures and in [42] for survival signatures. The approach using decomposition into subsystems for a general system, computing signatures of the subsystems, and their subsequent integration was explored in [64] for system signatures and in [2] for survival signatures.

Significant improvement in computation allowing evaluating signatures for larger systems was achieved in [65, 66]. These propose a prior transformation of the structure function into a binary decision diagram form, which leads to significant reduction of the space which needs to be traversed by the computational algorithm. This leads to a recursive scheme, which traverse the binary decision diagram by component indices, computes the survival signatures for both cases - component i is working and component i is failed, and integrate the partial results using efficient tensor operations.

Nevertheless, computation of exact signatures for general non-trivial systems inevitably lead to combinatorial explosion of traversed states. Recently, approximation methods are being introduced. In [41, 67], heuristic methods (dijkstra algorithm to identify the minimal amount of working components and percolation theory for identification of path sets) are first used to identify non-significant entries of the

signatures which are omitted in the subsequent approximation of the remaining entries using Monte Carlo estimates. Limiting the amount of Monte Carlo simulations is addressed in [68] using entropy-driven methods to focus the sampling efforts to relevant entries. Another option is limit the amount of entries for which survival signature is computed and use machine learning methods for estimating the remaining entries as was done in [69] using ensembles of neural networks.

2.3.7 Summary

Both the system and survival signatures allow us to greatly reduce the complexity of system reliability analyses for general systems. They allow us to compress the system specification independent of its component failure distributions. This is especially advantageous in situations when multiple evaluations of system reliability is required, like survival function reconstruction, importance analysis, redundancy optimization or statistical inference as demonstrated in Section 2.3.5. Computation of the signatures still suffers from combinatorial explosion of states, but can be carried in advance and reused in multiple analyses. Some efficient methods for survival signature computation have already been developed (see Section 2.3.6) and more are being investigated. Signatures also provide a level of system annonymization. They are sufficient for carrying reliability inferences, yet identification of the exact system structure from its signature is not possible.

Chapter 3

Uncertainty quantification

3.1 Introduction to uncertainty

The desired outcome of an engineering project is a system which provides the service it was designed for. But the exact future behaviour of a system in the real world is unknown until the system is built and tested. This applies also to the use of familiar systems that operate under novel environmental conditions. This poses a dilemma: how to design systems so that they meet our requirements once deployed? We need some procedure(s) to aid us with the decision making.

We would like to be able to assess consequences of various actions - making predictions of the future behaviour. Science is a field which explores relations between various aspects of reality and constructs models, upon which we may base our predictions. But there is no guarantee that these models are totally accurate. Mathematical models are usually simplifications of the occurring phenomena, often various additional simplifications have to be made in order to make the computation tractable, and the numerical evaluation itself may introduce additional error (e.g. when simulating processes described by differential equations). Furthermore, another type of error is introduced when providing numerical inputs for the models, their parameters. These

also come from scientific inference and, therefore, suffer from the similar issues as the models themselves. Their usual sources are: measurements with finite resolution, statistical inference, and expert elicitation. All these are subjected to uncertainty.

The field of modelling uncertainties is nowadays dominated by two complementary theories, probability theory and interval arithmetic. Although these two allow us to model many situations, they contain several drawbacks, which make their application difficult for practical applications. Interval arithmetic, used for worst-case scenario modelling, is often overly conservative and fails to capture correlations among quantities of interest. Probability theory requires us to specify how likely the occurrence of each possible outcome is, which can be impossible up to the required level of precision required to construct the mathematical models.

Combining interval arithmetics with probability theory yields a relatively novel theory of *Imprecise Probability* (IP), which allows greater flexibility in modelling uncertainties and ignorance. We will provide some applications of imprecise probability theory for systems reliability engineering. Imprecise models are already well established in the field of reliability engineering. Barlow and Proschan [10] constructed bounds on survival functions based on qualitative properties of the hazard functions (increasing/decreasing hazard rate). Recently, it has been argued that precise probability models are not appropriate for risk quantification [70] and proposed an approach resembling hypothesis testing. The issue with the precise probability model can be reduced to the following example: consider evaluating risk of a falling ball hitting a point at a ground. If the ball is directly above the mark, it will hit it with certainty. Nevertheless, if the initial position of the ball is uncertain, say modelled by a Gaussian distribution, the increasing uncertainty (variance in position) will actually decrease the evaluated probability of hitting the target - hence the risk evaluated by probability risk assessment methodology. Therefore, using precise probability models for risk assessment can lead to false confidence in system safety.

Regarding probability theory, we will show two complementary approaches for building an axiomatic theory of probability. The fist one is based on Kolmogorov's formulation [71] in which a probability distribution is represented by a positive additive measure (Section 3.2.1). Such measure, which is a set function, will directly encode the modelled probabilities of various assertions about the outcomes of random experiments, allows us to assess an *expected value* of a random variable, and also extends the models to derived random quantities. This approach to probability theory has became dominant across fields because it offers an intuitive description of random outcomes and enables us to construct efficient general algorithms for solving many practical problems (Monte Carlo algorithms, Bayesian inference, etc.). Some extensions of the measure theoretic formulation for imprecise probability will be shown in Section 3.3.1.

Another approach for constructing an axiomatic base for probability theory is based on a functional representation of random quantities [72, 73] described in Section 3.2.3. Uncertainty models are represented by a functional, the *prevision*, which corresponds to the expected value operator in the measure theoretic approach. A model is specified by assessing the expected values for several selected functions - the random variables. This allows us to pose less assumptions on the models since the underlying probability measure is not required to be specified exactly, but also pose limitations in extending the assessments to derived quantities. These extensions will (mostly) result only in bounds on the expectations of the derived random variables. Generalizations of this approach, the lower prevision theory, is described in Section 3.3.4.

Finally, we will show how imprecise probability models naturally arise in statistical inference. Imprecise probability formulation allows us to decrease the amount of necessary additional modelling assumptions and lead to more trustworthy - valid - inferences. Specifically, we will discuss a less-known theory of fiducial inference in Section 3.4.3, which aims to provide direct statistical inversion without the necessity to specify Bayesian priors. The aim of fiducial inference is to provide posterior

distributions with proper calibration properties and validity guarantees, which are further elaborated in Section 3.4.5 which discusses a promising theory of Inferential Models.

3.2 Probability theory

Suppose that we are to perform an experiment. We will denote the set of all its possible outcomes as the *sample space*, Ω. Let us further assume that an outcome cannot be exactly determined prior to its actual observation - it is uncertain. But even though the experimental outcome can be random (i.e. we are not able to predict it with any finite algorithm), multiple repetitions of the same experiment may follow some predictable *law*. Probability theory aims to describe these laws.

3.2.1 Basics of probability theory

Definition 3.2.1 (Probability.)**.** Let Ω be a sample space and $\mathcal{A}$ a σ-field (a collection of subsets) over Ω. We will call a set function $P : \mathcal{A} \to \mathbb{R}$ a *probability measure*, if

- $\forall E \in \mathcal{A} : P(E) \in [0, 1]$,
- $P(\Omega) = 1$,
- $\forall E_i \in \mathcal{A}$, which are mutually disjoint, $: P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$.

The tuple $\mathcal{K} := (\Omega, \mathcal{A}, P)$ from Definition 3.2.1 is called a *probability field*.

Let us again consider the experiment with the set of possible outcomes Ω. If the law of outcomes can be described by a probability distribution P, it represents that over multiple repetitions of the (exactly the same) experiment, for any chosen $E \in \mathcal{A}$, the relative number of outcomes which will be elements of E, will converge to $P(E)$ as the number of repetitions increases.

Engineering problems usually contain multiple quantities of interest and models describing their connection. If these quantities are uncertain, and the law governing

their observations can be described by a probability distribution, we could specify a sample space for each of them and derive assessments about the outcomes for each of them. But in order to provide a connection among them, including possible correlations, we will consider the actual quantities as *random variables*.

Definition 3.2.2 (Random Variable.)**.** Let $(\Omega, \mathcal{A}, P)$ be a K-probability field and $(\Omega_X, \mathcal{A}_X)$ a measurable space. Any $\mathcal{A}_X$-$\mathcal{A}$-measurable function $X : \Omega \to \Omega_X$ will be called a *random variable* (RV).

Let us also denote $\mathcal{L}(\Omega)$ the set of all random variables on $(\Omega, \mathcal{A})$.

Now we are interested in how can we derive the distribution law of the random variables. The probability measure from Definition Definition 3.2.1 measures probabilities of events in the sample space. We need to *propagate* this model in order to assess statements about random variables. The answer will also allow us to specify distributions of other derived quantities, like $Z = f(X, Y)$.

Definition 3.2.3 (Preimage.)**.** For an arbitrary mapping $f : X \to Y$, for any set $E \subset Y$ its *pre-image* is

$$f^{-1}(E) := \{a \in X : f(a) \in E\}. \tag{3.1}$$

So a random variable, a mapping from a probability space, induces its own probability space with set of outcomes being the image of original sample space, $\Omega_X = X(\Omega)$ and $\mathcal{A}_X = \{X(E) : E \in \mathcal{A}\}$.

For a set $E \in \mathcal{A}_X$, the probability that the RV X obtains a value in E is

$$P_X(E) := P(X \in E) = P(X^{-1}(E)) = P(\{\omega : X(\omega) \in E\}). \tag{3.2}$$

Each random variable induces its own probability field $(\Omega_X, \mathcal{A}_X, P_X)$, where $P_X = P \circ X^{-1}$. We will use notation $X \sim P_X$ for denoting that P_X is the (induced) measure of X. For the derived quantities, say $Z = f(X, Y)$, we may proceed analogically.

Furthermore, the situation simplifies if the sample spaces Ω and Ω_X are the real lines and the mapping defining the new random variable is monotone.

Theorem 3.2.1 (Random variable transform.)**.** *Let $X : (\mathbb{R}, \mathcal{B}, P) \to (\mathbb{R}, \mathcal{B})$ be a strictly increasing random variable and $[a, b] \subset \mathbb{R}$ an interval. Because there exists unique classical inverse X^{-1} of X, which is also an increasing function, we can express the distribution of X as*

$$\begin{aligned} P_X([a,b]) = P(X \in [a,b]) &= P(X < b) - P(X < a) \\ &= P([X^{-1}(a), X^{-1}(b)]). \end{aligned} \tag{3.3}$$

And similarly for a decreasing X, where the interval the for pre-image is given by swapping the bounds, i.e. $[X^{-1}(b), X^{-1}(a)]$.

Special attention in probability theory is given to the cumulative distribution functions (CDFs), which represent the probability that a real valued RV attains a value smaller than the argument. A CDF $F_X : \mathbb{R} \to [0, 1]$ represents, for each $a \in \mathbb{R}$, the probability of event $\{X \in [-\infty, a]\}$. The probability of all the other events in the Borel algebra on the real line, $\mathcal{B}$, can be derived through the axioms of probability measures. CDF therefore uniquely represent whole probability measure, which would be intractable to work with otherwise.

In special cases, the CDF of a derived quantity can be easily derived from Theorem 3.2.1. For increasing functions f, the CDF of an extended RV $Y = f(X)$ can be calculated as

$$F_Y(y) = P(f(X) < y) = F_X(f^{-1}(y)). \tag{3.4}$$

Similarly for the case of a decreasing function f, where $F_Y(y) = 1 - F_X(f^{-1}(y))$.

3.2.2 Repeated sampling

Compared to the analysis on the real line, it is convenient to also introduce some summaries of the random variables. This can be done through the *expected value* functional ($\mathbb{E} : \mathcal{L}(\Omega) \to \mathbb{R}$), where $\mathcal{L}$ is space of all real random variables on the sample space Ω.

Definition 3.2.4 (Expectation of a ranfom variable.)**.** Let X be a RV on $(\Omega, \mathcal{A}, P)$. Then

$$\mathbb{E}_{P_X}[X] := \int_{\Omega_X} x dP_X = \int_{\Omega} X(\omega) dP, \tag{3.5}$$

will be called the *expected value* of random variable X. The subscript representing the underlying distribution (P_X) is usually omitted, and we will also do so for precise random variables. We introduce it because of the necessity of computing expected values for various probability measures later in the chapter.

The expected value represents a *typical value*, the limit of average values of multiple draws. By the *law of large numbers*, the average of the finite amount of draws from X will converge towards $\mathbb{E}X$.

Theorem 3.2.2 (Law of Large Numbers.)**.** *Let $X_1, \ldots, X_n$ be a series of random variables such that each of them is distributed according to the same law with a finite expected value $\mathbb{E}X_1 =: \mu \in \mathbb{R}$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu, \tag{3.6}$$

where the real number μ can be viewed as a degenerate random variable M s.t. $\forall \omega \in \Omega : M(\omega) = \mu$.

The law of large number is the core principle which allows us to perform statistical inference. It guarantees that we will approach correct assessments about the sampling distributions (means, other moments, probability statements, etc.) with increasing number of observations.

Another important result of the probability theory, the *central limit theorem*, states what is the asymptotic convergence rate towards these values. It also provides theoretical guarantee for convergence of numerical Monte Carlo algorithms [74] which are used to approximate probability distributions of random variables.

Theorem 3.2.3 (Central Limit Theorem.)**.** *Let $X_1, \ldots, X_n$ be a series of random variables such that each of them is distributed according to the same law with a finite expected value $\mathbb{E}X_1 =: \mu \in \mathbb{R}$ and a finite variance $\mathbb{E}(X_1 - \mu)^2 =: \sigma^2 \in \mathbb{R}^+$. Then*

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}\sigma} \sum_{i=1}^{n} (X_i - \mu) \sim \mathcal{N}(0, 1), \tag{3.7}$$

where $\mathcal{N}(0, 1)$ is the standard normal distribution. The convergence is meant in the distributional sense:

$$X_i \xrightarrow{d} X \Rightarrow \forall Y \in \mathcal{L}(\Omega) : \mathbb{E}_{P_{X_i}}[Y(X)] \to \mathbb{E}_{P_X}[Y(X)]. \tag{3.8}$$

3.2.3 Probability via expectation

Another approach to build an axiomatic theory of probability is explained by Whittle [72]. The idea is that, instead of focusing on the probability distributions as measures on a sample space, we investigate the random variables from the functional perspective. Whittle therefore describes a system based on axioms posed on the expected values, instead of on the probability measures. He further shows how to reproduce results of the standard approach (measure-based) to probability theory.

Definition 3.2.5 (Expected Value.)**.** Let $(\Omega, \mathcal{A})$ be a measurable space and $\mathcal{L}(\Omega)$ the set of all random variables on Ω. We will call a functional $\mathbb{E} : \mathcal{L}(\Omega) \to \mathbb{R}$ the *expected value*, if

- $\forall X \in \mathcal{L}(\Omega) : X \geq 0 \Rightarrow \mathbb{E}(X) \geq 0$,
- $\forall a, b \in \mathbb{R}, \forall X, Y \in \mathcal{L}(\Omega) : \mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$,
- $\mathbb{E}(\mathbf{1}) = 1$,

- if, $\forall \omega \in \Omega$, a sequence $X_n(\omega)$ increases monotonically to $X(\omega)$, then $\mathbb{E}(X) = \lim \mathbb{E}(X_n)$.

Starting from the axioms of the expected value functional leads to the same theoretical system as Kolmogorov's approach. In the precise probability case, both the approaches are equivalent, but this is no longer the case with imprecise probabilities, where the theory of Lower Previsions [75] allows us to represent a larger set of models than the IP extension of the measure theoretic approach.

A similar approach to formulate probability theory was also explored by de Finetti and Savage [73, 76] from a decision making perspective. A notable difference introduced by de Finetti is that the expected values are called *previsions*, but denote the same object. Also, in the notation, de Finetti further does not distinguish between the symbol for probability of an event and for an expectation of a random variable, both are P, because there exists a one-to-one mapping between events and binary random variables. We can obtain the probability of any event $E \subset \Omega$ by calculating the expected value of its indicator function $I_E \in \mathcal{L}(\Omega)$. The meaning is usually evident from the context.

$$I_E(\omega) := \begin{cases} 1, & \omega \in E, \\ 0, & \omega \notin E, \end{cases} \tag{3.9}$$

$$P(E) = \mathbb{E}(I_E).$$

An interesting feature, that both Whittle and de Finetti have presented, is the possibility to extend our partial knowledge to other random variables. Given a set of known expectations for RVs $\mathcal{K} \subset \mathcal{L}(\Omega)$, we can derive bounds for the expectation of another RV $Y \notin \mathcal{K}$.

To do this, we can use corollaries of the expectation axioms (Definition 3.2.5). Given two RVs (mappings from the sample space) s.t. $\forall\omega \in \Omega : X(\omega) \leq Y(\omega)$, we can derive, due to the linearity and positivity of the expectation functional, that $\mathbb{E}(X) \leq \mathbb{E}(Y)$. Therefore, if we already know the expectations $\mathbb{E}(X_n)$ of several RVs X_n, we trivially know the expectation of their arbitrary countable linear combination $Z = \sum a_i X_i$, which is $\mathbb{E}(Z) = \sum a_i \mathbb{E}(X_i)$.

For an arbitrary RV Y, we can acquire the lower bound on its expectation by taking the supremum over all the RVs in the span of $X_1, ..., X_n$, which are strictly lower than Y. Similarly for the upper bound.

Theorem 3.2.4 (Extension of the expected value operator.)**.** *Let $\mathcal{K} = \{X_1, ...\}$ be a countably infinite set of random variables with known expectations and $\mathcal{Z} := \{\sum_{i=1}^{n} a_i X_i + b : X_i \in \mathcal{K}, a_i, b \in \mathbb{R}, n \in \mathbb{N}\}$ the linear span of $\mathcal{K} \cup \{\mathbf{1}\}$. The consistent (coherent in de Finetti's treatment) bounds for the expected value $\mathbb{E}(Y)$ can be obtained as*

$$\sup_{Z \in \mathcal{Z}, Z \leq Y} \mathbb{E}(Z) \leq \mathbb{E}(Y) \leq \inf_{Z \in \mathcal{Z}, Z \geq Y} \mathbb{E}(Z), \tag{3.10}$$

where by $Z \geq Y$ we mean that $\forall\omega \in \Omega : Z(\omega) \geq Y(\omega)$.

For $\Omega, \mathcal{K}$, which are both finite, where $|\Omega| = N, |\mathcal{K}| = n$, the extension is a linear program.

$$\sup_{Z \in \mathcal{Z}, Z \leq Y} \mathbb{E}(Z) = \sup_{\substack{b \in \mathbb{R}; \quad a \in \mathbb{R}^n \\ \forall \omega \in \Omega : b + \sum_{i=1}^{n} a_i X_i(\omega) \leq Y(\omega)}} b + \sum_{i}^{n} a_i \mathbb{E}(X_i). \tag{3.11}$$

Its dual is

$$\sup_{Z \in \mathcal{Z}, Z \leq Y} \mathbb{E}(Z) = \inf_{\substack{p \in (\mathbb{R}_0^+)^n \\ \sum_i^n p_i = 1 \\ \forall X_j \in \mathcal{K} : \sum_{i=1}^{n} p_i X_j(\omega) = \mathbb{E}(X_j)}} \sum_{i}^{n} p_i Y(\omega_i), \tag{3.12}$$

which effectively means, that we are extremizing the expectation over some *set of admissible distributions* p, which would yield the known expectations for all $X_n \in \mathcal{K}$.

3.3 Imprecise probability theory

The history of imprecision in probabilistic assessments dates back to Boole's work on inductive logic [77], who identified certain situations in which the available information is not sufficient to construct precise probability statements. For some probability assessments, only bounds could be constructed.

Bounds on probability statements can also be used when precise values were intractable (e.g. Markov inequality, Jensen's, Chebyschev's). Some early examples may also be found in the field of sensitivity analysis for statistical inference [78]. Nevertheless, by the mid 20th century, a separated theory of imprecise probabilities began to emerge as a generalisation of probability theory. This includes, among other, introduction of non-additive measures by Choquet [79], generalisation of statistical inference by Dempster [80], Walley's work on statistical inference with imprecise probabilities [81], and development of the theory of lower previsions [75].

Special emphasis is put on *extending* our (partial) specification of the model to answer enquiries about derived quantities in a consistent manner. This means that given some claims about some aspects of some random variables, we are interested in what other claims can be deduced about transformed random variables.

3.3.1 Interval probabilities

We begin by introducing an extension to measure-theoretic probability given by Weichselberger [82] and similarly by Walley [83]. The idea is to assign to every event $E \in \Omega$ a pair of real numbers $\in [0, 1]$, which represent lower and upper bounds for the probability of that event. These lower and upper bounds will reoccur within the rest of the theory of imprecise probabilities because they are extreme points of an underlaying convex set of precise probability distributions. They also have an epistemological interpretation - the lower bound measure the evidence supporting the occurrence of E, and the upper bound measure the evidence, which contradicts

E. Their difference can be used as a measure of our ignorance.

Definition 3.3.1 (R-probability)**.** An interval valued set function P on a measurable space $(\Omega, \mathcal{A})$ will be called a *R-probability*, if - $\forall E \in \mathcal{A} : P(E) = [L(E), U(E)]$ s.t. $0 \leq L(E) \leq U(E) \leq 1$, - the set $\mathcal{M} := \{p : p \text{ is K-probability }, \forall E \in \mathcal{A} : p(E) \in P(E)\} \neq \emptyset$.

We will call the tuple $\mathcal{R} := (\Omega, \mathcal{A}, L, U)$ a *R-probability field*. The set $\mathcal{M}$ from the second axiom is called the *structure* (Weichselberger) or the *credal set* (Walley) of $\mathcal{R}$.

The letter R- represents *reasonable*. It corresponds to the property of *avoiding sure loss* introduced in Section 3.3.4. The definition directly implies, that for a R-probability P, $L(\emptyset) = 0$ and $U(\Omega) = 1$ through the non-emptiness of the credal set.

An R-probability is directly connected to a set of precise probabilities via its credal set. It is also apparent, that for any set of precise probabilities we may construct an R-probability s.t. this set will be a subset of its credal set, but such construction may not be unique. In order to specify an R-probability, we would need to define the L and R functions for all the elements in the respective algebra $\mathcal{A}$. Precise probability theory uses cummulative density functions of probability density functions (Radon-Nikodym derivatives of the probability measure) for this purpose, from which probabilities of particular events can ba derived. Analogues for the interval models exist in the form of P-boxes [84] or pairs of probablity density functions [85].

Simpler IP models can be used to assess judgements about more complex ones. If we specify an R-probability via (possibly finite) set of precise models, we automatically also include all their convex combinations in the structure of such an R-probability. Given two K-probabilities $p, q \in \mathcal{M}$, the credal set of some R-probability, for all their convex combinations $r = \lambda p + (1 - \lambda)q, \quad \lambda \in [0, 1]$, $r(E)$ will take values in between $p(E), q(E)$ for all the events $E \in \mathcal{A}$ due to the axioms of probability measures. Thus, $L(E) \leq r(E) \leq U(E)$ so $r \in \mathcal{M}$.

Therefore, the credal set of a R-probability is equal to its convex hull.

A one-to-one correspondence between probability bounds and the underlying credal set is a desired feature in IP theory. In Weichselberger's treatment, this may be achieved by tightening the requirements on the probability bounds L, U.

Definition 3.3.2 (F-probability)**.** An R-probability P, which also satisfies

$$\forall E \in \mathcal{A} : \left(\inf_{p \in \mathcal{M}} p(E) = L(E)\right) \quad \wedge \quad \left(\sup_{p \in \mathcal{M}} p(E) = U(E)\right),$$

is called an *F-probability* and the corresponding tuple $\mathcal{F} := (\Omega, \mathcal{A}, L, U)$, the *F-probability field*.

The letter F- represents *feasible*. It corresponds to the notion of *coherence* introduced in Section 3.3.4. For every F-probability, we also immediately obtain that $L(\Omega) = 1$ and $U(\emptyset) = 0$, because this is true for each of the elements of the credal set, therefore also their infimum and supremum. The axioms of F-probabilities directly imply a relation which is reoccurring throughout many parts of IP theory and which enables us to focus our attention solely on either the L or the U function. The other follow through the conjugacy property

$$\forall E \in \mathcal{A} : U(E) + L(E^c) = 1. \tag{3.13}$$

Example 3.3.1 (Interval probabilities.)**.** An example of F- and R- probabilities is given in Table 3.1. L_F, U_F and L_R, U_R correspond to F- and R- probability bounds respectively. A K-probability is also shown in column P to demonstrate non-emptiness of the respective credal sets.

Since an F-probability defines an underlying credal set and the probability assessments can be obtained through extremization over this set, we may define also the lower and upper expected values for derived random variables through similar extremization. The lower expectation would be given by Equation 3.14 and the upper one similarly with taking a supremum instead of the infimum.

$$\underline{\mathbb{E}}[X] := \inf_{p \in \mathcal{M}} \mathbb{E}_p[X]. \tag{3.14}$$

Table 3.1: Example of F- and R- probabilities on a simple finite sample space $\Omega = \{0, 1, 2\}$. $L \leq P \leq U$.

E	L_F	U_F	P	L_R	U_R
$\emptyset$	0.00	0.00	0.00	0.00	0.33
$\{0\}$	0.33	0.60	0.50	0.33	0.60
$\{1\}$	0.30	0.50	0.40	0.30	0.50
$\{2\}$	0.00	0.37	0.10	0.00	0.55
$\{0, 1\}$	0.63	1.00	0.90	0.00	1.00
$\{0, 2\}$	0.50	0.70	0.60	0.00	0.90
$\{1, 2\}$	0.40	0.67	0.50	0.00	0.83
$\{0, 1, 2\}$	1.00	1.00	1.00	0.50	1.00

A desirable property in the imprecise probability framework is the possibility to derive probability bounds on all the events $E \in \mathcal{A}$ from the knowledge of the bounds only for some events $E' \in \mathcal{A}' \subset \mathcal{A}$. This operation is commonly referred to as an *extension* (with some adjectives corresponding to the theories and actual definitions). In Weichselberger's treatment, a F-probability may be derived from the underlying set of K-probabilities.

Definition 3.3.3 (Support.)**.** Let $(\Omega, \mathcal{A})$ be a measurable space. Denote $\mathcal{A}' = \mathcal{A}/\{\emptyset, \Omega\}$ and let $\mathcal{A}_L, \mathcal{A}_U \subset \mathcal{A}'$.

If there exist a non-empty set $\mathcal{M}$ of probability distributions and set functions L, U s.t.

- $\forall E \in \mathcal{A}_L, p \in \mathcal{M} : L(E) \leq p(E)$,
- $\forall E \in \mathcal{A}_U, p \in \mathcal{M} : U(E) \geq p(E)$,
- $L(\emptyset) = U(\emptyset) = 0, \quad L(\Omega) = U(\Omega) = 1$,

then $P = (L, U)$ is called a *partially determinate R-probability.*\ We will call $(\mathcal{A}_L, \mathcal{A}_U)$ the *support* of P.

Definition 3.3.4 (Partially determinate F-probability)**.** Let $(\Omega, \mathcal{A}, L, U)$ be a partially determinate R-probability field with support $(\mathcal{A}_L, \mathcal{A}_U)$. If also

- $\forall E \in \mathcal{A}_L : L(E) = \inf_{p \in \mathcal{M}} p(E)$,
- $\forall E \in \mathcal{A}_U : U(E) = \sup_{p \in \mathcal{M}} p(E)$,

then P is called *partially determinate F-probability.*

For partially determined F-probabilities, there exist a straightforward way of calculating probability bounds for events outside of their support. The procedure is called *normal completion* in Weichselberger's and *natural extension* in Walley's treatment. It simply exploits the extremizing property of F-probabilities over their respective credal sets $\mathcal{M}$, thus

$$\forall E \in \mathcal{A} : L(A) = \inf_{p \in \mathcal{M}} p(A). \tag{3.15}$$

Example 3.3.2 (Partial F-Probability)**.** Let us consider a partially determinate F-probability on $\Omega = \{0, 1, 2\}$, $\mathcal{A}_L = \mathcal{A}_U = \{\{0\}, \{1\}\}$ with L, U on $\mathcal{A}_L$ given in Table 3.1, and a credal set $\mathcal{M}$.\ For an arbitrary event $E \in \mathcal{A}$, we can calculate its lower probability by solving the optimization problem

$$L(E) = \min_{p \in \mathcal{M}} p(E).$$

Especially, denoting $a := P(\{0\}), b := P(\{1\}), c := P(\{2\})$,

$$L(\{0,1\}) = \min_{\substack{0.33 \leq a \leq 0.67 \\ 0.1 \leq b \leq 0.17 \\ a+b+c=1}} a + b = 0.67.$$

The bounds for the rest of events in $\mathcal{A}$ are given in Table 3.1.

Weichselberger's treatment provides a useful framework to define the lower and upper probabilities and even desired extending properties for partial specifications. Probability bounds given for all the elementary events $E \in \mathcal{A}' \subset \Omega$ can be extended into bounds for arbitrary event $E \in \Omega$ by solving a linear optimization problem.

For the derived random variables $Y = f(X)$, we may calculate the imprecise probabilities that they will obtain a value in an element of their respective σ-algebras, similarly as in the precise case through Equation 3.2.

The imprecision in the distributions will also manifest in the imprecision in the expected values. We can calculate the lower and upper expectations, but for the structures introduced in this section, we can only do so through an optimization of the expected value over the credal set.

3.3.2 Capacities

F-probabilities offer good starting point for reasoning about probability bounds but provide only limited space for modelling and other manipulation of interval probabilities. Capacities originated in the work of Choquet [79] on the generalization of measure theory for non-additive measures. They provide further useful properties and structure to imprecise probability models which allows us to simplify specification of the bounds for arbitrary event $E \in \mathcal{A}$ and the expectations of random variables.

Definition 3.3.5 (Super-, Sub-additivity.)**.** Let $(\Omega, \mathcal{A})$ be a measurable space. A set function $g : \mathcal{A} \to \mathbb{R}$ is called a *capacity*, if it is monotone, i.e.

$$\forall A, B \in \mathcal{A} : A \subset B \Rightarrow g(A) \leq g(B). \tag{3.16}$$

A capacity g is further called *super-additive* if

$$\forall A, B \in \mathcal{A} : A \cap B = \emptyset \Rightarrow g(A \cup B) \geq g(A) + g(B). \tag{3.17}$$

If the inequality is reversed, it is instead called *sub-additive*.

Note that if the structure $\mathcal{M}$ of a F-probability is closed (i.e. arginf belongs to $\mathcal{M}$), then both L and U are super- and sub-additive capacities, respectively.

Definition 3.3.6 (N-monotonocity.)**.** A capacity g is said to be *n-monotone* if for any collection $E_n \subset \mathcal{A}$ of n elements

$$g\left(\bigcup_{E \in E_n} E\right) \geq \sum_{\mathcal{E} \subset E_n} (-1)^{|\mathcal{E}|+1} g\left(\bigcup_{E \in \mathcal{E}} E\right). \tag{3.18}$$

If g is monotone for every $n \in \mathbb{N}$, then it is called ∞-monotone.

Any n-monotone capacity is also $n > m$-monotone.

2−monotone capacities are coherent (as defined in Section 3.3.4). A pair of super- and sub-additive capacities, such that the sub-additive one dominates the super-additive one, constitute an F-probability.

Definition 3.3.7 (Möbius Inverse.)**.** For a super-additive capacity g defined on a *finite* space Ω, we define, for every event $E \subset \Omega$, a function $m_g : 2^\Omega \to \mathbb{R}$, the *Möbius inverse*, as

$$m_g(E) := \sum_{A \subset E} (-1)^{|E \setminus A|} g(A). \tag{3.19}$$

The benefit is that there exists an inverse mapping, which enables us to reconstruct the capacity from its Möbius inverse as

$$g(E) = \sum_{A \subset E} m_g(A). \tag{3.20}$$

The dual capacity, the upper probability, can be reconstructed as

$$g^*(E) = \sum_{\{A \in 2^\Omega : A \cap E \neq \emptyset\}} m_g(A). \tag{3.21}$$

A special class of models is composed of ∞-monotone lower probabilities on finite spaces. Their Möbius inverses (aka the *mass functions*) are non-negative for every event. Conversely, any normalized ($\sum m = 1$) non-negative function $m : 2^\Omega \to \mathbb{R}$ with a finite support induces a ∞-monotone lower and upper probabilities by Equation 3.20 and Equation 3.21, respectively. The mass function is explicitly used in the evidence theory [86] for basic probability assignments on the focal elements.

Let us assume, that we have a collection of interval-valued measurements: $\{(0.2, 0.6), (0.4, 0.8), (0.1, 0.3)\}$.

By the Laplace indifference principle, we assign to each of the interval an equal mass $m = \frac{1}{3}$. With such a mass function, we can construct lower and upper probabilities via Equation 3.20 and Equation 3.21. For example

$$\begin{aligned} L([0.3, 1]) &= m((0.4, 0.8)) = \frac{1}{3} \\ U([0.3, 1]) &= m((0.2, 0.6)) + m((0.4, 0.8)) = \frac{2}{3}. \end{aligned} \tag{3.22}$$

Capacities also provide us means for calculating not only the bounds on probabilities of events, but also on the expected values of random variables. If an F-probability is viewed as a pair of capacities, we can define an integration functional which will enable us to compute the bounds on the expected value, similar to which we are used to from the precise probability theory (Definition 3.2.4).

Definition 3.3.8 (Choquet integral)**.** For a capacity $g : \mathcal{A} \to \mathbb{R}$ and a real valued function f measurable on $\mathcal{A}$, the *Choquet integral* is defined as

$$(C)\int f dg = \int_0^\infty g(f \geq x)dx + \int_{-\infty}^0 (g(f \geq x) - 1)dx, \tag{3.23}$$

where $f \geq x$ denotes $\{t \in \Omega : f(t) \geq x\}$.

Theorem 3.3.1 (Lower expectation of a function via capacities)**.** *For a coherent 2-monotone lower probability $g : \mathcal{A} \to \mathbb{R}$, the lower expectation of a function f is given by the Choquet integral.*

$$\underline{P}(f) = (C)\int f dg. \tag{3.24}$$

Remark 3.3.1. Remark: if the capacity represents a 2-monotone upper probability, the integration would yield the upper expectation.

3.3.3 Random sets

A set-valued evidence may be encountered in many practical scenarios, may it be the error bounds of measuring devices or an interval valued expert elicitation. There exists an approach on how to handle these within precise probability theory in the case that we know that the imprecision is not inherent to the actual realisation of the experiment and only comes as a *coarsening* of precise values via our imperfect methods. In such a case, we may introduce an additional assumption on the stochastic nature of how the coarsening occurs, a conditional model on where the actual value lies in the set (e.g. as in the treatment of censored data in reliability theory). But in some cases, this assumption may be unjustifiable and bias our assessments. The imprecision may also be caused by the very nature of the of the experiment, where the random observation itself is set-valued and cannot be treated by the mentioned method. In order to rigorously address these situations, probability theory may be generalized for the set valued observations into theory of random sets [87–89].

Random set theory is a basis for Dempster-Shafer theory of evidence [86] and some modern statistical methods [90]. Random set models have also been used for sensitivity analysis [91] and uncertainty modelling in general [92].

Definition 3.3.9 (Random Set.)**.** Let $(\Omega, \mathcal{A}, P)$ be a probability space, $\mathcal{S}$ a collection of subsets of Ω_Φ, and $\Phi : \Omega \to \mathcal{S}$ a map.

If

$$\{\omega : \Phi(\omega) \cap K \neq \emptyset\} \in \mathcal{A}; \quad \forall \text{ compact } K \subset \Omega_\Phi, \tag{3.25}$$

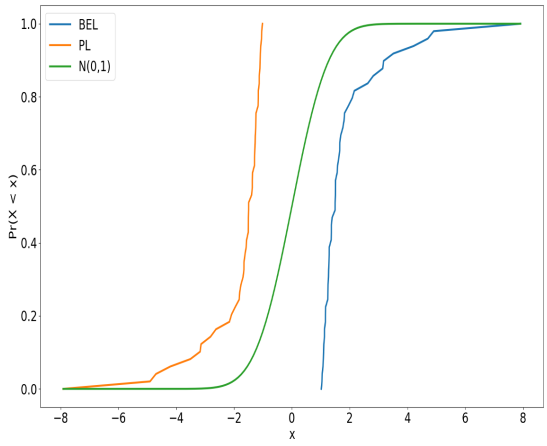
then we will call Φ a *random set*.

The definition of a random set is almost identical to that of a random variable (Definition 3.2.2) we just need to impose proper measurability properties. Nevertheless, the treatment of the random sets is slightly different. For a random set Φ, we can assess several claims.

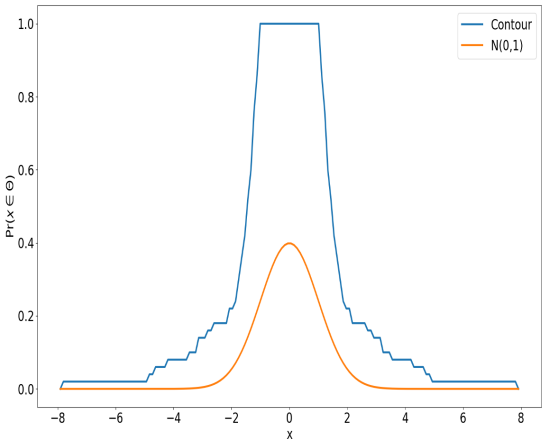
Definition 3.3.10 (Random Set Descriptors.)**.** Let $\Phi : \Omega \to \mathcal{S} \subset 2^{\Omega_\Phi}$ be a random set derived from probability space $(\Omega, \mathcal{A}, P)$. Then for $E \in \mathcal{S}$ and $x \in \Omega_\Phi$. We define:

- the *belief function* $Bel(E) := P(\Phi \subset E)$,
- the *plausibility function* $Pl(E) := P(\Phi \cap E \neq \emptyset)$,
- the *contour function* $C(x) := P(x \in \Phi)$.

The belief and plausibility functions corresponds to lower and upper probabilities and L and R functions from Section 3.3.1, respectively. As imprecise probability models, the belief and plausibility functions induced by random sets are $\infty-$monotone capacities and, therefore, coherent lower probabilities. Therefore, we also know the form for the derived lower and upper expectations via Theorem 3.3.1.



(a) Belief and Plausibility function of a Random set compared to the Cumulative Distribution function of Standard Normal distribution.



(b) Contour function of a Random set compared to the Probability Density function of Standard Normal distribution.

Figure 3.1: An example of a random set constructed from Chebyshev's inequality with $\mu = 0, \sigma = 1$.

Due to the coherency,

$$\begin{aligned} &\forall A \in \mathcal{B}: \\ &\qquad Bel(A) \leq P(A) \leq Pl(A) \\ &\qquad Bel(A) + Pl(A^c) = 1 \\ &\qquad Bel(A) + Bel(B) \leq Bel(A \cup B) \end{aligned} \tag{3.26}$$

Random sets can be used for statistical inference with little assumptions. The models can (and have been [91]) be constructed from Chebyshev's inequality (Theorem 3.3.2) if only the population mean and variance are know. This represents the tightest bounds for the respective probabilities over all possible probability distributions compliant to these assumptions. If the population mean and variance are unknown, Saw [93] has proposed a variant of the Chebyshev's inequality based on their sample estimates.

The construction from Chebyshev's inequality defines a random set induced by an uniformly distributed random variable, say $U \sim Uni([0, 1])$. The random sets, as models for a random variable with mean μ and variance σ^2, are constructed via

mapping $\Theta_C(u) := [\mu - \frac{\sigma}{\sqrt{u}}, \mu + \frac{\sigma}{\sqrt{u}}]$. An example of such a constructed random set is depicted in Figure 3.1.

Theorem 3.3.2 (Chebyshev inequality.)**.** *For a random variable X with finite expectation $\mu = \mathbb{E}(X)$ and finite non-zero variance $\sigma^2 = \mathbb{E}((X-\mu)^2)$ and $\forall a \in \mathbb{R}$: $a > 0$,*

$$P(|X - \mu| \geq a\sigma) \leq \frac{1}{a^2}. \tag{3.27}$$

Since they are analogical to precise probability theory, some of the useful results are also available for the theory of random sets. Especially variants of the law of large numbers (Theorem 3.2.2) and the central limit theorem (Theorem 3.2.3) can be generalized for random sets [89]. [94, 95] provide means on how to perform Monte Carlo simulation with random set models via constructing an *empirical* random set from drawn samples such that the approximations of the Bel and Pl functions are unbiased estimates and converge almost surely to the population ones.

3.3.4 Lower previsions

The expectation treatment can be also specified for imprecise models in the form of super-(and conjugated sub-)linear functionals in the theory of lower previsions. The theory was studied by Walley [81] with further extensive theoretical treatment given by Troffaes and de Cooman [75] and is build upon de Finetti's behavioral interpretation of uncertainty [73]. Random variables are therefore referred to as *gambles* in this line of work to emphasise that they represent prior uncertain rewards of bets and their combinations. Constructed lower previsions of gambles can then be interpreted as bounds on expected values of the respective random variables with respect to an underlying credal set - the set of precise probability models.

Properties of lower previsions are derived from certain rationality criteria for a gambler which aims to specify a (partial) preference relation amongst a set of gambles. These criteria specify how to construct reasonable subsets of the set of

all gambles which ensure that any combination of gambles ensure the gambler to at least *avoid sure loss* and *accept sure gain*. These minimal requirements ensure that a gamble (or combination of them) which would result in negative utility gain regardless of the outcome is deemed as not acceptable or desirable. Similarly, a gamble with only positive utility gain is always acceptable or desirable. Further criteria consider dominance and combination of gambles and allow construction of extensions of partially specified preference models - in this case a completion of a set of consistent acceptable gambles by all the gables that dominate them and their non-negative linear combinations.

The lower previsions for a gamble are then defined as the supremum price for which a gambler would be willing to pay for the gamble with uncertain reward.

Definition 3.3.11 (Lower Prevision.)**.** Given $\mathcal{G}$ the set of all gambles and $\mathcal{A} \subset \mathcal{G}$ a set of some consistent gambles (any non-negative linear combination of them avoids sure loss), their extension $\mathcal{D} := \{g \in \mathcal{G} : g \geq \sum_1^n \lambda_k f_k, n \in \mathbb{N}, \lambda \in \mathbb{R}_0^+, f \in \mathcal{A}\}$,

and the coherent lower prevision $\underline{P}(f) := sup\{\lambda \in \mathbb{R} : f - \lambda \in \mathcal{D}\}$,

Avoiding sure loss and *coherency* have special behavioral implications. From the gambling perspective, once a gambler specifies his previsions, he is assumed to accept any gable or their combination which he evaluates as acceptable. This becomes more obvious once we assume a casino or an insurance company in the gambler's position. *Avoiding sure loss* ensures that no combination of gambles would lead to negative gain irrespective of the actual realization of the random event. But only avoiding sure loss may lead to overly conservative assessments. Therefore *coherency* ensures that the bounds on prevision are minimal yet compatible with gambler's knowledge. From the mathematical perspective, coherency allows to specify the imprecise previsions as extreme points of some underlaying convex set of precise probability distributions (the credal set) and consequently evaluation of the lower and upper previsions by means of optimization over the credal set.

Remark 3.3.2. The conjugated upper prevision $\overline{P}(f) = -\underline{P}(-f)$.

The lower and upper probabilities for an event can be again constructed as previsions on its indicator function.

If the lower prevision is specified only for some gambles $f_1, ..., f_n$, it can be extended for arbitrary gambles as

$$\underline{P}(f) = sup\{\mu \in \mathbb{R} : f - \mu \geq \sum_{1}^{n} \lambda_k (f_k - \underline{P}(f_k)), \lambda_k \in \mathbb{R}_0^+\}$$

In the case that for each gamble its supremum buying price equals the infimum price for which the gambler would also sell the respective gamble, i.e. $\underline{P}(f) = \overline{P}(f)$, the prevision corresponds to de Finetti's linear previsions and such prices would represent their *fair prices*.

3.4 Methods of statistical inference

In this section, we will briefly revise the basic methods of statistical inference for data analysis. We will hereafter assume that we have a set of measurements $\mathbf{x} = \{x_1, ..., x_n\}$, independent and identically distributed (i.i.d.) samples, which were generated according to some precise *ground truth* distribution $\hat{P}$. Our intention is to construct probabilities of various events of interest based/conditioned on this dataset. Common practice is to construct an approximation (model) P of the sampling distribution $\hat{P}$ and estimate the desired probabilities from P.

A simple way of inferring probability distributions from a set of samples is given by *non-parametric methods*. Here, for an arbitrary event E, the probability is estimated as $P(E) = \frac{1}{n}\#\{x_i; x_i \in E\}$, the relative ratio of observations which comply with E. Distributions inferred in this way are usually labelled as empirical and constitute models with least additional assumptions. An example of an empirical CDF is depicted in Figure 3.2 with a label "empirical".

3.4.1 Inference about model parameters

Alternatives to the non-parametric methods search for an approximative distribution by *inverting* an apriori-selected set of low-dimensional sampling distributions, say $\mathbb{P} := \{P_\theta, \theta \in \Theta\}$. Based on the axiomatic theory of probability, precise methods mainly comprise of two methodologies - frequentist and Bayesian. Nevertheless, the common inference scheme for constructing distributional point-estimates, i.e. selecting a single *best-fitting* probability distribution, is simply this:

1. Choose (subjectively) a set of plausible sampling distributions $\mathbb{P}$.
2. Construct the *likelihood function* $\mathcal{L}(\theta; \mathbf{x})$, which models the probability of observing the collection $\mathbf{x}$ for each parameter θ.
3. Select $\hat{\theta}$ that best fits the observations and approximate $\hat{P}$ by $P = P_{\hat{\theta}}$ (the point estimate approach), or construct a mixture of distributions from the chosen family with mixing weights $w(\theta) \propto \mathcal{L}(\theta; \mathbf{x})\pi_0(\theta)$ and approximate $\hat{P}$ by $P = w(\theta)P_\theta$(the Bayesian approach; π_0 is called the prior distribution).
4. Evaluate the approximations of desired probabilities from inferred distribution P.

The data samples may come in various forms. Most commonly, they are considered precisely specified (e.g. real values for a random variables which obtains value on real line) in which case the likelihood function for inference from set of independent samples will take the form

$$\mathcal{L}(\theta; \mathbf{x}) = \prod_{i=1}^{n} f_\theta(x_i), \tag{3.28}$$

where f_θ is the probability density function of distributions from the chosen family $\mathbb{P}$ indexed by θ.

As mentioned earlier, the observations may also be imprecisely specified. There are multiple reasons which lead to this situation. For example, some observations may be censored or, in multivariate analysis, one of the covariates might be completely missing). Such observations may be included in the likelihood function via their

partial likelihood (which is the value of probability density function at observed point in case of precise observations in Equation 3.28):

$$\mathcal{L}_i(\theta; O_i) = Pr_\theta(X_i \in O_i), \tag{3.29}$$

where O_i denotes observed set value, and X_i RV representing i-th observation.

Example 3.4.1 (Maximum likelihood inference.)**.** Let us assume that we have a set of observations $\mathbf{x} := \{x_1, ..., x_N\}$ of a positive random variable X. We choose the set of admissible sampling distributions of the RV X to be the set of all exponential distributions ($F(x; \theta) = 1 - \exp(-\theta x)$). Using methods of statistical inference, we want to estimate the value of distribution parameter θ.

Frequently used frequentist method is the so-called *maximum likelihood estimation* (MLE) which has convenient asymptotic properties. Here we seek such value of θ which maximises the likelihood function Equation 3.28. Thus

$$\theta_{MLE} = \text{argmax}_{\theta \in \Theta} \quad \mathcal{L}(\theta; \mathbf{x}), \tag{3.30}$$

and construct $P = P_{\theta_{MLE}}$.

Resulting estimated distribution CDF is depicted in Figure 3.2 with label "MLE".

Example 3.4.2 (Bayesian inference.)**.** Let us assume the same scenario as in Example 3.4.1. We again select the set of admissible sampling distributions of the RV X, $\mathbb{P}$, to be the set of all exponential distributions. But now, we apply the Bayesian procedure, which results in a model capturing our uncertainty about the distribution parameter θ in a form of probability distribution in contrast to the single value estimated by MLE.

First, a *prior* distribution $\pi_0(d\lambda)$, which represents our knowledge about λ before observing the data, has to be elicited. Then, our knowledge is refined via the Bayes

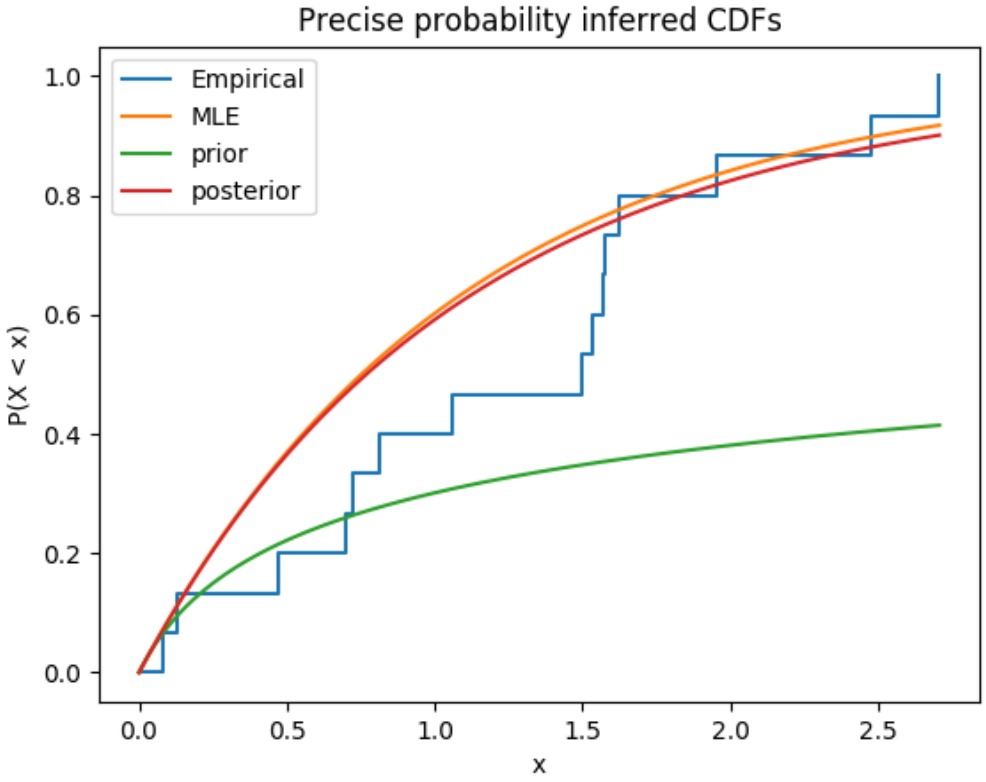


Figure 3.2: Example of precise probability inferences: an empirical distribution, a maximum likelihood estimate, and prior and posterior predictive distributions from the Bayesian inference.

updating rule to construct our *posterior* knowledge about θ as

$$w(\theta) = p(\theta|\mathbf{x}) \propto \mathcal{L}(\theta; \mathbf{x}) p_0(\theta). \tag{3.31}$$

Equation 3.31 specifies the posterior probability density function (the mixing weight), up to a normalization constant. From that, we can construct the predictive distribution for a future sample X_{n+1} as a weighted average of predictions of all the models in $\mathbb{P}$. Thus

$$p(x_{n+1}|\mathbf{x}) = \frac{\int_\theta p(x_{n+1}|\theta) p(\theta|\mathbf{x}) d\theta}{Z(\mathbf{x})}, \tag{3.32}$$

where $Z(\mathbf{x})$ is a normalization constant.

An example of the Bayesian inference is depicted in Figure 3.2 with CDFs labeled as the "prior" and the "posterior" for prior and posterior predictive distributions respectively.

For particular choices of families of likelihood functions, we can find a family of prior distributions which is closed under the Bayes' updating. This means that the posterior distribution lies in the same family, so we only need to update its parameters. We call these the *conjugated families*. For the particular choice of the exponential model for observed samples in Example 3.4.2, the conjugated family of distributions of θ are the Gamma distributions. This class also induces a closed form for the posterior predictive distribution (Equation 3.32). If no conjugate form can be found for the Bayesian inference, the problem needs to be solved numerically, usually by Monte Carlo algorithms [74].

3.4.2 Robust Bayesian inference

One application of IP theory was to provide means for sensitivity analysis for various decision making problems under uncertainty. In the case of Bayesian inference, it got labeled *Robust Bayesian analysis* [78]. In the Bayesian framework, we can analyse the sensitivity on both the prior distribution and/or the observation model (on the likelihood function). A straightforward solution is to consider sets of functions (priors and likelihoods) instead of just a single one in the analysis. The set of prior distributions would define an F-probability with the respective credal set. The credal set for the posterior F-probability would be given by the set of all updated prior distributions. All the assertions of interested would then be given by extremization over the posterior credal set (Equation 3.14).

Example 3.4.3 (Robust Bayes inference.)**.** Assume the same observations as in Example 3.4.1 and Example 3.4.2 and the same set of admissible sampling models, $\mathbb{P}$. This leads to the same likelihood function (Equation 3.28). Now, assume that we cannot properly specify one prior distribution for the Bayesian analysis as in Example 3.4.2. Instead, let us consider a set of prior distributions, again conjugated with our likelihood (i.e. Gamma distributions).

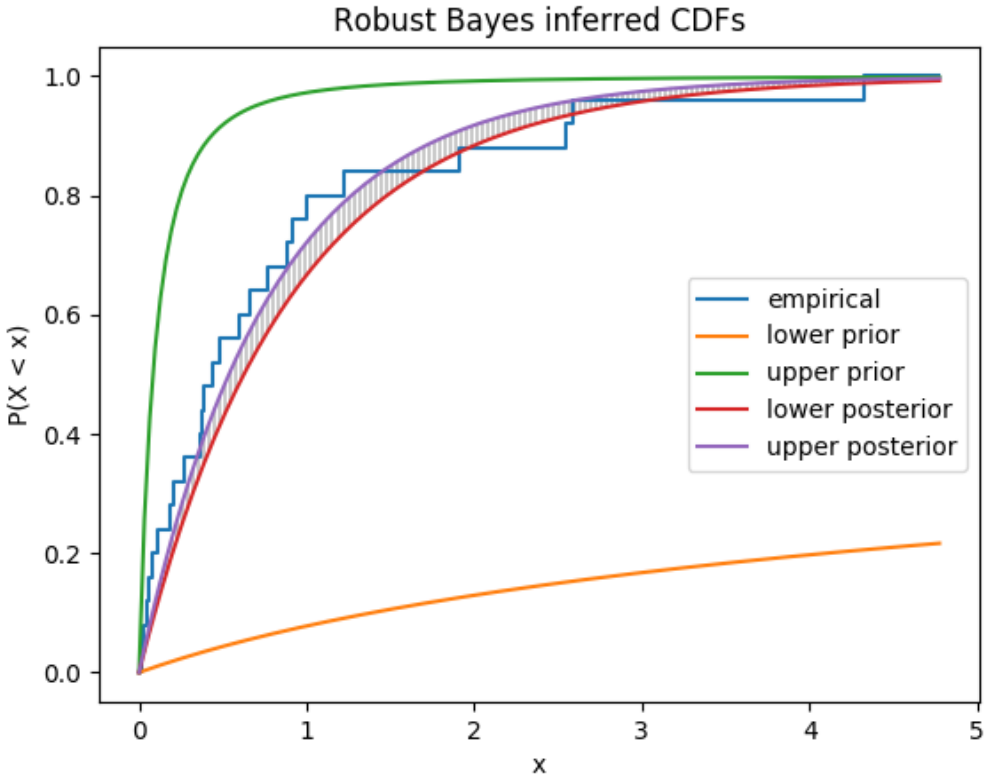


Figure 3.3: Bounds for prior and posterior predictive CDFs resulting from Robust Bayesian inference compared to the empirical distribution of samples.

For drawing inferences from the IP model, we need to consider answers from all the singular models in the credal set. In the case of reconstructing the predictive CDFs, we construct the bounds for all the CDFs from the set of all the updated prior distributions. Therefore

$$\underline{F}(x) := \min_{\pi \in \Pi_0} \frac{1}{Z_\pi(\mathbf{x})} \int_0^x \int_\theta p(x_{n+1}|\theta)\mathcal{L}(\theta;\mathbf{x})\pi(\theta)d\theta dx_{n+1}. \tag{3.33}$$

An example of a Robust Bayesian inference is depicted in Figure 3.3, where the lower and upper bounds are given for both the prior and the posterior predictive distributions.

A powerful result of the robust Bayesian inference is a closed form solution for the *imprecise dirichlet model* for inference about parameters of multinomial distribution [81]. The power lies in two features. First, the model is constructed for multinomial sample distributions. It can therefore be used for any inferential scenario with finite

set of outcomes, therefore also for inferences in general spaces after a finite grouping of its elements. The inferred parameters are the probability masses for the considered categories. The second feature is that the imprecise dirichlet model can model entirely vacuous prior previsions. This means that the prior probability bounds for an observation to be of arbitrary category is $[0, 1]$.

Compare this with Example 3.4.3, where even though we have used a set of prior distributions, the prior previsions of $P(X < 4)$ would be approximatelly $(0.2, 1) \neq (0, 1)$. Imprecise dirichlet model employs the ideal *non-informative* prior for Bayesian inference, which cannot be obtained as any precise probability distribution.

Robust Bayesian inference provides a tool for detecting and analysing conflict - the *inconsistencies* among observations and our prior distribution. It may occur, that our prior previsions about distribution parameters are off, in which case they will bias all the subsequent assessments and make them not reliable. Sometimes, we can interpret one of the parameters of prior distributions proportional to our *confidence* about the prior judgement. By choosing the set of priors such that they include distributions with various levels of this confidence, we can observe how the actual observations change our predictions w.r.t. these levels. If there is no conflict between our prior assessments and the observations, the imprecision in answers of the model tends to decrease with increasing sample size and the predictive sets will be nested if ordered by the confidence parameter. But if there is a conflict, the imprecision may increase and we may observe a *shift* in the predicted values - loss of the nestnessness property. The consistency can be analyzed graphically by plotting the predictions against the parameter for the confidence in prior [44]. An example is show in Figure 3.4 for the inference about the success probability in a Bernoulli trial with a set of Beta prior distributions which is a class which possesses this *confidence* parameter.

Example 3.4.4 (Conflicting evidence.)**.** Let us consider a collection of i.i.d. Bernoulli trials (realization of one trial is either 1 or 0, or success or failure). The distribution of number of successes s in n trials can be described by parametric Binomial distribution

$P(S = s) = \binom{n}{s}\theta^s(1-\theta)^{n-s}$, where parameter θ represents probability of success in singular trial.

For Bayesian updating, there exist a conjugated family of distributions of θ for Binomial observation model, the Beta distribution family, with PDF:

$$p(\theta|y,m) = \frac{\theta^{ym-1}(1-\theta)^{m-ym-1}}{B(ym, m-ym)}, \tag{3.34}$$

where y represents "point estimate" of θ, m can be seen as the "confidence measure" mentioned earlier, and $B(.,.)$ is the Beta function.

Bayesian updating (Equation 3.31) of parameters y, m after observing s successes in n trials is given by:

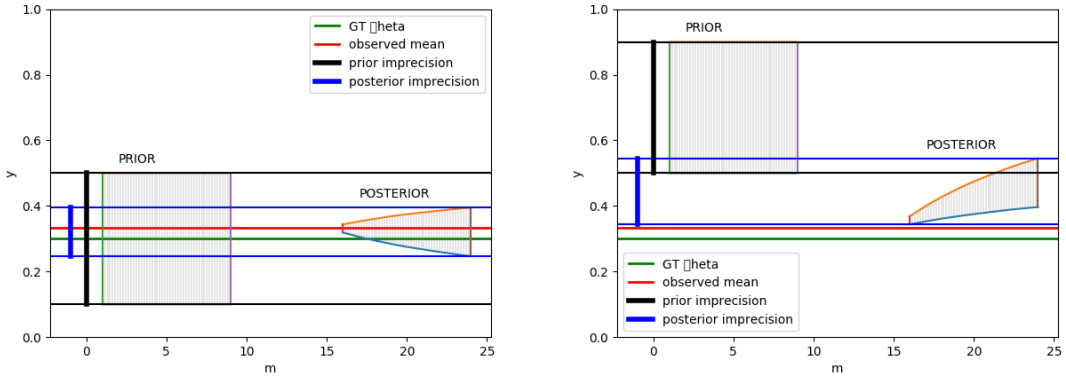
$$m_n = m_0 + n, \qquad y_n = \frac{m_0 y_0}{m_0 + n} + \frac{s}{m_n + n}, \tag{3.35}$$

where subscripts n and 0 denote updated and prior beta parameters respectively.

We can distinguish two basic scenarios w.r.t the relation of "consistence" of prior previsions and observations. For the same, randomly generated observation (n, s), transformation of the set of prior beta distribution parameters into a set of posterior beta distribution parameters is depicted in Figure 3.4 for cases with and without prior-data conflict.

In case of no conflict among the two information sources, we can observe a collection of nested previsions of interest (bayesian optimal point estimates of posterior binomial parameter θ, i.e. y) in the direction of increasing confidence in prior (increasing m), whereas in case of conflicting evidence, a shifting trend may be observed in the same direction. The conflict also results into greater posterior imprecisions compared to the inference in non-conflicting information situation.

Sensitivity analysis, similarly to the robust Bayesian statistics, was also explored in the frequentist framework. Frequentist induction, with precise probabilities, assumes that there exist a precise sampling distribution from which the i.i.d. observations



(a) Without conflicting information. (b) With conflicting information.

Figure 3.4: Comparison of prior and posterior sets of parameters of beta distribution - mean value y and (pseudo)count m. The box labeled 'PRIOR' represents chosen set of prior parameters for the Robust Bayesian inference. The shape labeled 'POSTERIOR' represents imprecise posterior distribution constucted as a collection of precise Bayesian posteriors when applied on individual pairs of values from the 'PRIOR'. 'GT' represents the truth mean value used to sample the 'observed mean'. Prior and posterior imprecision refers to imprecision in the mean value parameter y.

are generated. This assumption may be weakened, as has been done in [96], where a possibility of imprecise sampling distributions and desirable subsequent properties of imprecise frequentist inference. Strong motivation for this extension is theoretical impossibility of observing *identically* distributed samples due to variability in experimental setting (although it might be negligible). Their approach include precise distributions as a special case and in the case of a precise sampling process, the imprecisions in the inferred distributions converge to zero - the inferred IP law converges to a precise law.

3.4.3 Fiducial inference

Another important considerations of frequentist inference procedures lies in the validity principle - e.g. bounding the type I error in hypothesis testing and proper calibration of confidence intervals. Generally we want to ensure that "It's unlikely that the procedure leads to incorrect conclusions". If our statistical procedure constructs a system of nested confidence sets indexed by their coverage probability

α, we require that

$$P_{X|\theta}(\theta \in C_X(1-\alpha)) \geq 1-\alpha. \tag{3.36}$$

Bayesian procedures can only comply with these asymptotically and can be severely biased by the information supplied through the prior distribution in the cases when only small number of observations is available. On the other hand, the results of Bayesian procedures, the posterior distributions, can straightforwardly be propagated to obtain assertions about derived quantities, $f(X)$.

The aim of fiducial inference is to construct posterior distributions, similar to Bayesian inference, but without the need to specifying prior distribution which can bias the inference. The problem of obtaining distributional estimate in the frequentist framework was introduced by Fisher in his work on *fiducial inference* [97] which later inspired the development of the standard theory of confidence intervals and Neyman-Pearson hypothesis testing procedures. Alternative fiducial procedures were introduced by Fraser [98] who studied statistical inference from the perspective of noisy measurement of a precise quantity. This allowed to seamlessly shift the uncertainty model on the parameter space and provide probabilistic statements about the quantity in question. Eventually, the approach was developed into a mature theory allowing to preform prior-free inference with distibutional estimates. [99–101] formulated the inference for general distribution parameters where modelling the sampling process as a transformation of an ancillary random variable with known distribution. Using a copy of the ancillary variable allows us to construct posterior fiducial distribution modelling uncertainty about the unknown model parameters. Confidence intervals can be constructed from such posteriors in the same way as credible set from Bayesian posterior distributions. Additional information can be included in the inference using Bayes rule.

Further advancements in fiducial inference were enabled by Dempster [80, 102] by introducing set-valued inversions of the sampling model. Such ideas were further

developed into the **evidence theory** by Dempster and Shafer [86]. A review of the development, properties and some applications of fiducial inference theory is given by Hannig in [103].

3.4.3.1 Basic fiducial inference procedure

Generally, fiducial inference can be carried by inverting the structural equation. This equation provides an alternative to the likelihood function representation from Bayesian or Fisher's approach. Assume that we can write the observed variable as a (parametrized) transformation of a random variable with known distribution.

$$X = A(\theta, U), \qquad U \sim P_U \tag{3.37}$$

The fiducial inverse is then taken as the inversion of Equation 3.37 sampled through an independent copy of U, $U^* \sim P_U$.

$$\begin{aligned} \Theta &= G_x(U^*) | G_x(U^*) \neq \emptyset \\ G_x(u) &:= \{\theta : x = A(\theta, u)\}. \end{aligned} \tag{3.38}$$

The fiducial procedure using structural equation can be summarized as follows. Given a random variable X:

1. Construct the sampling model $X = A(\theta, U)$
   - where U is an ancillary RV with known distribution,
   - and θ are model parameters.
2. Construct fiducial inversion $\Theta(x, u) := \{\theta : x = A(\theta, u)\}$.
3. If X is known, and U would be observed, Θ represents *plausible* model parameters.
4. Observation x and the distribution of U induce a (random set) model for θ.
5. If Θ is singleton, it is equivalent to fiducial inference with precise posterior distribution.

3.4.3.2 Considerations about generalization

When the dimension of U and Θ are the same, the inversion can usually be constructed, although we may need to use numerical methods. When the dimensions of the ancillary random variable U is higher that that of Θ, the inversion may result in an empty set with probability 1. Fraser solved this problem for a class of sampling models by dimension-reduction. Hannig presents a general methodology which indroduces conditioning the posterior on existence of the inversion (since the existence of observations implies the existence of fiducial inversion) [103]. The observation space of N random variables is divided into low-dimensional orbits specified by e.g. sufficient statistics as in the above case of the normal sampling model.

Generally, the structural model can always be constructed from the likelihood function. In the case of i.i.d. observations, the structural equation can be specified by parametrized quantile functions and the ancillary random variables are i.i.d. uniform:

$$\begin{aligned} X_i = A(\Theta, U_i) = F_\Theta^{-1}(U_i), \qquad U_i \sim Unif(0,1) \\ \Theta = G_x(U_1^*, \ldots, U_n^*) | G_x(U_1^*, \ldots, U_n^*) \neq \emptyset \end{aligned} \tag{3.39}$$

3.4.3.3 Examples of Fiducial inference

Example 3.4.5 (Fiducial inference for Gaussian random variable.)**.** For a set of n i.i.d. observations from the Gaussian distribution with sample mean $\langle X \rangle$:

$$\begin{aligned} X_i &\sim \mathcal{N}(\mu, \sigma) \\ \sqrt{n}\left(\frac{\mu - \langle X \rangle}{\sigma}\right) &\sim \mathcal{N}(0,1) \\ \sqrt{n}\left(\frac{\mu - \langle X \rangle}{\sigma}\right) &= U, \end{aligned} \tag{3.40}$$

where U is the ancillary random variable from Equation 3.37 with standard normal

distribution. Considering that the variance σ is know and the mean value μ is the parameter of inference interest, we can invert the Gaussian model algebraically to

$$\mu = U\frac{\sigma}{\sqrt{n}} + \langle X \rangle. \tag{3.41}$$

This inversion in Equation 3.41 corresponds to the usual procedure used to derive confidence intervals for the mean of Gaussian distribution with known variance:

$$\begin{aligned} \text{frequency:} \quad & P_{X|\mu}\left(\sqrt{n}\left(\frac{\mu - \langle X \rangle}{\sigma}\right) > a\right) = q_a \\ \text{confidence:} \quad & P_{X|\mu}\left(\mu > \frac{a\sigma}{\sqrt{n}} + \langle X \rangle\right) = q_a \end{aligned} \tag{3.42}$$

where q_a is the quantile of ancillary random variable $A \sim N(0,1)$.

Example 3.4.6 (Fiducial binomial inference)**.** Let us have samples $y_1, y_2 \sim F$, where F is a general distribution. Define random variable $X = \sum_{y_i<\theta} 1$, where θ represents the median of distribution F and an ancillary random variable $U \sim Bi(2, 0.5)$ The fiducial inversion can be constructed even though the random variable X is unobserved using Equation 3.38:

$$\Theta = \left\{\theta : \sum_{y_i<\theta} 1 = U\right\}, \tag{3.43}$$

which can be expressed explicitly as:

$$\Theta((y_1, y_2), u) = \begin{cases} (-\infty, y_{(1)}) & , u = 0 \\ [y_{(1)}, y_{(2)}) & , u = 1 \\ [y_{(2)}, \infty) & , u = 2 \end{cases} \tag{3.44}$$

Thus, the induced random set model is:

$$\Theta \sim \begin{cases} (-\infty, y_{(1)}) & , p = 0.25 \\ [y_{(1)}, y_{(2)}) & , p = 0.5 \\ [y_{(2)}, \infty) & , p = 0.25 \end{cases} \tag{3.45}$$

The inferred posterior is imprecise and can be used to derive belief and plausibility functions (Definition 3.3.10):

$$\begin{aligned} &\forall B \in \mathcal{B}_\theta : \\ &\qquad Bel_x(\theta \in B) = P_U(u : \Theta((y_1, y_2), u) \subset B) \\ &\qquad Pl_x(\theta \in B) = P_U(u : \Theta((y_1, y_2), u) \bigcap B \neq \emptyset) \end{aligned} \tag{3.46}$$

3.4.4 Nonparametric predictive inference

Another example of a fiducial method is **non-parametric predictive inference** (NPI) [104]. The method assumes exchangeability of the observations and bases its indifference principle on Hill's assumption about posterior distribution of percentiles [105] - indifference among all possible orderings.

After observing N real valued observations $\mathbf{x} = \{x_1, ..., x_N\}$, NPI constructs a predictive random set for the next observation X_{N+1} by placing mass $1/(N+1)$ on each of the intervals $(x_i, x_{i+1}), i = 0...N$, where x_0, x_{N+1} are some bounds of the X_{N+1} support (possibly infinite). The corresponding lower and upper probabilities may be derived by Equation 3.20 and Equation 3.21. It was shown that the result of the inference is an ∞-monotone capacity and an F-probability [106].

The procedure may continue to derive predictive distribution about multiple future observation by extending the Hill's assumption.

NPI has also been extended to include situations with censored observations without the need of including any restrictive censoring assumptions [107, 108] and are well

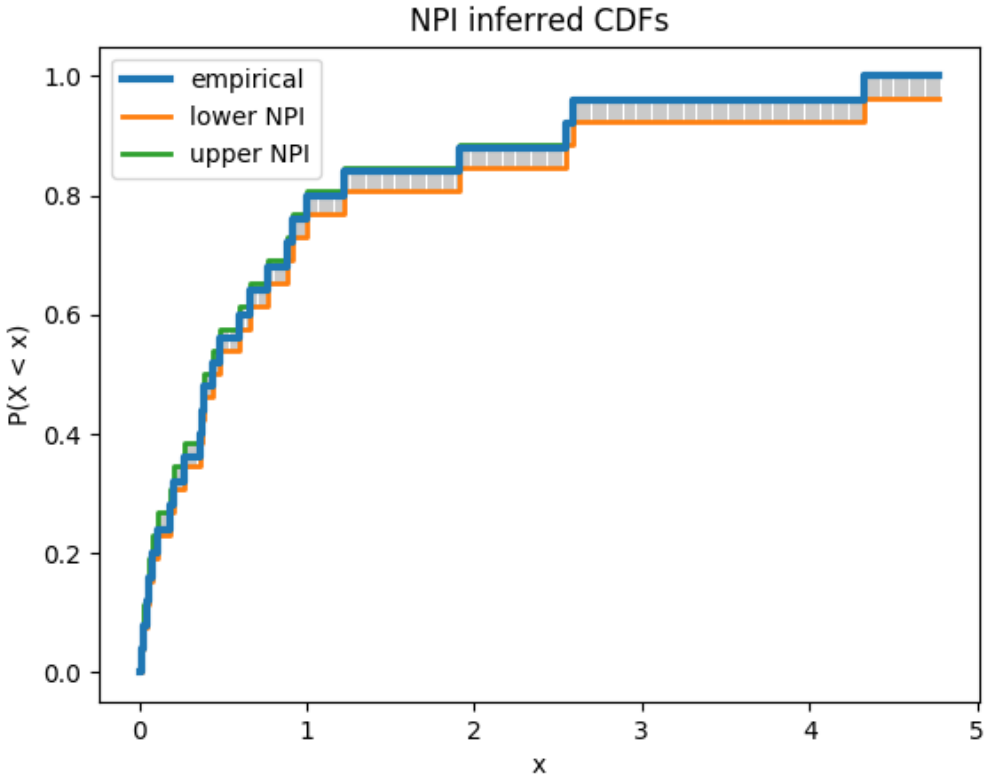


Figure 3.5: Lower and upper cumulative distribution function obtained via NPI.

suited for applications together with survival signatures [42] or survival analysis in general [109].

An example of NPI lower and upper CDFs is show in Figure 3.5 compared to the empirical distribution.

A weaker analogue of non-parametric predictive inference can be derived using the fiducial argument. The difference is that the fiducial argument assumes independence of observations instead of the weaker property of exchangeability. The fiducial inference about distribution median (Example 3.4.6) can be extended to general quantiles. Consider N i.i.d. ordered observations, $x_1, ..., x_N$, including extra bounds x_0, x_{N+1}. The fiducial posterior on quantile q place probability mass on individual intervals $[x_k, x_{k+1}]$ equal to binomial probability mass function $\binom{N}{k} q^k (1-q)^{N-k}$. Distribution of quantiles can be used to construct a predictive inference using the inverse transformation sampling:

$$X_{N+1} = Q(A), \qquad A \sim Unif(0,1) \tag{3.47}$$

where $Q(A)$ is the fiducial posterior on quantile q_A.

Theorem 3.4.1 (Fiducial predictive is NPI.)**.**

$$P(X_{N+1} \in [x_k, x_{k+1}]) = \frac{1}{N+1}. \tag{3.48}$$

Of Theorem 3.4.1. For a fixed quantile q and any of the intervals $[x_k, x_{k+1}]$, the fiducial posterior probability quantile q belonging to the interval is given by binomial probability mass function. The total posterior mass placed on the interval can be expressed as the expected value over the randomly sampled quantiles.

$$P(X_{N+1} \in [x_k, x_{k+1}]) = \int_0^1 \binom{N}{k} q^k (1-q)^{N-k} dq. \tag{3.49}$$

For $k \in 0, N$, the integrand becomes a simple polynomial function and the integral equals to $\frac{1}{N+1}$.

For $k \in \{0..N\}$ all the integrals are equal since, using per partes integration,

$$\begin{aligned} I[N,k] &:= \int_0^1 \binom{N}{k} q^k (1-q)^{N-k} dq \\ &= \left[\binom{N}{k} \frac{1}{k+1} q^{k+1} (1-q)^{N-k} \right]_0^1 - \quad (-1)\binom{N}{k} \frac{N-k}{k+1} \int_0^1 q^{k+1} (1-q)^{N-k-1} dq \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad + \binom{N}{k+1} \int_0^1 q^{k+1} (1-q)^{N-k-1} dq \\ &= I[N, k+1] \end{aligned} \tag{3.50}$$

Hence the predictive mass placed on each interval $[x_k, x_{k+1}], k \in 0..N$ is $\frac{1}{N+1}$. □

3.4.5 Inferential models

Specifying uncertainties with precise probability models has more subtle issues. [70] shown how a precise probability model can lead to paradoxical situation in risk assessment, where increasing uncertainty about observation lead to decreasing calculated risk when computing probability of satellite collisions. The inferences which lead to precise posterior distribution turn out to fail to satisfy the frequentist calibration properties in non-trivial cases [110].

The Fundamental Frequentist's Principle:

"It's unlikely that the procedure leads to incorrect conclusions".

When constructing posterior predictive intervals C_X on a chosen significance level α for an uncertain value, we would like to comply with the **coverage property**:

$$P_{X|\theta}(\theta \in C_X(1-\alpha)) \geq 1-\alpha \tag{3.51}$$

Martin and Liu [90] presented **inferential models** - a statistical framework built on Fraser's structural models, but replacing the copy of the ancillary variable by a random set with superior calibration properties. The idea for reliably capturing a value of the predicted random variable is to construct a random set which covers it with reasonable probability. The validity of inferences can be obtained by using this principle on the fiducial ancillary variable A, such that

$$P(S \ni A) \geq_{st} Unif(0,1), \tag{3.52}$$

where S is the *valid predictive random set* for A.

This leads to construction of random set structures, which can be used to obtain valid confidence intervals on any level of significance and to carry out hypothesis testing for arbitrary derived assertions. The result of these inferences are generally random sets, belief and plausibility functions of which can be used to bound the

inferences about the investigated random variable. This leads to random set posterior distributions, which are *valid*, leading to desirable properties of the derived belief and plausibility functions:

$$\begin{gathered} \forall \theta \in \Omega_\theta, B \in \mathcal{B}_\theta, \alpha \in [0,1] \\ \sup_{\theta \notin B} P_{X|\theta}(\{Bel_x(\theta \in B) \geq 1-\alpha\}) \leq \alpha, \\ \sup_{\theta \in B} P_{X|\theta}(\{Pl_x(\theta \in B) \leq \alpha\}) \leq \alpha, \end{gathered} \tag{3.53}$$

Such posterior random sets can be used for construction of hypothesis testing procedures. For given $H_0 : \theta \in B$, reject H_0 if $Pl_x(\theta \in B) \leq \alpha$.

Besides of that, inferential models allow us to develop derived methods for situations with additional knowledge, propagate the resulting random sets to obtain assessments about derived quantities, and naturally analyse imprecise observations without additional modelling assumptions.

Chapter 4

System reliability assessment examples

4.1 Computing remaining useful life of a system

This section introduces a method to estimate residual lifetime with knowledge of the number of functioning components. It can be applied in situations where we directly observe the number of functioning components or when we model a hidden process of the number of functioning components based on some proxy observations (power input, waste output, system performance, etc.).

We will assume the following properties of the system:

- Lifetimes of components of the same type are exchangeable
- Lifetimes of components of different types are independent
- The system is coherent (the structure function is non-decreasing)
- The components are non-repairable (their state processes are non-increasing)

Note that the last two imply that: The system state process is non-increasing.

The residual lifetime, conditioned only on the knowledge that the system is functional

at time t, can be computed as

$$\begin{aligned} P(S_t|S_0) &= \frac{P(S_t \cap S_0)}{P(S_0)} \\ &\overset{S_0 \subseteq S_t}{=} \frac{P(S_t)}{P(S_0)} \\ &= \frac{\sum_l \Phi(l) P(L_t = l)}{\sum_l \Phi(l) P(L_0 = l)} \end{aligned} \tag{4.1}$$

We will denote S_t the event that the system is functional at time t and L representing the number of functioning components. In the following derivations, we will assume only one component type. The extension to systems with multiple component types is possible by assumming L as a vector representing the number of functioning components of respective types.

$$\begin{aligned} P(S_t|S_0, L_0 = k) &= \frac{P(S_t, S_0|L_0 = k)}{P(S_0|L_0 = k)}, \\ &\overset{S_0 \subseteq S_t}{=} \frac{P(S_t|L_0 = k)}{P(S_0|L_0 = k)}, \end{aligned} \tag{4.2}$$

where $P(S_0|L_0 = k) = \Phi(k)$, the survival signature for k functioning components.

Next, we derive the conditional probability $P(S_t|L_0)$. Denote $D_X = \{0,1\}^N$ the domain of component states, $D_L = [0, N]$ the domain of number of functioning components $L(x)$ and $D_X(k) = \{x \in D_X : L(x) = k\}$ the set of all combinations of component states such that exactly k components are functional.

$$\begin{aligned} &P(S_t|L_0 = k) = \\ &= \sum_{x \in D_x} P(S_t|X_t = x, L_0 = k) P(X_t = x|L_0 = k) \\ &= \sum_{x \in D_x} \varphi(x) P(X_t = x|L_0 = k) \\ &= \sum_{x \in D_x} \varphi(x) \sum_{y \in D_x} P(X_t = x|X_0 = y, L_0 = k) P(X_0 = y|L_0 = k) \end{aligned} \tag{4.3}$$

Due to the exchangeability assumption,

$$P(X_0 = y|L_0 = k) = \delta_{(L(y)=k)} \frac{1}{|D_X(k)|} = \delta_{(L(y)=k)} \frac{1}{\binom{N}{k}}, \tag{4.4}$$

therefore we sum only $y \in D_X(k)$.

The exchangeability assumption can be used to calculate the transition probability $P(X_t = x|X_0 = y, L_0 = k)$. If $x \leq y$, $L(y) = L_0 = k$ and $L(x) = l$, then the transition represents failure of $k - l$ and survival of l components, which is equal for any x, y and depends only on l, k.

Considering $L(X_t) = l$, denote $P_t(k \to l) := P(L(X_t) = l|L_0 = k)$ the transition probability from l to k functioning components in time t. With an additional i.i.d. assumption on component failure times, $P_t(k \to l) = R(t)^l F(t)^{k-l}$ representing survival of l and failure of $k-l$ components. This transition probability $P_t(k \to l) = 0$ for $l > k$ due to the non-repairability assumption.

$$P(S_t|L_0 = k) = \frac{1}{\binom{N}{k}} \sum_{l \in D_L} \binom{N-l}{k-l} P_t(k \to l) \sum_{x \in D_x(l)} \varphi(x) \tag{4.5}$$

The conditional residual lifetime can be calculated as:

$$P(S_t|S_0, L_0 = k) = \frac{\sum_{l \in D_L} \binom{N-l}{k-l} P_t(y_k \to x_l) \mathcal{N}(l)}{\mathcal{N}(k)}, \tag{4.6}$$

where $\mathcal{N}(l)$ is the number of distinct states $x \in D_X(l)$ s.t. $\varphi(x) = 1$. The $\mathcal{N}(l)$ is equivalent to the counting signature introduced in Definition 4.3.1, so $\Phi(l) = \frac{\mathcal{N}(l)}{|D_X(l)|}$.

Substituting for $\mathcal{N}$ in Equation 4.6, we get a representation in terms of survival signatures:

$$P(S_t|S_0, L_0 = k) = \frac{\sum_{l \in D_L} \binom{N-l}{k-l} \binom{N}{l} \Phi(l) P_t(y_k \to x_l)}{\Phi(k)|D_X(k)|}. \tag{4.7}$$

The procedure is demonstrated on a simple example of a 5-component bridge system with single component type with independent exponential lifetimes with mean time

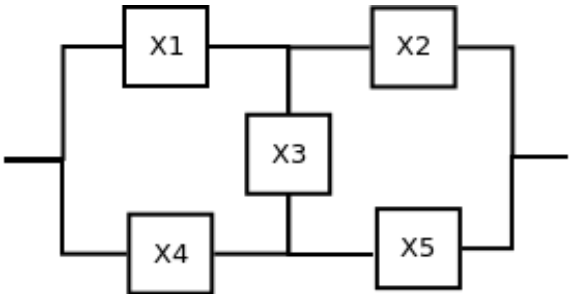


Figure 4.1: RBD of a bridge system used in the demonstration example.

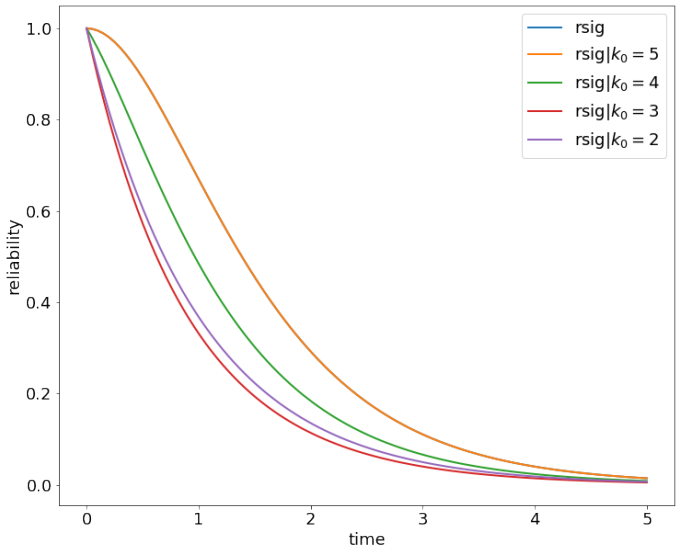


Figure 4.2: Survival functions representing the time to system failure based on the number of functioning components. Standard method for computing the survival function assumes that all the components are functional at the initial time, therefore it overlaps with the $k_0 = 5$ curve.

to failure of 2 time units. The RBD of the system is depicted in Figure 4.1. Survival functions of the remaining life are depicted in Figure 4.2.

In conclusions, residual life can be inferred with no additional demands on computation of any special survival signatures when the system state process is non-increasing. This is a benefit since complexity such computations are exponential. There are possible applications for condition monitoring based on observation of the number of functioning components or an ancillery variable with number of functioning components modelled as a hidden process. These results are not applicable to cases with repairable components because of the requirement that the event $S_0 \subset S_t$ in Equation 4.2.

4.2 Phased mission analysis with survival signatures

This section is an exerpt of a conference paper submitted for SECESA 2018, [1].

Sometimes, a system may be required to perform multiple tasks in a sequence, like an airplane, which has to (at least) take-off, cruise, and land again. Such situations are denoted as Phased Missions and are considered successful if the system successfully completes all the phases. Generally, in each of the phases, different conditions and failure modes may apply. Assessing the mission reliability requires us to merge individual specifications of all the phases into a specification of the whole mission.

The method for computation phased mission reliability using survival signatures was introduced in [52]. This section describes how phased missions can be modelled using survival signatures and investigates how the ordering of the phases influence mission reliability.

4.2.1 Phased mission reliability

For an arbitrary system, its reliability is defined as the probability that it will complete its mission. If we require a system to stay functioning for its whole mission time T_M, the reliability can be calculated by assessing the probability of the event:

$$X_M := \left\{ \inf_{t \geq 0} \{t : X_S(t) = 0\} > T_M \right\}, \tag{4.8}$$

where X_S corresponds to a random variable (RV) representing the system state with $X_S = 0$ meaning that the system is not functioning.

A system is a device composed of distinctive subsystems. The reliability of each of the system components may be assessed individually. To assess the reliability of a system, it is necessary to model dependencies between states of the components and the states of the system.

For a binary system S (i.e. we only distinguish two states - functioning and failed) composed of N binary components, this dependency may be described by a structure function $\varphi_S : \{0,1\}^N \to \{0,1\}$. With dependency modelled by a structure function and states of components being random, the probability of the system functioning may be obtained by calculating the expected value of its structure function:

$$P(X_S = 1) = \mathbb{E}\{\varphi(X_1, \ldots, X_N)\}, \tag{4.9}$$

where $X_S \in \{0,1\}$ represents the state of the system and $X_i \in \{0,1\}$ the state of component i.

The component state can evolve in time and in that case we need to provide a stochastic model for its state at any time point t considered, i.e. describe a stochastic process of component state $\{X_i(t)\}|_t$. Having this description for each of the system components, we may describe the stochastic evolution of the system state. For each time t, the probability of the system functioning can be obtained by calculating the expectation in Equation 4.9 for (random) component states at time t.

If we assume that a component is non-repairable, i.e. it remains failed once failed, description of the stochastic process of its state simplifies greatly. In such case, it suffices to model time to failure (TTF) of a component as a positive random variable, say T_i. The probability that a component functions at an arbitrary positive time t is then equivalent to the probability that the component survived up to time t, i.e. probability of event $\{T_i > t\}$. This probability is modelled by a survival function $R_i(t) := P(T_i > t)$. In this section, we will consider systems with non-repairable components.

4.2.2 Survival signatures

Evaluation of Equation 4.9 requires us to sum over the set of all possible component states, the cardinality of which increases exponentially with the number of components. Survival signatures, described in Section 2.3 can decrease the complexity.

For system a composed of K distinct types with exchangeable failure types in each group and total amount of M_j component of type j in the system, distribution of system failure time can be expressed via its survival signature as:

$$P(T_S > t) = \sum_{l_1=0}^{M_1} \dots \sum_{l_K=0}^{M_K} \Phi_S(\mathbf{l}) P(L(t) = \mathbf{l}), \tag{4.10}$$

where

$$P(L(t) = \mathbf{l}) = \prod_{j=1}^{K} \left[\binom{M_j}{l_j} [R_j(t)]^{l_j} [1 - R_j(x)]^{M_j - l_j} \right], \tag{4.11}$$

where $R_j(.)$ denotes the survival function common to components of type j.

Once we can model a system solely by modelling its TTF with a survival function $R_S(.)$, the event describing successful mission completion can be simplified to an equivalent event $X_M = \{T_S > T_M\}$. Mission reliability can then be assessed as:

$$Rel_S = R_S(T_M). \tag{4.12}$$

A phased mission consists of a series of W phases. The whole system mission time T_M may be divided into intervals $[\tau_{i-1}, \tau_i]$, for $i = 1, \dots, W$ representing time frames of the mission phases, with $\tau_0 = 0$. For each of the phases we have to consider that:

- different components might be used in different ways, which can be modelled by separate structure functions for each of the phases, and
- components might be subjected to different conditions, which affects their deterioration rates and, therefore, failure probabilities.

A phased mission is considered successful if it successfully completes all its phases.

We can reflect this by constructing a structure function for the whole mission. Taking the time evolution of (non-repairable) component states into account, the structure function of a mission can be expressed as:

$$\varphi_{PM}(x(\tau_1), \ldots, x(\tau_W)) = \prod_{i=1}^{W} \varphi_i(x(\tau_i)), \tag{4.13}$$

where φ_i is structure function describing requirements on phase i and vector $x(\tau_i)$ represents states of components at the end of phase i.

To simplify PM reliability computation using survival signatures, minor adjustments have to be considered.

Phased mission state space

The PM structure function is a function of component states at the end of the phases. This means that every component may be functional or not at each of phase end. Care must be taken to include our assumption of non-repairability of the components, i.e. to exclude states for which some component is non-functional at one time and functional later. The new component state evolution space, Ω_X, will be a subset of $\{0,1\}^{K \cdot W}$ which excludes these.

Component grouping

Grouping of components into types with exchangeable lifetimes have to take into account whether components of the group are stressed in exactly the same way throughout the whole mission. In simplified case, where all the components are stressed in all the phases, we may employ the same discrimination as in single phased systems, i.e. grouping them by they physical type. Further discussion on component grouping is provided in [52].

State space decomposition

The auxiliary RV L introduced via the law of total expectation for convenient state space decomposition, will just change form reflecting the change of state space and component grouping. It is a vector of values which represent the number of functioning components in a group, now also for the end of each phase. The dimension of the vector will therefore be $K \cdot W$, where K is the number of component groups and W the number of phases. The ordering can be made arbitrary, e.g.:

$$\mathbf{l} = (\mathbf{l}_1, \ldots, \mathbf{l}_W), \tag{4.14}$$

where $\mathbf{l}_i$ is a K-dimensional vector, elements of which represents number of respective functioning components at the end of phase i.

Taking the changes introduced earlier, the survival signature is defined exactly in the same way. We only need to reflect the change of Ω_X and what does it mean that l holds (i.e. x is such that l_{ij} components of type j function at the end of phase i).

The PM reliability can be computed as the expectation of the structure function for RV $\mathbf{X} := (X(\tau_1), \ldots, X(\tau_W))$. We can decompose the expectation via law of total expectation through the augmented auxiliary RV L as:

$$\begin{aligned} Rel_{PM} &:= \mathbb{E}\varphi_{PM}(\mathbf{X}) \\ &= \sum_{l_1=0}^{M_1} \ldots \sum_{l_{KW}=0}^{M_{KW}} \Phi_{PM}(\mathbf{l}) P(\mathbf{L} = \mathbf{l}), \end{aligned} \tag{4.15}$$

where the mixing probability $P(\mathbf{L} = \mathbf{l})$ represents the probability that $l_i^j - l_{i-1}^j$ failures of component of group j occur in phase i. In the case where all the components are present in all the phases, it can be calculated as:

$$P(\mathbf{L} = \mathbf{l}) = \prod_{i=1}^{W} \prod_{j=1}^{K} \binom{l_{i-1}^j}{l_i^j} p_{ij}^{l_i^j} (1 - p_{ij})^{(l_{i-1}^j - l_i^j)}, \tag{4.16}$$

where p_{ij} denotes probability that a single component of group j survives phase i, given that it has been functioning at the beginning of the phase, and l_0^j is the total

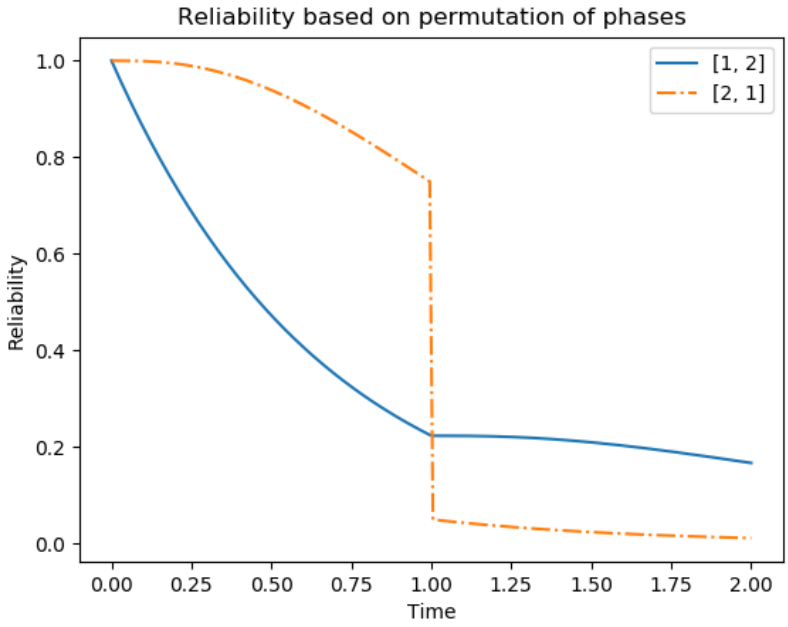


Figure 4.3: Survival function of a simple PM with serial (1) and parallel (2) phases for both possible orderings of the phases.

amount of components of group j in the system (i.e. the amount functioning at the beginning of the first phase).

4.2.3 Ordering of the phases

In robust design, we are interested in how to design our system to increase their performance, one of which is their reliability. Sometimes, we require a system to perform multiple actions, but we might not be dogmatic about which order it does it. This section addresses the question of whether, and how does the reliability of a PM depend on the ordering of the phases.

We will demonstrate on two examples that the order may matter. The presented figures depict mission reliability for an arbitrary time $t \in [0, T_M]$ (the means of conducting such analyses is described in [52] and will be omitted here).

The first example depicts (in Figure 4.3) differences in reliability functions for a PM consisting of 3 components with exchangeable TTF distributions (i.e. K=1) and equal failure rate in each of the phases. One of the phases requires all the

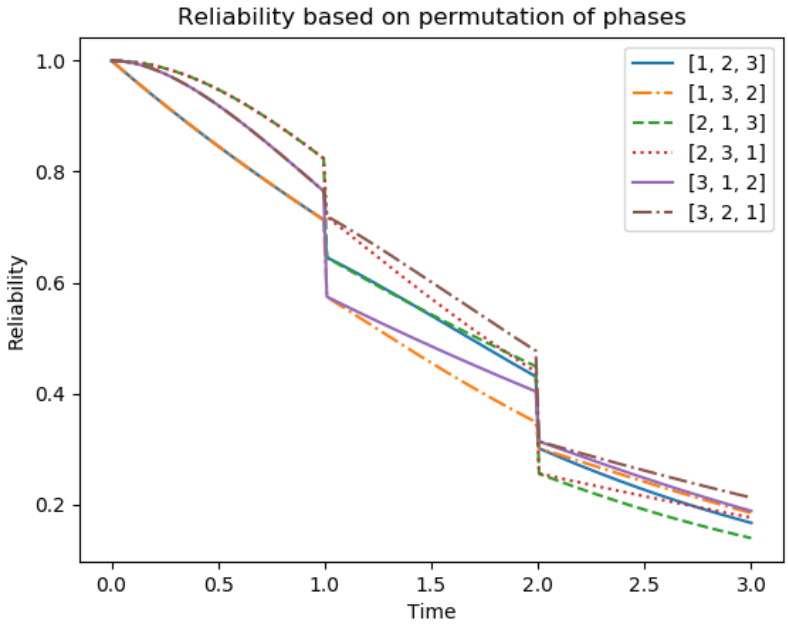


Figure 4.4: Survival function of a PM with three phases for all the possible permutations of phase ordering.

components to be functional, the other just one of them (serial and parallel systems, respectively). Reliability of the whole PM is equal to reliability at the time $t = 2$, the end of mission. Even from this simple example it is apparent, that ordering of the phases plays a significant role in system design.

Second example (in Figure 4.4) does the similar, just to provide a slightly more complex result for comparison. Now, the PM consists of 3 phases with the same component failure rates. Structure functions for each phase are described by reliability block diagrams depicted in Figure 4.5.

The examples show a phenomenon which was observed irrespective of the actual structure functions of the phases. The survival function of the PM systems differs greatly for possible orderings. The same applies for the mission reliability (survival function at the mission time T_M). Therefore, the phase ordering should be optimised whenever it is possible.

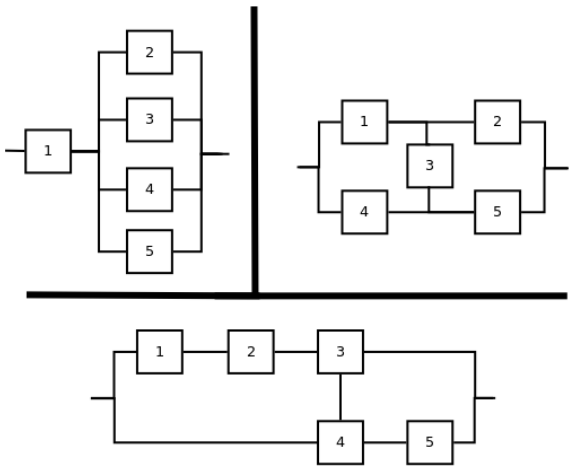


Figure 4.5: Reliability block diagrams for mission phases. Top-left: phase 1, top-right: phase 2, bottom: phase 3. The numbers identify individual physical components.

4.2.4 Properties of the ordering

No mathematical relations, that could easily decide which ordering is better has been found. In order to find the best ordering, a full optimisation process has to be conducted over all possible permutations of phase orders. This means, that for each of the permutation, we need to calculate the PM system survival signature. Nevertheless, a simple relation has been found, which can either serve as a heuristics to avoid necessity of calculating signatures for each ordering, or as an tool to restrict our search space.

A simple bound on PM system reliability can be obtained by applying the chain rule. Since:

$$Rel_{PM} = P(F_1)P(F_2|F_1)\dots P(F_W|F_1,\dots,F_{W-1}), \tag{4.17}$$

where event F_i represents successful completion of phase i.

Given that all $P(F_i|F_{1:i-1})$ are probabilities (i.e. ≤ 1), the PM reliability can be bounded by $P(F_1)$, and consequently also by $P(F_i)$ for any i. Therefore:

$$Rel_{PM} \leq \min_{i\in 1,\dots,W} P(F_i), \tag{4.18}$$

where $P(F_i) = \mathbb{E}\varphi_i(X(\tau_i))$.

4.2.5 Conclusions

We have reviewed survival signatures for system reliability modelling and its generalisation for phased missions (PMs). Further, we have introduced preliminary work on dependency of PM reliability on ordering of the phases. Examples provided in Figure 4.3 and Figure 4.4 clearly show that the ordering matters, which introduces further possibilities for designing reliable missions. So far, in order to find the most reliable phase ordering, an optimisation has to be run over the set of all permutations of mission phase orderings.

A simple heuristic method has been introduced which aims to help with PM system design (Equation 4.18). It provides no guarantees of optimality, but may be used as a heuristics, or to restrict the search space for reliability optimisation. The derived bound in Equation 4.18 depends on the ordering of the phases only through the hazard function which influences the probability that a component is functional at time τ_i and can be computed without taking into account the actual structure of a PM. This means that these partial (smaller and easier) survival signatures can be calculated only once before the optimisation process and used to discard some orderings without the need to construct the (more expensive) survival signature of the whole PM system. Consequently, the bound may also be used for preliminary reliability analysis to discard inadmissible designs or identify critical phases and components.

4.3 Decomposition approach for survival signatures computation

The following section is an exerpt of a conference contribution submitted for ESREL2019, [2].

We introduce a formula for computing system survival signatures by the means of merging survival signatures of multiple subsystems. The algorithm extends previous results for survival signatures on merging systems with no common components to systems with common components. It also extends existing results on merging dependent systems for calculating Samaniego's system signatures to survival signatures. Apart from these trivial extensions, we also introduce a novel decomposition method which allows us to decouple the dependencies among subsystems according to a wider class of events which may lead to savings of computational resources.

The survival signatures prove to be a very powerful computational tool for systems composed of groups of multiple components with similar failure laws, e.g. various networks with many hubs and transitions.

Suppose we have N random (basic) events on an underlying probability space $(\Omega, \mathcal{A}, P)$. From herein, we will represent these events by binary random variables, where $X_i = 1$ if the ith event obtains and 0 otherwise. Denote $\mathcal{X}$ $(=: \{0,1\}^N)$ the joint state space of these random variables and $\mathbf{X}$ the random state vector. With the notation $[N] := \{1, \ldots, N\}$, suppose we have a system, $([N], \varphi)$, with a structure function $\varphi : \mathcal{X} \to \{0,1\}$ representing the system state based on the occurrence of the basic events. We will model this system by a success tree with the top level event, "the system is functioning", represented by a binary random variable X_{top}. The system reliability can be calculated as the expected value of the structure function,

$$P(X_{\text{top}} = 1) = \mathbb{E}\{\varphi(X)\}. \tag{4.19}$$

In the rest of the section, we will describe a methodology which will simplify the computation of Equation 4.19, especially for procedures needing its repeated evaluation.

The computation of the expected value in Equation 4.19 is NP-hard, since it requires us to sum over the whole state space $\mathcal{X}$. Survival signature allows us to construct the reliability polynomial of a system, evaluation of which can be done with a polynomial, instead of the original exponential, complexity.

Suppose that we have another random variable Y on the common probability space. We will denote $\mathcal{D}^Y$ a decomposition of the state space $\mathcal{X}$, s.t.

$$\mathcal{D}^Y_{\mathcal{X}}(y) := \{X \in \mathcal{X} : Y = y\}. \tag{4.20}$$

The subscripts and superscripts will sometimes be omitted if they are evident from the context. We will usually denote the respective arguments by lower case letter, so the reader should be able to deduce them.

By the law of total expectation, Equation 4.19 can be expressed as

$$P(X_{\text{top}} = 1) = \sum_y \mathbb{E}\{\varphi(X)|X \in \mathcal{D}^Y(y)\}P_Y(y), \tag{4.21}$$

where by subscripts of the common probability measure P, we denote induced measures for respective random variables.

Suppose that we can group the basic events into K groups by their type, such that events in the same group are exchangeable, meaning that their joint probability distribution is invariant under permutation of event indices. Define function $\mathcal{C} : [N] \to [K]$, which assigns each basic events its type and a vector function $\mathbf{C} : \mathcal{X} \to \mathbb{N}_0^K$,

$$\mathbf{C}_k(x) = \sum_i \delta_{(\mathcal{C}(i),k)} x_i, \tag{4.22}$$

where $\delta_{(i,j)}$ is the Kronecker delta function.

For the ease of presentation, suppose that the basic events represent whether some system components are functioning ($X_i = 1$) or not. Define a K-dimensional random vector $L = \mathbf{C}(X)$, which represents, for each element, the random number of functioning components of the respective types. The exchangeability assumption become an assumption that the component failure times are exchangeable among the components of the same type.

Applying the decomposition in Equation 4.21 by L yields

$$\begin{aligned} P(X_{\text{top}} = 1) &= \sum_{l} \mathbb{E}\{\varphi(X)|L = l\} P_L(l), \\ &= \sum_{l=\mathbf{0}}^{\mathbf{M}} \Phi(l) P_L(l), \end{aligned} \tag{4.23}$$

where $\mathbf{M} := (M_1, \ldots, M_K)$ denotes the vector of total number of components of each of the types in the system and Φ is the survival signature introduced in [37]. The summation is over all $l \in \mathbb{N}_0^K$ such that $\mathbf{0} \leq l \leq \mathbf{M}$ in the sense of vector partial ordering.

The original survival signatures are conditional probabilities. Since the procedures described later in the section will require us to perform multiple arithmetical operations with them, we might arrive to accumulating vast numerical error over the course of the computation. For this reason, and in order to simplify some relationships, in this section we will describe their equivalent representation through "conditional counts", the *counting signatures*.

Due to the exchangeability assumption we have made earlier, the survival signature can be expressed as a classical probability. Recall the definition of state space decomposition $\mathcal{D}$ in Equation 4.20. Since the total number of states for which $L = l$ is dependent only on the number of components in the system and not its topology, we can focus solely on describing the cardinality of the set $\mathcal{D}^{L,X_{\text{top}}}(l, 1)$ in order to specify the survival signature, where the double superscript represents decomposition by the joint event $L \cap X_{\text{top}}$.

Definition 4.3.1 (Count signature)**.** For a system with top level event X_{top}, denote

$$\mathcal{N}(l, e) := \left|\mathcal{D}^{L,X_{\text{top}}}(l, e)\right|, \tag{4.24}$$

the system's counting signature. If the second argument e is omitted, it will denote $e = 1$.

The survival signature can be computed from the counting signature as

$$\Phi(l) = \frac{\mathcal{N}(l)}{|\mathcal{D}^L(l)|} = \frac{\mathcal{N}(l)}{\prod_{j=1}^{K}\binom{M_j}{l_j}}. \tag{4.25}$$

4.3.1 Independent systems

Construction of the survival signature of a system constructed as a series or a parallel system of two subsystems with known signatures was already introduced in [108]. Here, we introduce a generalization of those results for arbitrary system connection. We will state them for the counting signatures $\mathcal{N}$, of which the survival signatures Φ may be obtained via Equation 4.25. Manipulating with the counting signatures allows us to avoid necessity of evaluating probability mass functions of hypergeometric distribution as in [37] in order to normalize the results.

Let $(\bigcup_{i=1}^{R}\mathcal{I}^i, \varphi)$ be the system of interest composed of subsystems $(\mathcal{I}^i, \varphi^i)$, such that they have no components in common, i.e. $i \neq j \Rightarrow \mathcal{I}^i \cap \mathcal{I}^j = \emptyset$.

We will denote $\mathcal{X}_{\mathcal{I}} = \{0, 1\}^{|\mathcal{I}|}$ the state space of components indexed by elements of an index set $\mathcal{I}$. The subscript on $\mathcal{X}$ will often imply that we refer to a subspace of some larger joint space. Let $\widehat{\varphi}$ denote the top level gate structure function, that is $\varphi(X) = \widehat{\varphi}(X^1_{\text{top}}, \ldots, X^R_{\text{top}})$, where $X^i_{\text{top}} = \varphi^i(X_{\mathcal{I}^1})$ represent the top level events of the respective subsystems. Denote $\mathbf{M}^i$ the vector of the total number of components of each type in subsystem i and $\mathcal{N}^i$ their counting signatures.

The following lemmas will aid us with the proofs to come.

Lemma 4.3.1 (Law of total counts.)**.** *Let us have a system and random variables* A, B*.*

$$\left|\mathcal{D}^A(a)\right| = \sum_b \left|\mathcal{D}^{A,B}(a,b)\right|. \tag{4.26}$$

Proof. $\mathcal{D}^B$ is a total disjoint decomposition of the state space and the set cardinality is an additive measure. □

Lemma 4.3.2 (Product counts.)**.** *Let us have a system with index set* $\mathcal{I}$ *and random variables* A, B*, s.t.* A *is independent on all the basic events indexed by indices in some set* $\mathcal{K} \subset \mathcal{I}$ *and* B *is independent on all in* $\mathcal{I} \setminus \mathcal{K}$*. Then*

$$\mathcal{D}^A_{\mathcal{X}}(a) \cap \mathcal{D}^B_{\mathcal{X}}(b) = \mathcal{D}^A_{\mathcal{X}_{\mathcal{I}\setminus\mathcal{K}}}(a) \otimes \mathcal{D}^B_{\mathcal{X}_{\mathcal{K}}}(b). \tag{4.27}$$

Proof. If A is independent on basic events indexed by indices in $\mathcal{K}$, then

$$\mathcal{D}^A_{\mathcal{X}}(a) \cap \mathcal{D}^B_{\mathcal{X}}(b) = \mathcal{D}^A_{\mathcal{X}_{\mathcal{I}\setminus\mathcal{K}}}(a) \otimes 2^{\mathcal{K}}. \tag{4.28}$$

And similarly for B. The intersections can then be executed space-wise. □

Now, we can state the formula for the counting signature of a system composed of independent subsystems.

Proposition 4.3.1 (Combination of independent systems.)**.**

$$\mathcal{N}(l,e) = \sum_{\substack{\alpha \\ \sum_j \alpha^j = l}} \sum_{\substack{z \\ \widehat{\varphi}(z)=e}} \prod_{i=1}^{R} \mathcal{N}(\alpha^i, z_i). \tag{4.29}$$

Proof. We need to show that $\mathcal{N}(l,e) = \left|\mathcal{D}^{L,X_{\text{top}}}_{\mathcal{X}}(l,e)\right|$ for each l, e.

Denote $A = (L^1, \ldots, L^R)$, where $L^i = \mathbf{C}(X_{\mathcal{I}^i \setminus \mathcal{K}})$, and $Z = (X^1_{\text{top}}, \ldots, X^R_{\text{top}})$, where $X^i_{\text{top}} = \varphi^i(X_{\mathcal{I}^i})$ are the respective top level events of the subsystems.

$$
\begin{aligned}
\left|\mathcal{D}_{\mathcal{X}}^{L,X_{\text{top}}}(l,e)\right| &= \sum_{\alpha,z}\left|\mathcal{D}_{\mathcal{X}}^{L,X_{\text{top}}}(l,e)\cap\mathcal{D}_{\mathcal{X}}^{A,Z}(\alpha,z)\right|,\\
&= \sum_{\substack{\alpha\\ \sum_j \alpha^j=l}}\sum_{\substack{z\\ \widehat{\varphi}(z)=e}}\left|\mathcal{D}_{\mathcal{X}}^{A,Z}(\alpha,z)\right|,\\
&= \sum_{\substack{\alpha\\ \sum_j \alpha^j=l}}\sum_{\substack{z\\ \widehat{\varphi}(z)=e}}\left|\bigcap_{i=1}^{R}\mathcal{D}_{\mathcal{X}}^{L^i,X_{\text{top}}^i}(\alpha^i,z_i)\right|,\\
&= \sum_{\substack{\alpha\\ \sum_j \alpha^j=l}}\sum_{\substack{z\\ \widehat{\varphi}(z)=e}}\left|\bigotimes_{i=1}^{R}\mathcal{D}_{\mathcal{X}_{\mathcal{I}^i}}^{L^i,X_{\text{top}}^i}(\alpha^i,z_i)\right|,\\
&= \sum_{\substack{\alpha\\ \sum_j \alpha^j=l}}\sum_{\substack{z\\ \widehat{\varphi}(z)=e}}\prod_{i=1}^{R}\left|\mathcal{D}_{\mathcal{X}_{\mathcal{I}^i}}^{L^i,X_{\text{top}}^i}(\alpha^i,z_i)\right|,\\
&= \sum_{\substack{\alpha\\ \sum_j \alpha^j=l}}\sum_{\substack{z\\ \widehat{\varphi}(z)=e}}\prod_{i=1}^{R}\mathcal{N}^i(\alpha^i,z_i),\\
&= \mathcal{N}(l,e).
\end{aligned}
\tag{4.30}
$$

□

4.3.2 Dependent systems

The results from the last section required the subsystems to be independent. In this section, we will introduce a generalized procedure which will allow us to compute counting signatures for a larger class of systems. The procedure is similar to that of [111], who introduced a decoupling scheme for Samaniego's signatures.

The method in [111] allows us to split the state space into disjoint parts in which we can treat the systems as independent and use similar results to those described in Section 4.3.1. They decomposed the state space according to all the possible states of the shared components. Once the states of all the shared components are known, a new structure function is induced for which the subsystems are independent.

We proceed in a similar fashion but 1) provide results for systems with multiple types of components via their counting signatures and 2) our generalization allows us to decompose according to a wider class of events which allows us to choose among

them according to the computational complexity.

Suppose we have a system $(\mathcal{I}, \varphi)$ and a random vector, we will refer to it as the *boundary*, $\mathcal{B}$. Define a set $\mathcal{I} \setminus \mathcal{K}$ of all the indices of basic events (component states) such that $\mathcal{B}$ is independent on all the events indexed by elements of $\mathcal{I} \setminus \mathcal{K}$.

Definition 4.3.2 (Decomposed counting signature)**.** If there exists, for each b in the state space of $\mathcal{B}$, a structure function $\tilde{\varphi}_b : \mathcal{X}_{\mathcal{I}\setminus\mathcal{K}} \to \{0, 1\}$, such that $\forall y \in \mathcal{D}_{\mathcal{K}}^{\mathcal{B}} : \tilde{\varphi}_b(x_{\mathcal{I}\setminus\mathcal{K}}) = e \Rightarrow \varphi(z) = e$, where $z \in \mathcal{X}_{\mathcal{I}} : z_i = y_i$ if $i \in \mathcal{K}$ and $x_{\mathcal{I}\setminus\mathcal{K},i}$ otherwise, then we define the decomposed counting signature

$$\mathcal{N}_b(l, e) := \left| \mathcal{D}_{\mathcal{X}_{\mathcal{I}\setminus\mathcal{K}}}^{L_{\mathcal{I}\setminus\mathcal{K}}, X_b}(l, e) \right|, \tag{4.31}$$

where $X_b = \tilde{\varphi}_b(X_{\mathcal{I}\setminus\mathcal{K}})$ and $L_{\mathcal{I}\setminus\mathcal{K}} = \mathbf{C}(X_{\mathcal{I}\setminus\mathcal{K}})$.

The reasoning behind the decomposed counting signature is this. We require, that the relation between $x_{\mathcal{I}\setminus\mathcal{K}}$ and x_{top} becomes deterministic for each possible realization of b, such that x_{top} is some function of $x_{\mathcal{I}\setminus\mathcal{K}}$ and b. If the requirement from Definition 4.3.2 holds, then this can be achieved by construction of the introduced new structure functions $\tilde{\varphi}_b$, for each b. These take as arguments the states of the components indexed by elements in $\mathcal{I} \setminus \mathcal{K}$.

The interpretation of the decomposed counting signature is the number of states in the restricted space state $\mathcal{X}_{\mathcal{I}\setminus\mathcal{K}}$, which will render the system into state e such that exactly l_k components of each respective type k, from those indexed by $\mathcal{I} \setminus \mathcal{K}$, are functional.

Suppose that we have a system $(\mathcal{I}, \varphi)$ composed of R subsystems $(\mathcal{I}^1, \varphi^1), \ldots, (\mathcal{I}^R, \varphi^R)$. Suppose that the system state is described by a top level structure function $\hat{\varphi} : \{0, 1\}^R \to \{0, 1\}$ such that $\forall x \in \mathcal{X} : \varphi(x) = \hat{\varphi}(\varphi^1(x_{\mathcal{I}^1}), \ldots, \varphi^R(x_{\mathcal{I}^R}))$, where the subscripts of x denote its projection on the respective subsystem state spaces. Suppose further that the subsystems are sharing some components, that is that for some $i, j : \mathcal{I}^i \cap \mathcal{I}^j \neq \emptyset$.

Let $\mathcal{B}$ be a boundary, as introduced in Definition 4.3.2, such that all the indicis of the shared components in all the subsystems belong to the set $\mathcal{K}$ induced by $\mathcal{B}$. Denote $\mathcal{X}_{\mathcal{B}}$ the state space of $\mathcal{B}$, M_b the supremum of $\mathbf{C}(y)$ over $y \in \mathcal{D}^{\mathcal{B}}_{\mathcal{K}}(b)$ (the number of components Definition 4.3.2 of each type among those indexed by $\mathcal{K}$) and $\mathcal{N}^{\mathcal{B}}(h,b) := \left|\mathcal{D}^{L_{\mathcal{K}},\mathcal{B}}_{\mathcal{X}_{\mathcal{K}}}(h,b)\right|$ the counting signature of the boundary.

Proposition 4.3.2 (Boundary decomposition.)**.** *If, for each of the subsystems there exists an augmented structure function $\tilde{\varphi}^i_b$ for each $b \in \mathcal{X}_{\mathcal{B}}$, as required in Definition 4.3.2, then*

$$\mathcal{N}(l,e) = \sum_{b\in\mathcal{X}_{\mathcal{B}}}\sum_{h=\mathbf{0}}^{M_b}\mathcal{N}^{\mathcal{B}}(h,b)\Big[\sum_{\substack{(\alpha^1,\dots,\alpha^R)\\ \sum_i \alpha^i+h=l}}\ \sum_{z\in\{0,1\}^R\hat{\varphi}(z)=e}\ \prod_{i=1}^{R}\mathcal{N}^i_b(\alpha^i,z_i)\Big], \tag{4.32}$$

where $\mathcal{N}^i_b$ are the respective counting signatures of the subsystems.

Proof.

$$\begin{aligned}
\left|\mathcal{D}^{L,X_{\text{top}}}(l,e)\right| &= \sum_b \left|\mathcal{D}^{\mathcal{B}}(b)\cap\mathcal{D}^{L,X_{\text{top}}}(l,e)\right| \\
&= \sum_b\sum_h \left|\mathcal{D}^{L_{\mathcal{K}},\mathcal{B}}(h,b)\cap\mathcal{D}^{L_{\mathcal{I}\setminus\mathcal{K}},X_b}(l-h,e)\right| \\
&= \sum_b\sum_h \left|\mathcal{D}^{L_{\mathcal{K}},\mathcal{B}}_{\mathcal{X}_{\mathcal{K}}}(h,b)\otimes\mathcal{D}^{L_{\mathcal{I}\setminus\mathcal{K}},X_b}_{\mathcal{X}_{\mathcal{I}\setminus\mathcal{K}}}(l-h,1)\right| \\
&= \sum_b\sum_h \left|\mathcal{D}^{L_{\mathcal{K}},\mathcal{B}}_{\mathcal{X}_{\mathcal{K}}}(h,b)\right|\cdot\left|\mathcal{D}^{L_{\mathcal{I}\setminus\mathcal{K}},X_b}_{\mathcal{X}_{\mathcal{I}\setminus\mathcal{K}}}(l-h,e)\right|
\end{aligned} \tag{4.33}$$

The $\left|\mathcal{D}^{L_{\mathcal{K}},\mathcal{B}}_{\mathcal{X}_{\mathcal{K}}}(h,b)\right|$ is the counting signature of the boundary and the first two sums correspond to the first two sums in proposition Proposition 4.3.2.

Denote $A = (L^1,\dots,L^R)$, where $L^i = \mathbf{C}(X_{\mathcal{I}^i\setminus\mathcal{K}})$, and $Z = (X^1_b,\dots,X^R_b)$, where $X^i_b = \tilde{\varphi}^i_b(X_{\mathcal{I}^i})$.

$$\begin{aligned}\left|\mathcal{D}^{L_{\mathcal{I}\setminus\mathcal{K}},X_b}_{\mathcal{X}_{\mathcal{I}\setminus\mathcal{K}}}(l-h,e)\right| &= \sum_{\alpha}\sum_{z\in\{0,1\}^R}\left|\mathcal{D}^{A,Z,L_{\mathcal{I}\setminus\mathcal{K}},X_b}_{\mathcal{X}_{\mathcal{I}\setminus\mathcal{K}}}(\alpha,z,l-h,e)\right| \\ &= \sum_{\substack{\alpha\\ \sum_i \alpha^i+h=l}}\sum_{\substack{z\in\{0,1\}^R\\ \hat{\varphi}(z)=e}}\left|\mathcal{D}^{A,Z}_{\mathcal{X}_{\mathcal{I}\setminus\mathcal{K}}}(\alpha,z)\right| \\ &= \sum_{\substack{\alpha\\ \sum_i \alpha^i+h=l}}\sum_{\substack{z\in\{0,1\}^R\\ \hat{\varphi}(z)=e}}\prod_{i=1}^{R}\left|\mathcal{D}^{L^i,Z^i}_{\mathcal{X}_{\mathcal{I}^i\setminus\mathcal{K}}}(\alpha^i,z_i)\right|.\end{aligned} \tag{4.34}$$

Since $Z_i = \tilde{\varphi}^i_b(X_{\mathcal{I}^i})$ are the respective augmented top events, $\left|\mathcal{D}^{L^i,Z^i}_{\mathcal{X}_{\mathcal{I}^i\setminus\mathcal{K}}}(\alpha^i,z_i)\right| = \mathcal{N}^i_b(\alpha^i,z_i)$.

□

It can be seen that Proposition 4.3.2 is a generalization of the Proposition 4.3.1. For independent systems, the boundary may be chosen as a constant event, say $\mathcal{B}\equiv b_0$, thus independent of all the basic events of the system and with $\mathcal{N}^{\mathcal{B}}(\mathbf{0},b_0)=1$ and 0 otherwise. In such a case, the first two sums in Proposition 4.3.2 disappear and we arrive to the same summation as in the Proposition 4.3.1.

Although it might be difficult to ensure existence of the augmented structure functions $\tilde{\varphi}_b$ in a general system, the situation gets easier in fault and success trees. In their hierarchical structure, the boundaries may be composed of some macro events or a collection of them. Nevertheless, an arbitrary choice still would not guarantee the existence of $\tilde{\varphi}_b$ and further investigation has to be performed. We will not present sufficient conditions for a validity of boundaries, but an example of both a valid and an invalid choices are demonstrated on an example in Section 4.3.4.

4.3.3 An example of decoupling systems with shared components

We will demonstrate the formula in Proposition 4.3.2 on a simple example. Suppose that a success tree of a system is such as on Figure 4.6. Suppose that components $\{1,2\}$ are of type 0 and the component x_3 is of type 1. The total number of components of each type in the system will therefore be $M=(2,1)$. We identify

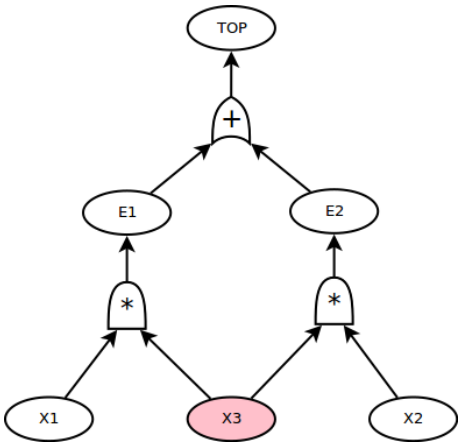


Figure 4.6: An example system. The shared component is highlighted in red color.

two subsystems $(\{1,3\}, \text{AND})$ and $(\{2,3\}, \text{AND})$ with E_1, E_2 being the respective subsystem top level events. The original system top event structure function $\hat{\varphi} = \text{OR}$. The set of shared basic events among the subsystems is $\mathcal{K} = \{3\}$.

For the computation of the original system's counting signature, we need to choose a suitable boundary and compute the subsystems decomposed counting signatures. A natural boundary to choose is $\mathcal{B} = \{X_3\}$ with state space $\mathcal{X}_{\mathcal{B}} = \{0, 1\}$. The boundary counting signature is shown in Table 4.1. The induced augmented structure functions for E_1, E_2 are shown in Table 4.2 and $\mathcal{N}_b^1 = \mathcal{N}_b^2$ in Table 4.3.

Table 4.1: The boundary counting signature $\mathcal{N}^{\mathcal{B}}(l, (x_3))$.

l \ x_3	(0)	(1)
(0,0)	1	0
(0,1)	0	1

Table 4.2: Augmented structure functions φ_b^i of subsystems $(\{1,3\}, \text{AND})$ and $(\{2,3\}, \text{AND})$.

x_1	$\tilde{\varphi}^1_{(0)}(x_1)$	$\tilde{\varphi}^1_{(1)}(x_1)$	x_2	$\tilde{\varphi}^2_{(0)}(x_2)$	$\tilde{\varphi}^2_{(1)}(x_2)$
0	0	0	0	0	0
1	0	1	1	0	1

Table 4.3: Decomposed counting signatures (Definition 4.3.2) of the subsystems $(\{1,3\}, \text{AND})$ and $(\{2,3\}, \text{AND})$. Only $\mathcal{N}_b^1$ is shown since they are identical. The subscript of $\mathcal{N}$ denotes the realization of $b = (x_3)$.

l	$\mathcal{N}^1_{(0)}(l,0)$	$\mathcal{N}^1_{(0)}(l,1)$	$\mathcal{N}^1_{(1)}(l,0)$	$\mathcal{N}^1_{(1)}(l,1)$
$(0,0)$	1	0	1	0
$(1,0)$	1	0	0	1

According to Proposition 4.3.2:

For $l < (1,1)$, the counting signature is zero.

For $l = (1,1)$,

$$\begin{aligned}\mathcal{N}((1,1)) = &\sum_{x_3\in\{0,1\}} \sum_{\alpha:\sum_j \alpha^j+(0,1)=(1,1)} \\ &\cdot \left[\mathcal{N}^1_{x_3}(\alpha^1,1)\mathcal{N}^2_{x_3}(\alpha^2,1) + \mathcal{N}^1_{x_3}(\alpha^1,0)\mathcal{N}^2_{x_3}(\alpha^2,1) + \mathcal{N}^1_{x_3}(\alpha^1,1)\mathcal{N}^2_{x_3}(\alpha^2,0)\right] \\ &= \mathcal{N}^1_1((0,0),0)\mathcal{N}^2_1((1,0),1) + \mathcal{N}^1_1((1,0),1)\mathcal{N}^2_1((0,0),0) \\ &= 2\end{aligned} \tag{4.35}$$

For $l = (2,0)$, it implies that $\alpha^1 = \alpha^2 = (1,0)$ and $b = (x_3) = (0)$, so the only contribution to Equation 4.35 can come from $\mathcal{N}^1_{(0)}((1,0),1)\mathcal{N}^2_{(0)}((1,0),1)$, but that is null (see Table 4.3).

For $l = (2,1)$, all the components are functional, $\alpha^1 = \alpha^2 = (1,0)$ and $b = (x_3) = (1)$, so the only non-zero contribution to the counting signature comes from $\mathcal{N}^1_{(1)}((1,0),1)\mathcal{N}^2_{(1)}((1,0),1) = 1$. Therefore $\mathcal{N}((2,1)) = 1$.

For all other $l \in \mathbb{N}_0^K$, the counting signature is zero.

Both the counting signature and the survival signature Equation 4.25 of the example system are shown in Table 4.4.

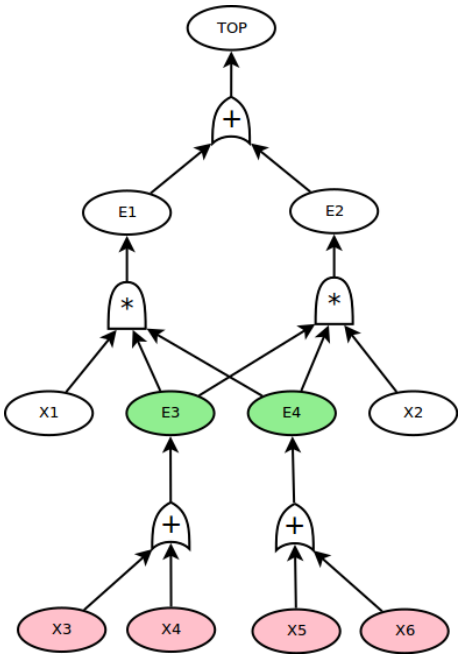


Figure 4.7: An example success tree. Shared components are highlighted in red, and the chosen boundary for decomposition in green color.

Table 4.4: The counting signature $\mathcal{N}$ and the survival signature Φ of the system from Figure 4.6.

l	$\mathcal{N}(l)$	$\Phi(l)$
(0,0)	0	0.
(1,0)	0	0.
(2,0)	0	0.
(0,1)	0	0.
(1,1)	2	2/3
(2,1)	1	1.

4.3.4 An example of decoupling by boundary in a success tree

In this section we wish to demonstrate the advantage that our method provides over decoupling only over the state space of the shared components. For larger sets of shared components, the decomposition would decouple the the state space

into $2^{|\mathcal{K}|}$ subsets for which the subsystems would be independent. The generalized procedure introduced in Section 4.3.2 allows us to, sometimes, identify a more efficient decomposition and thus decrease the CPU cost of our computation.

In the example to follow, we would need to construct 2^4 augmented success trees and merge them independently, if we were to decompose according to the set of shared components. Instead, by using the boundary decomposition, we only need to do this 2^2 times.

Suppose that a success tree of a system is such as on Figure 4.7. Suppose that, for the sake of simplicity, all the components are of the same type. We identify two subsystems $(\{1,3,4,5,6\}, \text{AND})$ and $(\{2,3,4,5,6\}, \text{AND})$ with E_1, E_2 being the respective subsystem top level events. The original system top event structure function $\hat{\varphi} = \text{OR}$. The set of shared basic events among the subsystems is $\mathcal{K} = \{3,4,5,6\}$.

We select boundary $\mathcal{B}$ be the macro-events E_3, E_4. Clearly, for each $b \in \mathcal{X}_{\mathcal{B}}$, if we fix $(e_3, e_4) = b$, a new success tree with only basic events $\{x_1, x_2\}$ is created. The augmented structure functions $\tilde{\varphi}_b^1, \tilde{\varphi}_b^2$ are taken to be the structure functions in this derived success tree in which $(e_3, e_4) = b$. That is $\tilde{\varphi}_b^1(x_1) = \text{AND}(x_1, e_3, e_4)$ and $\tilde{\varphi}_b^2(x_2) = \text{AND}(x_2, e_3, e_4)$. If $\tilde{\varphi}_b^1(x_{\mathcal{I}^1}) = 1$, then $\varphi^1(y) = 1$ for any choice of $y \in \mathcal{D}^{\mathcal{B}}(b)$, since all the choices of y result in the same derived success tree with $(e_3, e_4) = b$.

Such a choice of $\mathcal{B}$ therefore complies with the requirements of Proposition 4.3.2.

Remark 4.3.1 (Non-compliant boundary.). It is possible to choose a boundary which will not comply with the requirements of Proposition 4.3.2. Such a choice could be $\mathcal{B} = (E_1)$. Then, $\tilde{\varphi}_b^1(x_{\mathcal{I}^1}) = (b_1)$, which does not depend on which $y \in \mathcal{D}^{\mathcal{B}}(b)$ was chosen - that is fine. But $\tilde{\varphi}_b^2$ would, since, e.g. if we set $\tilde{\varphi}_{b=(0)}^2(x_2 = 1) = 1$, then we can find $y_1, y_2 \in \mathcal{D}^B((0))$, such that $\varphi^2(y_1) = 1$ and $\varphi^2(y_2) = 0$, namely $y_1 = (x_1, x_3, x_4, x_5, x_6) = (0,1,1,1,1)$ and $y_2 = (0,0,0,0,0)$.

Let us now compute the counting signature of the example system.

The boundary counting signature, $\mathcal{N}^{\mathcal{B}}$ is given in Table 4.5.

Table 4.5: The boundary counting signature $\mathcal{N}^{\mathcal{B}}(l,(e_3,e_4))$.

l \ $b=(e_3,e_4)$	(0,0)	(0,1)	(1,0)	(1,1)
0	1	0	0	0
1	0	2	2	0
2	0	1	1	4
3	0	0	0	4
4	0	0	0	1

The decomposed counting signatures of the subsystems are shown in Table 4.6, where only $\mathcal{N}_b^1$ is shown since $\mathcal{N}_b^1(l)=\mathcal{N}_b^2(l)$.

Table 4.6: $\mathcal{N}_b^1(l,e)=\mathcal{N}_b^2(l,e)$.

	$l=0$	$l=0$	$l=1$	$l=1$
$b=(e_3,e_4)$	$e=0$	$e=1$	$e=0$	$e=1$
(0,0)	1	0	1	0
(0,1)	1	0	1	0
(1,0)	1	0	1	0
(1,1)	1	0	0	1

Now, according to the Proposition 4.3.2,

$$\mathcal{N}(l)=\sum_b\sum_j\mathcal{N}^{\mathcal{B}}(j,b)\sum_{\substack{\alpha\\ \sum_i\alpha^i+j=l}}\left[\mathcal{N}_b^1(\alpha_1,1)\mathcal{N}_b^2(\alpha_2,1)+\mathcal{N}_b^1(\alpha_1,0)\mathcal{N}_b^2(\alpha_2,1)+\mathcal{N}_b^1(\alpha_1,1)\mathcal{N}_b^2(\alpha_2,0)\right] \tag{4.36}$$

Notice that in order for the whole system to be functional, at least one of the subsystems has to be. From Table 4.6, we can see that only choice of b, for which

$\mathcal{N}_b^i(\cdot, 1) > 0$ is $b = (1, 1)$. The rest would result into a summation of zeros.

Further, in order for $\mathcal{N}^1_{(1,1)}(\alpha_1, 1)$, $\mathcal{N}^2_{(1,1)}(\alpha_2, 1)$ to be positive, $(\alpha^1, \alpha^2) = (1, 1)$, so this term will only contribute when $j = l - 2$.

In order for $\mathcal{N}^1_{(1,1)}(\alpha_1, 0)\mathcal{N}^2_{(1,1)}(\alpha_2, 1)$ to be positive, $(\alpha^1, \alpha^2) = (0, 1)$ and for $\mathcal{N}^1_{(1,1)}(\alpha_1, 1)\mathcal{N}^2_{(1,1)}(\alpha_2, 0)$ it must be so that $(\alpha^1, \alpha^2) = (1, 0)$. So both these terms will only contribute when $j = l - 1$.

For a general l, we therefore get relation

$$\begin{aligned}\mathcal{N}(l) &= \mathcal{N}^{\mathcal{B}}(l-2,(1,1))\mathcal{N}^1_{(1,1)}(1,1)\mathcal{N}^2_{(1,1)}(1,1) \\ &\quad + \mathcal{N}^{\mathcal{B}}(l-1,(1,1))\Big[\mathcal{N}^1_{(1,1)}(0,0)\mathcal{N}^2_{(1,1)}(1,1) + \mathcal{N}^1_{(1,1)}(1,1)\mathcal{N}^2_{(1,1)}(0,0)\Big] \\ &= \mathcal{N}^{\mathcal{B}}(l-2,(1,1)) + 2\cdot\mathcal{N}^{\mathcal{B}}(l-1,(1,1)).\end{aligned} \tag{4.37}$$

The reconstructed counting signature and survival signature (derived via Equation 4.25) are shown in Table 4.7.

Table 4.7: The counting signature $\mathcal{N}$ and the survival signature Φ of the system from Figure 4.7.

l	$\mathcal{N}$	Φ
0	0	0
1	0	0
2	0	0
3	8	0.4
4	12	0.6
5	6	1
6	1	1

4.3.5 Conclusions

We have introduced an approach for constructing system signatures from (decomposed) signatures of its subsystems. It generalizes the results previously available for constructing survival signatures of independent subsystems and for constructing Samaniego's signatures for subsystems with shared components via decoupling over the state space of the shared components. Our method allows us to use a larger class of possible decompositions, which can result in saving of computational resources and, therefore, possibility to compute signatures of larger systems. The sufficient conditions for validity of the choice of a particular decomposition were not stated in this work.

Chapter 5

Investigations on set valued statistical methods

5.1 Interval inference on masked systems

This section is an exerpt of a conference contribution submitted to SMPS2018, [3].

Outside of controlled experiment scope, we have only limited information available to carry out desired inferences. One such scenario is when we wish to infer the topology of a system given only data representing system lifetimes without information about states of components in time of system failure, and only limited information about lifetimes of the components of which the system is composed. This scenario, masked system inference, has been studied before for systems with only one component type, with interest of inferring both system topology and lifetime distribution of component composing it. In this section we study similar scenario in which we consider systems consisting of multiple types of components. We assume that distribution of component lifetimes is known to belong to a prior-specified set of distributions and our intention is to reflect this information via a set of likelihood functions which will be used to obtain an imprecise posterior on the set of considered system topologies.

5.1.1 Introduction

Masked system inference concerns about carrying out inferences about the underlying system model from system failure time observations, rather than the more commonly studied situations where life test data is available on components. Our inference may concern lifetime distributions of system components or structure of the system. Also, the prior information may be available in various forms and sometimes prevents us from constructing suitable prior distributions for Bayesian inference.

We will study here a scenario in which we wish to infer unknown structure of the system from masked system lifetimes given prior distribution on system structure and a set of credible component lifetime distributions. System structures will be specified by survival signatures (introduced in [37]) and we will use theory of imprecise probabilities (IP; more in [112]) to describe and obtain inference results.

System reliability inferences with survival signatures based on component failure observations were described in [43] and further extended for IP framework in [44]. Masked system structural inference in Bayesian framework for single component type systems were studied by Aslett in [113], where further elaboration of the nature of inference on masked system with uncertain structure and its numerical solution by Monte Carlo algorithms is presented.

5.1.2 Masked system inference

Let Ω_S be a set of considered systems. We model underlying distribution of component lifetimes with a parametric model and we index collection of component lifetime distributions by multi-parameter $\theta \in \Omega_\Theta$. For each combination of system s and set of distributions indexed by θ we assume that we can calculate the system survival function $R(t|s,\theta) = P(T_{sys} > t)$.

We further assume that the observables, D, are distributed according to system lifetime distribution (elements d_i represent observations of system failure times, r.v.

T_{sys}). With additional assumptions about dependency among observations (e.g. i.i.d.), we can construct the observation model $f(d|\theta, s) \triangleq \mathcal{L}(\theta, s; d)$, for inference purposes:

$$\mathcal{L}(s, \theta; D) = \prod_i f(d_i|s, \theta) = \prod_i \left(- \left[\frac{\partial}{\partial_t} R_{sys}(t|s, \theta) \right]_{t=d_i} \right), \tag{5.1}$$

where specific form of $f(d_i|\theta, s)$ depends on our system model and shall be given by Equation 5.5.

The system design is considered unknown and is therefore included in the likelihood, which then enables joint inference about the reliability and the topology of the system.

5.1.3 Imprecise probability inference of masked systems

In IP inference we operate with set of models (set of priors, set of likelihoods). For each of singular model of this set, we can carry out standard inference and analyse the collection partial results. If our aim is to infer probability of some event of interest, in IP scenario we can calculate the bounds for coherent inferences - lower and upper probabilities, where lower probability is minimal inferred probability over the models in the set, and similarly for the upper probability.

In system inference with uncertainty about both component lifetime distributions and system structure, we can choose different uncertainty models for these respective variables. By imprecision we model situations in which we know only possible domain of random variable and are unable to specify prior distribution for Bayesian inference. Such case will lead to IP inference, where each particular value of imprecise random variable defines a stochastic model on which standard Bayesian inference can be performed and the results integrated.

In our case, we assume that we can construct prior distribution on system structures and know only set in which component lifetime distribution parameter θ lies. Lower bound on posterior predictive survival function can be obtained as:

$$\underline{P}(T_{sys} > t|D = d) = \min_{\theta \in \Omega_\Theta} \int_s R_{sys}(t|s,\theta) \frac{\mathcal{L}(s,\theta;d)}{Z(\theta,d)} f(s|\theta) ds, \tag{5.2}$$

where R_{sys} is the system lifetime survival function, $\mathcal{L}$ is the likelihood function described in Equation 5.1, f is prior density for Bayesian inference and factor Z is for posterior distribution normalization, i.e. $Z(\theta,d) = \int_s \mathcal{L}(s,\theta;d) f(s|\theta) ds$.

Upper bound is obtained via maximization of the same expression.

Similarly we can also introduce the lower posterior distribution on system structure as:

$$\underline{f}(S = s|D = d) = \min_{\theta \in \Omega_\Theta} \frac{\mathcal{L}(s,\theta;d)}{Z(\theta,d)} f(s|\theta), \tag{5.3}$$

with respective maximizations in case of upper bound.

5.1.4 Survival signatures for system state modelling

Via component state space decomposition, we can express the system survival function for systems consisting of K distinct types of components, with M_k components of type k with i.i.d. lifetimes for each component type k, as:

$$\begin{aligned} R_{sys}(t|s,\theta) &= \sum_{\mathbf{l}} P(T_{sys} > t|\mathbf{L}(t) = \mathbf{l}, s, \theta) P(\mathbf{L}(t) = \mathbf{l}|s,\theta) \\ &= \sum_{\mathbf{l}} \phi_s(\mathbf{l}) \prod_{k=1}^{K} \binom{M_k}{l_k} R_{\theta,k}^{l_k}(t) F_{\theta,k}^{M_k - l_k}(t), \end{aligned} \tag{5.4}$$

where $\phi_s(\mathbf{l}) = P(T_{sys} > t|\mathbf{L}(t) = \mathbf{l}, s)$ is called survival signature of system s, random vector $\mathbf{L}(t)$ represents number of functioning components of each respective type at time t (i.e. L_i is number of functioning components of type i), summation is over all possible combinations $\mathbf{l}$ of numbers of functioning component of each type. Survival functions $R_{\theta,k}$ and cumulative distribution functions (CDFs) $F_{\theta,k}$ indexed

by component type k and (multi-)parameter θ denote respective lifetime distribution characteristics for distinct component types.

The single observation density for systems described by survival signatures is therefore given by:

$$f(d_i|s,\theta) = -\left[\frac{\partial}{\partial_t} R_{sys}(t|s,\theta)\right]_{t=d_i} =$$
$$-\sum_{\mathbf{l}} \phi_s(\mathbf{l}) \sum_{k=1}^{K} \left\{ \binom{M_k}{l_k} \left[\frac{\partial}{\partial_t}\left(R_{\theta,k}^{l_k}(t) F_{\theta,k}^{M_k-l_k}(t)\right)\right]_{t=d_i} \prod_{k\neq j=1}^{K} \binom{M_j}{l_j} R_{\theta,j}^{l_j}(d_i) F_{\theta,j}^{M_j-l_j}(d_i) \right\}$$
$$= \sum_{\mathbf{l}} \left\{ \phi_s(\mathbf{l}) \prod_{k=1}^{K} \left[\binom{M_k}{l_k} F_{\theta,k}^{M_k-l_k}(d_i) R_{\theta,k}^{l_k}(d_i)\right] \sum_{k=1}^{K} \left[\left(\frac{l_k}{R_{\theta,k}(d_i)} - \frac{M_k-l_k}{F_{\theta,k}(d_i)}\right) f_{\theta,k}(d_i)\right] \right\}, \tag{5.5}$$

where $f_{\theta,k}(.)$ is probability density function of kth component type lifetime.

We have derived everything necessary to be able to compute both the imprecise posterior and posterior predictive distributions in the setting where only masked system lifetime data are available and when the system design may be unknown. This allows us to perform joint inference on the component lifetime parameters and the topology of the system using imprecise probability.

In the remainder of the section we will demonstrate the method for inference of system structure and predictive system lifetime.

5.1.5 Examples

In the experiments, we shall assume that the real system structure is one of those described by survival signatures in Table 5.1 (those are all simply connected systems of order 4, as defined and listed in [113], each with a random component type assignment). These systems consists of $K = 2$ types components, 2 components of each type ($M_1 = M_2 = 2$). Underlying component type lifetime distributions are assumed to be exponential with rates $\lambda_1 = 0.45$ and $\lambda_2 \in [0.06, 1.12]$ ($\Omega_\Theta = \Omega_{\Lambda_1} \otimes \Omega_{\Lambda_2}$). Prior distribution on systems ($f(s|\theta)$ in Equation 5.2 and Equation 5.3) is chosen to be uniform for all choices of λ_2 .

The data, observed system failures, for experiments are simulated from system labeled as 6, which will be hereon referred to as the "ground truth' ' system. Ground truth hazard rate for components of type 2 is chosen to be $\lambda_2 = 0.32$. We will not perform inference on λ_2 and considered it a fixed imprecise parameter for the purpose of investigating properties of the role of imprecision in the analyzed models. Although this is an artificial assumption, it may represent, for example, a situation in which components are provided by various manufacturers with contract-specified tolerances on component reliability.

Table 5.1: Survival signatures of systems in Ω_S. Zero row is being omitted ($\phi_s(\mathbf{0}) = 0$).

l_1	l_2	1	2	3	4	5	6	7	8	9	10	11
0	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.50	0.50	1.00
0	2	0.00	0.00	0.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00
1	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	1.00
1	1	0.00	0.00	0.25	0.50	0.50	0.75	0.50	0.50	0.75	1.00	1.00
1	2	0.00	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	1.00	1.00
2	1	0.00	0.50	0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

5.1.6 Survival predictions are not monotonic, nor convex

Since the predictions are defined by their bounds, it is necessary to acquire them by optimization. Optimization problems are greatly simplified for monotonic functions (we only need to investigate bounds of the set) or convex or concave functions (where efficient gradient based algorithms may be employed). Although the survival function predictions are monotone in case of know system structure ($|\Omega_s| = 1$), neither of these desired properties could be proven analytically in general for predictions with unknown system structure. Conducted experiment (see Figure 5.1) provides a

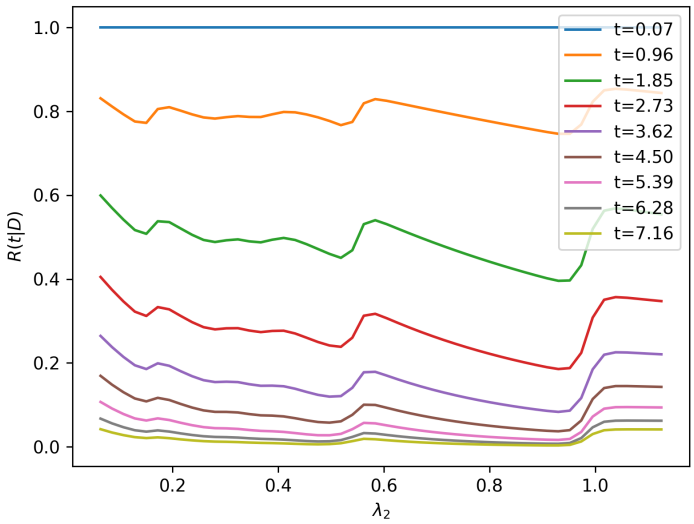


Figure 5.1: Dependency of posterior survival function predictions for various selected times on imprecise λ_2 obtained by analysing 250 data samples

counterexample for monotonicity, convexity and concavity of posterior predictions in case of unknown system structure. Furthermore, Figure 5.3 provides a counterexample for the same in case of system structure posterior inference.

5.1.7 Imprecise structure posterior and system identification

Two basic inferences of our interest are for the system lifetime survival function (via Equation 5.2), and for posterior system distribution (via Equation 5.3). An example of predictive and structure inferences are shown in Figure 5.2. On the left side, the intervals for each system represents lower and upper bounds for posterior on the set Ω_Θ. On the right picture, one set of prediction bounds for prior distribution on system structures (before updating by observations) and another for posterior obtained via Bayesian updating are compared with the Kaplan-Meier estimate and the ground truth survival function.

The system identification, which would be done by comparing system posterior probabilities in Bayesian decision making, has to be done in IP setting. As can be

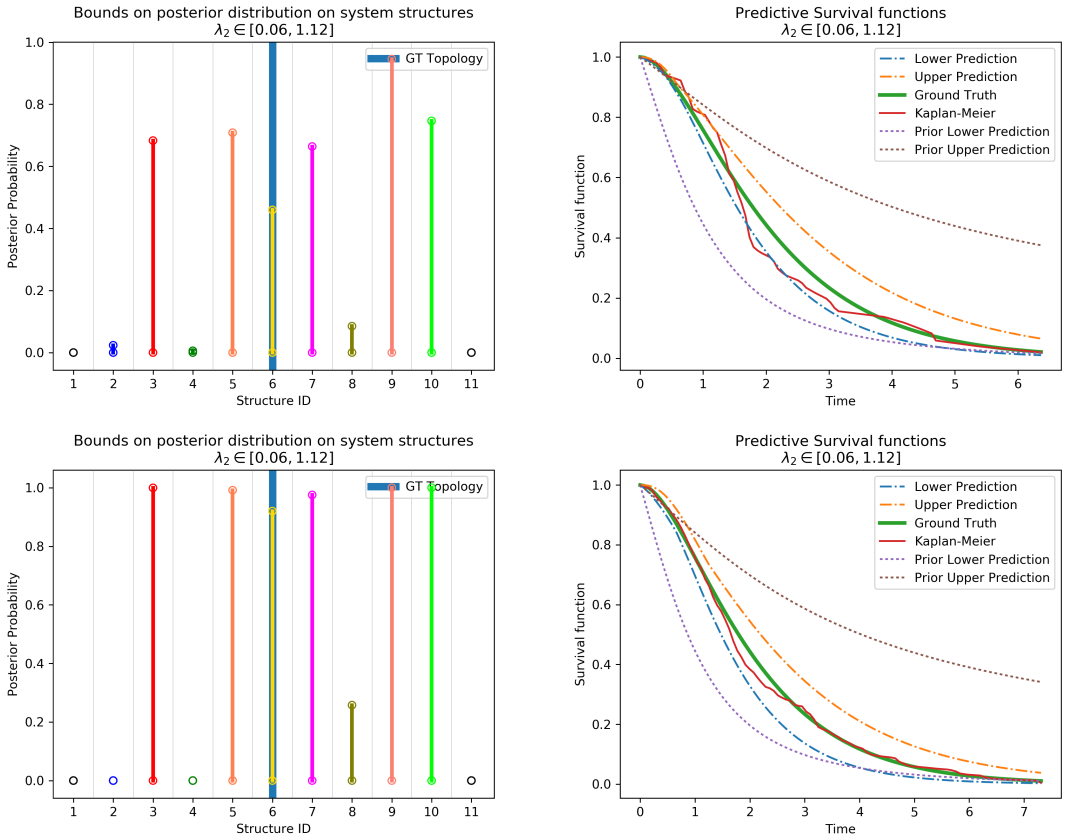


Figure 5.2: Inference results with imprecise distribution parameter λ_2 for sample sizes 50 (top) and 250 (bottom). Left: imprecise posterior distribution on systems. Right: predictions of system lifetime survival function.

seen in Figure 5.2, left, upper probabilities for multiple systems approach 1 in this experiment whilst the lower remain near 0. Therefore, there are several systems for which we are indecisive. The explanation of this wide range is illuminated in Figure 5.3, where we plot system posterior distributions obtained for various fixed λ_2 by standard Bayesian inference (i.e. inner function which is optimized in Equation 5.3). In different regions of Ω_Θ, one system becomes dominant over others and this effect is further increased with increasing sample size.

We can observe that an useful informative inference, we might obtain in IP setting, is that of rejection of several system structures. As is apparent from Figure 5.2 and also from Figure 5.4, upper posterior probability for some of the systems tends to approach 0, which indicates their unfitness to observations.

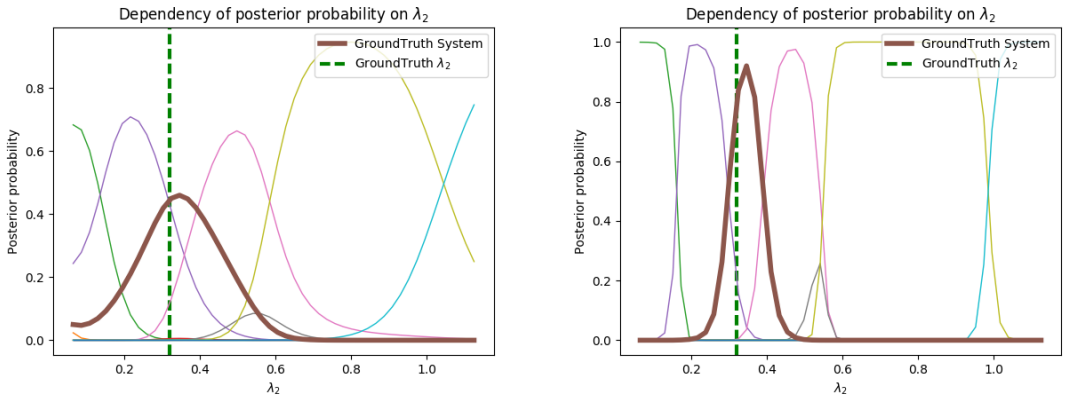


Figure 5.3: Dependency of system structure posterior distributions on fixed λ_2. Each vertical slice at selected λ_2 represents system posterior distribution (i.e. sums to 1). Left image for 50 data samples, right for 250. Thick curve denotes the evolution of posterior distribution of the ground truth system.

5.1.8 Response to varying the support

Next example investigates differences between disjoint choices of underlying support set Ω_Θ. We perform two imprecise inferences separately for λ_2 support divided by the value of (known) λ_1. The resulting imprecise system posteriors are shown in Figure 5.4.

From Figure 5.4 it is apparent, that some structures like 3 and 10, which were comparable by the means of inference in original support set (Figure 5.2, left), exhibit significant differences in case when the support is focused because the likelihood of these systems is small in these regions (see Figure 5.3). Similar behaviour was also observed in case of simply narrowing the λ_2 support where upper posterior probability of many systems approached 0. These results are being omitted here due to space limitations.

This scenario might be applicable for purposes of experimental design towards inference about adversarial systems. Proper choice of the support set Ω_Θ, and therefore the experimental settings, seems to influence identifiability of underlying unknown system structure.

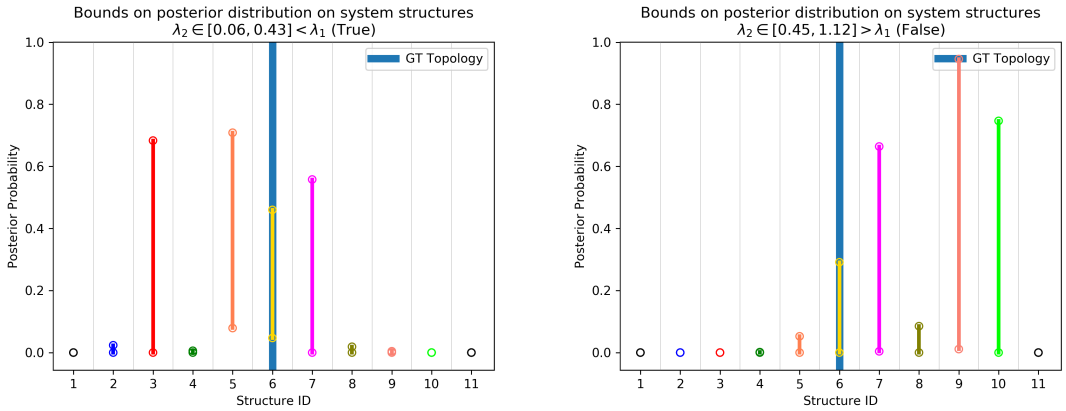


Figure 5.4: Influence of choice of the support set for λ_2 on structure posterior distribution. In the left picture, the GT λ_2 lies in investigated set, in the right one it does not.

5.1.9 Concluding remarks

We have demonstrated a novel methodology for inference in limited prior knowledge scenario, which allows us to avoid introducing some redundant and possibly unjustified modelling assumptions.

For the described situation, we have shown that the optimized functions of interest are nor monotone nor convex and, so far, have to be solved by general optimization procedures (in Section 5.1.6).

It has also been indicated in Section 5.1.7 and Section 5.1.8, that IP inference cannot generally serve for proper system identification, as IP reasoning allows for indecisiveness, but rather as a tool for system rejection in case of low upper posterior probability.

The behaviour which was presented was observed among multiple experiments that were conducted, although no analytical guarantees may be given at this stage of research. A follow-up generalizing study which would take into account even aspects which were only touched here (symmetrical properties of systems and rigorous IP decision theory) is necessary to further understand advantages and limitations of proposed methodology.

5.2 Phased mission systems with interval parameters

This section is a continuation of the contribution [1] presented at SECESA2018 conference. The objective is to introduce extension to phased mission model from Section 4.2 which include epistemic uncertainty on component failure rates and phase duration. In this section we present results which allow us to assess reliability of a PM system with epistemic uncertainty about component failure rates and phase durations.

5.2.1 Imprecision

Engineering, and also other applied fields, have to take uncertainties in our models into account in order to produce reliable conclusions. The first step is to reflect possible uncertainties in the value of parameters of our models. Although the uncertainty may be modelled by stochastic models and inferred via statistical methods, often we face a situation in which we need to include an additional, strong, assumption to be able to carry out the inference (e.g. random censoring). This may bias our analyses and make them unreliable. In cases of limited information (or also for the purpose of sensitivity analysis), interval models may be employed, which model our uncertainty of a parameter value by an interval of its possible range. Imprecisions in parameters of stochastic models are studied within the theory of Imprecise Probabilities [112, 114].

In this section, we will derive novel results which take into account interval uncertainty in the duration of mission phases and component failure laws. We will restrict ourselves to the specific scenario in which the sub-systems, representing individual phases, are coherent and the failure probability of components is modelled by exponential distributions, i.e. a constant failure rate law. Restriction to constant failure rate models provides us with a necessary simplification of the expression for

PM system reliability while still allowing us to address a wide class of real world situations.

5.2.2 System properties

We investigate a system with the following properties:

- structure function of each of the phases corresponds to a structure function of some coherent system. I.e. each $\varphi_i(.)$ is non-decreasing in x and therefore $\varphi_{PM}(.)$, a product of non-decreasing functions, is also non-decreasing in $(x(\tau_1), \ldots, x(\tau_W))$ (in the sense of partial ordering of component states),
- hazard rate of components of group j in phase i is constant and equal to λ_{ij}. I.e. the conditional probability from Equation 4.16 will take the form $p_{ij} = \exp(-\Delta_i \cdot \lambda_{ij})$, where $\Delta_i := (\tau_i - \tau_{i-1})$ is the duration of phase i.

5.2.3 Monotonicity properties

Theorem 5.2.1 (Monotonicity in hazard)**.** *Reliability of a PM (Equation 4.15) satisfying assumptions from Section 5.2.2 is non-increasing function of* $\lambda := (\lambda_{11}, \ldots, \lambda_{WK})$.

Proof.

1. All the RVs representing component states are binary, therefore:

$$Pr(X > x) = \begin{cases} 0 & ; x \geq 1 \\ 1 & ; x < 0 \\ Pr(X = 1) & ; x \in [0, 1) \end{cases} \tag{5.6}$$

2. For each component i

$$Pr(X_i(t) = 1) = Pr(T_i > t), \tag{5.7}$$

where T_i is the component failure time and $Pr(T_i > t) = \exp(-H_i(t))$, where

$H_i(t)$ is the hazard function of component i.

3. $H_i(t) = \sum_{j=0}^{\rho(t)} \lambda_{j,c(i)}\Delta_j$ for each time, which coincide with end of some of the mission phases. $\rho(.)$ is a mapping selecting which one is that, and $c(.)$ select to which group the component belongs.
4. Hazard rate is strictly increasing in $\lambda_{j,c(i)}$, and, therefore, $Pr(X_i(t))$ is strictly decreasing.
5. Given two processes $\tilde{X}(t), X(t)$ of component states, subjected to failure rates $\tilde{\lambda}, \lambda$, respectively, s.t. $\tilde{\lambda} \geq \lambda$, $\tilde{X}(t)$ is stochastically dominated by $X(t)$ at each time corresponding to an end of a phase.
6. Since PM reliability is an expectation of a monotone function, the expectation is lower or equal for the dominated process.

□

Theorem 5.2.2 (Monotonicity in time)**.** *Reliability of a PM Equation 4.15 satisfying assumptions from Section 5.2.2 is non-increasing function of* $\Delta := (\Delta_1, \ldots, \Delta_W)$.

Proof.

1. The reliability of a PM effectively depends on phase time through the conditional probability of component failure during a phase, $p_{ij} = \exp(-\Delta_i \cdot \lambda_{ij})$ (from Equation 4.16).
2. For any altered system with $\tilde{\Delta}_i = \alpha_i\Delta_i$, we can construct an auxiliary PM system with equivalent expression for reliability by, instead, assuming that $\tilde{\Delta}_i = \Delta_i$ and $\tilde{\lambda}_{ij} = \alpha_i\lambda_{ij}, \forall j$.
3. $\forall i : \alpha_i \geq 1 \Rightarrow \forall i, j : \tilde{\lambda}_{ij} \geq \lambda_{ij} \Rightarrow \tilde{\lambda} \geq \lambda$.
4. Reliability of PM is non-increasing in λ, therefore reliability of the altered system is lesser or equal to that of the former one.

□

5.2.4 Implications to interval analysis

We have proved, for our specific scenario, that the reliability of such a PM is monotone in both failure rates and phase durations, so interval analysis can be carried out at the cost of just two precise probabilistic analyses. If we are interested in assessing reliability of such PM across a set of plausible failure rates $\lambda \in \Omega_\lambda$ and phase durations $\Delta \in \Omega_\Delta$, the lower bound can be attained in the upper extreme corner of $\Omega_\lambda \times \Omega_\Delta$, i.e.:

$$\min_{(\lambda,\Delta)\in\Omega_\lambda\times\Omega_\Delta} Rel_{PM}(\lambda,\Delta) = Rel_{PM}\left(\overline{\lambda},\overline{\Delta}\right), \tag{5.8}$$

where $\overline{\lambda} := \max\{\lambda \in \Omega_\lambda\}$ and $\overline{\Delta} := \max\{\Delta \in \Omega_\Delta\}$.

Similarly like in Equation 5.8 for the upper bound, i.e.:

$$\max_{(\lambda,\Delta)\in\Omega_\lambda\times\Omega_\Delta} Rel_{PM}(\lambda,\Delta) = Rel_{PM}\left(\underline{\lambda},\underline{\Delta}\right), \tag{5.9}$$

where $\underline{\lambda} := \min \lambda \in \Omega_\lambda$ and $\underline{\Delta} := \min \Delta \in \Omega_\Delta$.

Corollary 5.2.1 (Using survival signatures.)**.** *For a PM system modelled via survival signatures, the lower bound on reliability is attained for respectively the largest failure rates of each component group and largest plausible phase durations. Analogically for the upper bound.*

5.2.5 Conclusions

We have provided theorems which allows us to assess reliability of phased missions subjected to epistemic uncertainty in both component failure rates and phase durations. Reliability of a special case of phased mission was shown to be a monotone function of these both (Theorem 5.2.1 and Theorem 5.2.2). This allows us to calculate lower and upper probabilities of successful completion of a mission with almost no further computational effort (Equation 5.8). The main contribution to computational complexity arises from calculating the survival signature and has to be done only once, i.e. the same amount as for assessing reliability of precisely specified system.

5.3 Fiducial inference with coarse data

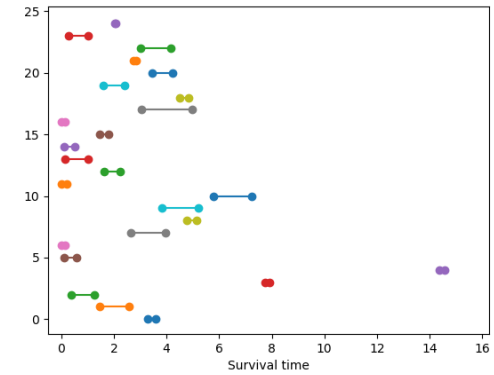
The section is inspired by Scott Ferson's talks on p-box inference with imprecise observations (his full report can be found in [115]) enriched by an application to system reliability with survival signatures.

It is often the case that the data we want to draw inferences from observations that are not precisely specified. This may be due to irreducible measurement errors or due to unspecified censoring. Coarse data are often handled by introducing assumptions on the censoring procedure, like censoring at random. This still requires introduction of additional modelling assumptions. In this section we will demonstrate how inference can be carried out without the assumptions about censoring by constructing a posterior random set. Theoretical properties of direct set-value observation inference are described in [87]. We will demonstrate a straightforward application of empirical random sets for system reliability assessment.

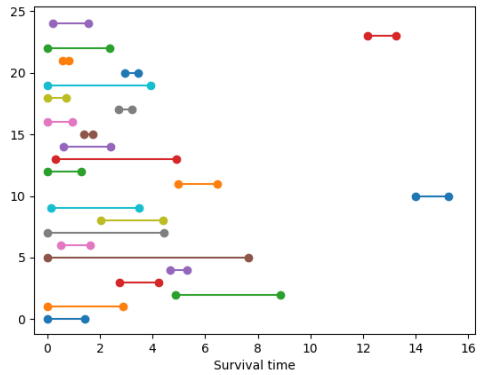
The method mimics using Monte Carlo method for constructing fiducial posterior distribution and propagating it to obtain inferences of interest. Empirical random sets consist of a collection of synthetic set-valued samples which may be used to estimate probability bounds on events of interests or even lower and upper expectations via Choquet's integration (Theorem 3.3.1).

We will assume that the sample is a set of i.i.d. observations where each observation may be specified by an interval (i.e. $x_i \in [\underline{x}_i, \overline{x}_i]$).

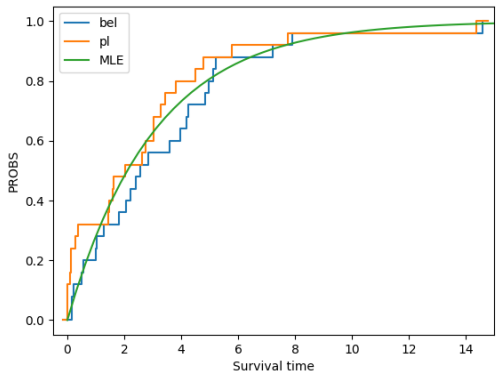
An example of two set-valued samples, which differ in the extent of coarsening, are shown in top row of Figure 5.5. The example was created using randomly sampled observations from exponential distribution with mean value 3.14. The coarsening is performed by randomly sampling the imprecision width from a uniform distribution with maximum value 1 (low imprecision) or 5 (high imprecision). We would expect that the difference in precision of the observation will be reflected in the inference result.



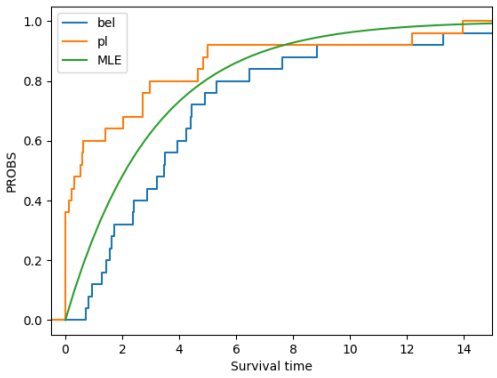
(a) Observed data with low imprecision.



(b) Observed data with high imprecision.



(c) Empirical distribution and MLE with censoring of low imprecision samples.



(d) Empirical distribution and MLE with censoring of high imprecision samples.

Figure 5.5: Observed data and inferred distributions.

5.3.1 Fiducial inference

Extension of the fiducial procedure (Section 3.4.3.1) to set-valued observations is straightforward. The idea behind fiducial inversion is to, for each sampled $u \sim U$, construct a set of plausible parameters θ. An extension towards set-valued observations can be constructed by considering a union of plausible parameters for all the particular values from the observed set sample instead.

Therefore, for X being the observed set,

$$\Theta(X, u) := \bigcup_{x \in X} \Theta(x, u). \tag{5.10}$$

The empirical random set is constructed by random sampling of the ancillary distribution U and constructing a collection of $\Theta(X, U)$.

The inferred posteriors may be used to also obtain assessments about transformed variables - like predictive inference. Empirical random sets (Monte Carlo) may be used for numerical propagation.

Let us have confidence structures for X, Y, constructed from ancillary RVs U_1, U_2, and a new RV $Z = g(X, Y)$. If X, Y are stochastically independent, we may construct a Cartesian product $XY(\mathbf{u}) = X(u_1) \times Y(u_2)$ and a confidence structure via Equation 5.11.

$$Z(\mathbf{u}) = \{g(x, y) : (x, y) \in XY(\mathbf{u})\}. \tag{5.11}$$

5.3.2 Exponential model

We will demonstrate how to apply this inference for a reliability problem. We consider the sample to be exponentially distributed and proceed to infer its mean value.

Let $X_i \overset{i.i.d.}{\sim} exp(\lambda)$. Therefore $\frac{\sum_i X_i}{\lambda} \sim \gamma(n, 1)$.

We can choose an ancillary random variable $U \sim \gamma(n, 1)$ which leads to the fiducial inversion on the exponential mean $\lambda = \frac{\langle X \rangle}{u} = \Theta(x, u)$.

If X_i are imprecise, so will be the sample mean, and we can set

$$\Theta(x, u) = \left\{ \lambda : \frac{\langle X \rangle^L}{u} \leq \lambda \leq \frac{\langle X \rangle^U}{u} \right\}, \tag{5.12}$$

where the supersripts L,U correspond to lower and upper bounds respectively.

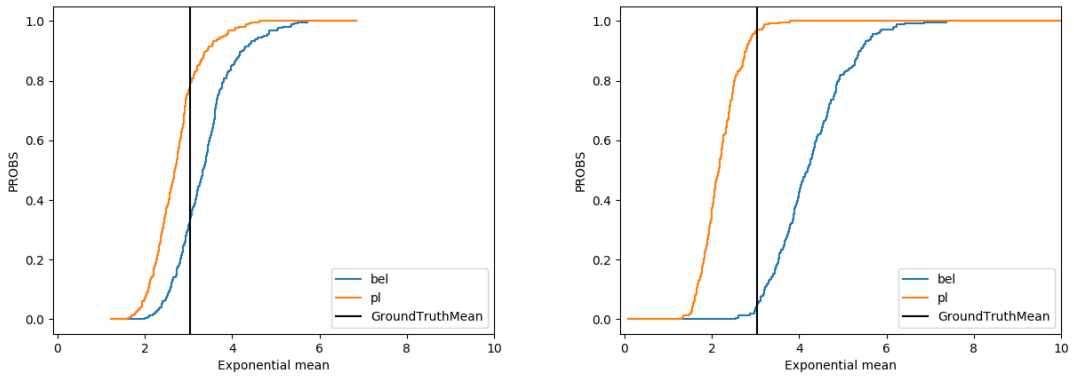


Figure 5.6: Fiducial posterior distribution for exponential mean for the samples with different extent of coarsening corresponding to Figure 5.5.

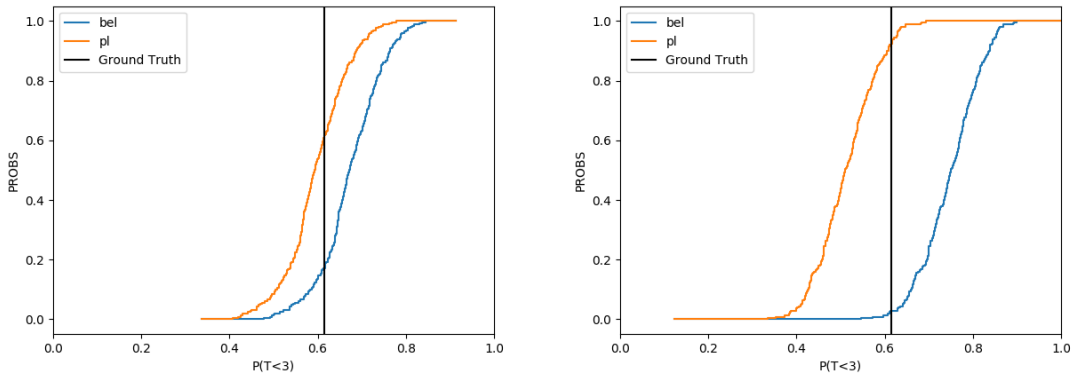


Figure 5.7: Posterior distribution of failure probability at $t = 3$ for the samples with different extent of coarsening corresponding to Figure 5.5.

Constructed fiducial posteriors of exponential mean for the samples depicted in Figure 5.5 are depicted on Figure 5.6.

Furthermore, the constructed posterior can be used to compute various inferences of interest. Figure 5.6 shows a transformed inference on failure probability at time $t = 3$ constructed by propagating the empirical fiducial posterior through analytical expression for exponential CDF.

The posteriors can also be used to infer reliability of systems. For a 5-component bridge system in Figure 5.8 assuming i.i.d. distributions of the component lifetimes, $P(S|\lambda) = \sum_{i=0}^{5} b_i \exp(-\lambda t)^i (1 - \exp(-\lambda t))^{(5-i)}$, where b is the survival signature of the system.

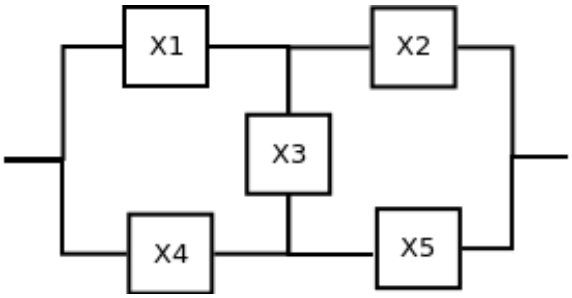


Figure 5.8: Reliability block diagram of a simple bridge system.

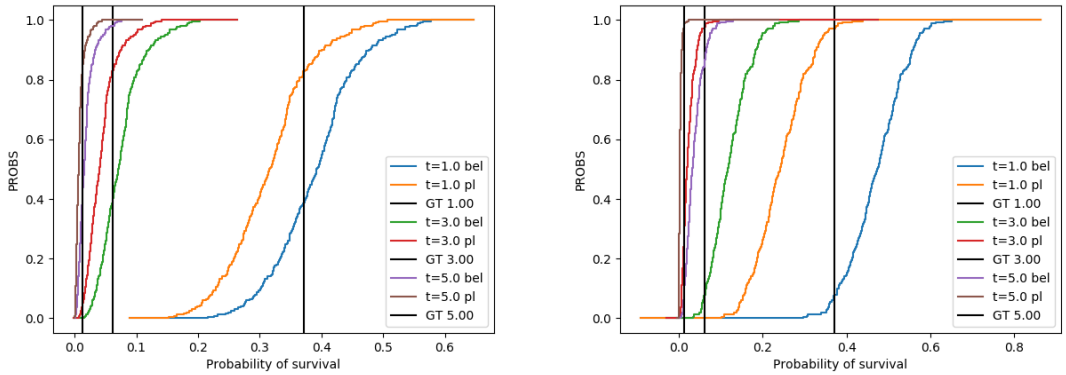


Figure 5.9: Predictive inference of system failure probabilities at selected times for observations corresponding to Figure 5.5.

Inferences of system failure probabilities at selected times are depicted at Figure 5.9.

5.4 Robust Markov analysis

The following section is an exerpt of a conference contribution submitted to IDT2019, [4].

Markov analysis is a wide-spread tool for modelling interactions among components in complex systems. It is based on modelling the evolution of system's component states by Markov Chains. But, as in many other uncertainty models, it might often be overly optimistic to assume that we can construct a precise stochastic model which properly captures the uncertainties present in the investigated system. This issue is addressed by the theory of Imprecise Probabilities and, specifically for stochastic processes, by the theory of Imprecise Markov Chains. In this section, we will demonstrate how Imprecise Markov Chains can not only serve as a robust

alternative to classical stochastic models, but also how they can facilitate analyses by the means of problem dimension reduction and also, by the means of deliberate model construction, enable analyses which would not be possible by using solely precise probability models.

5.4.1 Introduction

Modelling reliability of a complex systems is a daring task. One of the difficulties lies in properly capturing the interactions among the system components during the system's functional lifetime. In order to model these complex interactions, it is necessary to use flexible models to capture the dynamical evolution of the system component states. Markov analysis is used for this purpose. The stochastic evolution is modelled by (Continuous Time) Markov Chains [9, 116]. But the introduced complexity poses a challenge to properly capture all the possible interactions. We will therefore demonstrate the use of a framework for statistical inference of the Markov Models in order to ground the models by empirical evidence, and to robustify them by using models from the theory of Imprecise Probability (IP; [81, 112]).

One way of viewing the IP theory, is from the point of view of sensitivity analysis [78]. Instead of requesting to specify a single stochastic model, or a single sampling distribution, or a single prior distribution for Bayesian inference, we seek ways to analyse a whole set of them at once. Once working with a set of models, we lose the precision of the assertions we are used to from classical uncertainty analysis. Instead, we can infer bounds on the assertions of interest, guaranteed to include the correct answer (as long as the correct model is included in the set of assumed laws). The general aim of inference in the IP theory is to assess lower and upper bounds on functionals of interest (probabilities of events and expected values). For a set of probability measures $\mathbb{P}$ and some random variable Y, we define the lower and upper expectations as

$$\underline{\mathbb{E}}Y := \inf_{P\in\mathbb{P}} \mathbb{E}_P Y, \qquad \overline{\mathbb{E}}Y := \sup_{P\in\mathbb{P}} \mathbb{E}_P Y. \tag{5.13}$$

The analogies for lower and upper probabilities follow from the representation of events by indicator functions.

Due to construction of the IP theory, we only need to focus on either lower or upper expectation since, for so-called coherent models, one can be derived from the other via conjugacy relation $\overline{\mathbb{E}}[Y] = -\underline{\mathbb{E}}[-Y]$.

In this section, we will focus on an application of results from a sub-field of IP theory, on theory of Imprecise Markov Chains [117]. Recent development concluded in efficient computational schemes which enable us to conduct robust inferences on stochastic dynamical systems [118]. An outline of the numerical methods will be presented in Section 5.4.4. It has also been recently discovered how to perform Robust Bayesian Inference on the characteristics of the process dynamics [119], as will be shown in Section 5.4.4.1. Imprecise Markov Chains can also be used to reduce complexity of high-dimensional problems by the technique of lumping. This is not unknown to precise Markov Chain models, but its imprecise variant provides more flexibility because the lumping generally results in an Imprecise Markov Chain model [120]. This will be shown in Section 5.4.5.2.

Our selected application field is Reliability Theory. We will combine the above-mentioned advancements in the field of Imprecise Markov Chains with modelling techniques of Reliability Theory. The state-space lumping can be naturally combined with Survival Signatures, which also serve as a dimension reduction technique [37]. These will be covered in Section 5.4.5.1 and in Section 5.4.5 in general. We will further use the flexibility of the IP theory in order to propose a novel method to 1) enable statistical inference for Markov Reliability analysis for realistic scenarios at all, and 2) include additional high-level information into the models in order to decrease the imprecision of the answers. This will be explained in Section 5.4.6 and demonstrated on an example of reliability inference from a collection of observations for a simple system in Section 5.4.7.

5.4.2 Survival analysis

Many systems are engineered such that they are functional at the instant of deployment but their components will deteriorate over time. In such scenarios, we investigate how the system reliability changes in time. The component states are here modelled as stochastic processes $X_i(t)$ and the joint states as $X(t)$ for some $t \geq 0$. The simplest case is when all the X_i are non-increasing processes (once a component fails it will not recover its functionality). Then we can equivalently model just the time to the component's failure, say T_i, and define the stochastic process as $X_i(t) = \{T_i > t\}$.

The dynamics of the components' states will induce a stochastic process on the state of the system. Two basic assessments about the system are then usually of interest, the system availability $\mathcal{A}(t) = \mathbb{E}[\varphi(X(t))]$ (the probability that a system is operational at a given time instant, regardless of its history) and the system reliability $Rel(t) = P(\forall \tau \leq t : \varphi(X(\tau)) = 1)$, also called the survival function (the probability that the system has not failed before the enquired time instant). These two assessments are equivalent if the system is coherent and the component state processes are non-increasing.

5.4.3 Markov analysis

The evolution of component states may take complex form and include interaction among the components states and their deterioration rates. We will model the joint component state processes by Markov Chains on the joint component state-space since they can capture various interactions among component state dynamics (load-share system, spare components, component renewals, etc [9]).

In this section, we will focus on the Markov method for modelling the time evolution of component states. We will revise some of the basic elements of Markov processes, numerical methods for evaluating the availability of the system, and a prevision

centred formulation for evaluating functionals . $f \to \mathbb{E}[f(X(t))], t \in \mathbb{R}^+$, such as the availability. The space $\mathcal{X}$ is assumed to be finite. Instead of a single random variable for each of the components, we will now model the state evolution of each component as a stochastic process $X_i(t)$. For fixed t, $X(t)$, the joint state of all the system components, is a random vector with state space $\mathcal{X}$. Since the component states evolve over time, so will the system state and its reliability. In time analysis, we will be interested in the availability function $\mathcal{A} : t \to \mathbb{E}[\varphi(X(t))]$. $\mathcal{A}$ denotes the probability that system is functional if its service is required at time t.

If the component state process $X(t)$ is non-increasing in t, then we can use $\mathcal{A}(t)$ to express the distribution of system's time to failure

$$\begin{aligned} \mathcal{S}(t) &:= P(T_{sys} \geq t) = P(\{\omega | \forall \tau < t : \varphi(X(\tau)(\omega)) = 1\}) \\ &= P(\{\omega | \varphi(X(t))(\omega) = 1\}) = \mathcal{A}(t). \end{aligned} \tag{5.14}$$

We will be further interested in looking at stochastic processes X as a families of functionals $(\mathcal{X} \to \mathbb{R}) \ni f \to \mathbb{E}[f(X(t))], t \in \mathbb{R}^+$. Such as $\mathcal{A}$.

5.4.3.1 Continuous time Markov chains

Describing a random process corresponds to describing a random variable with an infinite dimensional co-domain, the functions. Markov chains provide a rich class of processes which can be described consistently and manipulated efficiently. We can evaluate functional $f \to \mathbb{E}[f(X(t))]$, if we know the distribution of $X(t)$.

The distribution $X(t)$ at any time point is uniquely characterised by an initial distribution $\pi_0(x) := P(X(0) = x)$ and a family of conditional distributions $T_s^t := P(X(t) = y | X(s) = x)$. The distribution of $X(t)$ at any $t \geq 0$ may be obtained via the law of total probability as $P(X(t) = y) = \sum_{x \in \mathcal{X}} P(X(t) = y | X(0) = x)\pi_0(x)$.

A **homogeneous continuous time Markov chain** (HCTMC) parametrized by a transition rate matrix Q (a stochastic matrix, i.e. each row sums to 0 and all non-diagonal elements are non-negative) induces a family of transition matrices $\{T_s^t : 0 \leq s \leq t\}$ such that

$$T_s^t(x, y) = \exp(Q(t - s))(x, y). \tag{5.15}$$

Together with an initial distribution π_0, Equation 5.15 specifies

$$P(X(t) = y) = \sum_{x \in \mathcal{X}} \pi_0(x) \exp(Qt)(x, y). \tag{5.16}$$

Let $\mathscr{L}(\mathcal{X})$ be a space of all functions $\mathcal{X} \to \mathbb{R}$. We may also look at the transition matrix as an operator $T_s^t : \mathscr{L}(\mathcal{X}) \to \mathscr{L}(\mathcal{X})$.

For a function $f \in \mathscr{L}(\mathcal{X})$ and $x \in \mathcal{X}$, denote

$$[T_s^t f](x) := \mathbb{E}[f(X(t))|X(s) = x]. \tag{5.17}$$

In the case of HCTMC, due to the homogeneity, $T_s^t f = T^{t-s} f$, where we will omit the subscript if it is zero, so $T^{t-s} := T_0^{t-s}$.

5.4.3.2 Evaluating functionals

There exist multiple ways for evaluating $\mathbb{E}[f(X(t)|X(0)]$. The commonly used one is based on the additive property of probability measures. It consists of computing the conditional probabilities $P(X(t) = y|X(0))$ for $y \in \mathcal{X}$. This allows us to construct $T^t f \in \mathscr{L}(\mathcal{X})$ as $T^t f(x) = \sum_{y \in \mathcal{X}} f(y) P(X(t) = y|X(0) = x)$. The expected value of $f(X(t))$ will be $\mathbb{E}[T^t f(X(0))]$. But this method is not applicable if we model the stochastic process by a non-additive measure. One way would be obtain a similar results as a Choquet integral $\mathbb{E}[f(X(t))|X(0) = x] = (C) \int f dP_{X(t)|X(0)=x}$, but we will use a different method in this section, which also enables us to use efficient computational methods introduced by [117] and [118] for non-additive Markov Processes.

Due to its definition, Equation 5.15, we can decompose any transition operator into a concatenation of partial operators. For arbitrary times $t > s > r \geq 0$,

$$\left[T_r^t f\right](x) = [T_r^s \circ T_s^t f](x). \tag{5.18}$$

Let $s = u_0 < u_1 < \ldots < u_n = t$ be a discretization of the time domain. Then combining Equation 5.15, Equation 5.17 and Equation 5.18, by induction,

$$\left[T_s^t f\right](x) = \left[T_s^{u_1} \circ \ldots \circ T_{u_{n-1}}^t f\right](x) = \left[T_s^{u_1} f_1\right](x), \tag{5.19}$$

where $f_n = f$ and $f_i = T_{u_i}^{u_{i+1}} f_{i+1}$. Therefore, we can evaluate $T_s^t f$ recursively by evaluating f_i, starting from $f_n = f$.

We can approximate the inference by a linearized scheme up to an arbitrary precision, similarly as we do with the Euler method for solving ordinary differential equations. Given the time short is small enough,

$$\begin{aligned} &\forall x \in \mathcal{X}: \\ &\quad f_n(x) = f(x) \\ &\quad \ldots\ldots \\ &\quad f_i(x) \approx \left[(I + (u_{i+1} - u_i)Q)\, f_{i+1}\right](x) \\ &\quad \ldots\ldots \\ &\mathbb{E}[f(X(t))|X(0) = x] \approx \left[(I + (u_1 - u_0)Q)\, f_1\right](x), \end{aligned} \tag{5.20}$$

where I denotes the identity operator.

Denote $L_s^t := \lim_{n\to\infty}(I + \frac{t-s}{n}Q)^n$, where by the exponent n we denote n operator concatenations. Then, under some regularity conditions, $L_s^t \to T_s^t$ [117].

5.4.3.3 An example

Suppose we have a transition diagram given by Figure 5.10 with state space $\mathcal{X} = \{0, 1\}^3$ and want to evaluate $P(X \in A = \{111, 101, 011\})$. If P is an additive

measure, then

$$P_X(A) = \sum_{y \in \mathcal{X}} \mathbb{I}_A(y) P_X(y) = \sum_{y \in A} P_X(y), \tag{5.21}$$

where $\mathbb{I}_A$ denotes the indicator function of set A ($\mathbb{I}_A(x) = 1$ if $x \in A$ and 0 otherwise). For fixed t, we can evaluate $P(X(t) = x)$ via Equation 5.16 and also $P(X(t) \in A | X(0) = x)$ via Equation 5.21 as

$$P_{X(t)|X(0)=x}(A) = \sum_{y \in A} T^t(x, y) = \sum_{y \in A} \exp(Qt)(x, y). \tag{5.22}$$

We can also proceed via the functional representation and use Equation 5.17. In Equation 5.21 we have used that $P_X(A) = \mathbb{E}[\mathbb{I}_A(X)]$. In the functional representation, $P_{X(t)|X(0)=x}(A) = (T^t \mathbb{I}_A)(x)$. We can assess $P(X(t) \in A | X(0) = x)$ also as $\mathbb{E}[\mathbb{I}_A(X(t)) | X(t(0)) = x]$. This can be solved numerically via a linearized discretization scheme based on Equation 5.20.

5.4.4 Imprecise Markov chains

Suppose that we cannot determine a single transition rate matrix Q and instead have a set of 'plausible' matrices $\mathscr{Q}$. Set $\mathscr{Q}$ induces a set of stochastic processes $\mathbb{P}$ and we may use these as our model of the system evolution. For computing expectations of $f \in \mathscr{L}(\mathcal{X})$ over a set of stochastic processes, we need to precisely specify, which processes are present in the set of measures $\mathbb{P}$. If $\mathbb{P}$ is a set of HCTMC with transition rates in $\mathscr{Q}$, then the lower expectation Equation 5.13 is defined, in accordance with Equation 5.13, as

$$\underline{\mathbb{E}} f(X(t)) = \inf_{Q \in \mathscr{Q}} \sum_{x \in \mathcal{X}} P(X(0) = x) \sum_{y \in \mathcal{X}} \exp(Qt)(x, y) f(y), \tag{5.23}$$

and similarly for the upper expectation.

But the optimisation problem in Equation 5.23 is generally intractable. Instead, we can take $\mathbb{P}$ to be a larger set of stochastic processes which comply with our knowledge. A stochastic process (not necessarily a HCTMC, see [117]) is said to be consistent

with $\mathscr{Q}$ if for each $t \in \mathbb{R}_0^+$

$$\lim_{\Delta \to 0} \frac{P(X(t+\Delta))T_t^{t+\Delta} - I}{\Delta} \in \mathscr{Q}. \tag{5.24}$$

This means that $\mathscr{Q}$ encapsulates transition rates locally, but the rates may differ over time.

Taking $\mathbb{P}$ to be a set of stochastic processes consistent with $\mathscr{Q}$, we can construct a family of sets of transition operators $\mathscr{T}_s^t$ such that

$$\begin{aligned} \left[\underline{T}_s^t f\right](x) &:= \underline{\mathbb{E}}[f(X(t))|X(s) = x] \\ &= \left[\inf_{T \in \mathscr{T}_s^t} Tf\right](x) \\ &= \lim_{n \to \infty} \left[\left(I + \frac{t-s}{n}\underline{Q}\right)^n f\right](x), \end{aligned} \tag{5.25}$$

where $\underline{Q} : \mathscr{L}(\mathcal{X}) \to \mathscr{L}(\mathcal{X})$ is the lower transition rate operator defined for each $x \in \mathcal{X}$ as

$$\left[\underline{Q}f\right](x) := \inf_{Q \in \mathscr{Q}} [Qf](x). \tag{5.26}$$

We can decompose the lower transition operator similarly to Equation 5.18. With $s < t$,

$$\underline{\mathbb{E}}[f(X(t))|X(0) = x] = \left[\underline{T}^s \circ \underline{T}_s^t f\right](x), \tag{5.27}$$

which would not hold if we would take only HCTMCs, since the transition rate matrices would need to be identical for $0 \to s$ and $s \to t$, but here we allow them to differ as long as they are consistent with $\mathscr{Q}$.

Similarly as Equation 5.18 and Equation 5.19 in the last section, Equation 5.27 allows us to construct an iterative scheme for computing lower expectations of functionals in $\mathscr{L}(\mathcal{X})$ (for more details and convergence conditions see [117] and [118]). For a discretization of the time dimension $s = u_0 < \ldots < u_n = t$,

$$
\begin{aligned}
f_n &= f \\
&\cdots\cdots \\
f_i &\approx \left(I + (u_{i+1} - u_i)\underline{Q}\right) f_{i+1} \\
&\cdots\cdots \\
\underline{\mathbb{E}}[f(X(t))|X(s) = x] &\approx \left[\left(I + (u_1 - u_0)\underline{Q}\right) f_1\right](x).
\end{aligned}
\tag{5.28}
$$

5.4.4.1 Inference of the transition rate operator

In this section, we will demonstrate how statistical inference can be used in order to construct the transition rates governing stochastic processes. We will show some results for inferences about the transition rate matrix of a HCTMC. Especially, we will focus on a Robust Bayesian inference method proposed in [119].

Suppose that we observe the trajectory, say ω, of a Markov chain. Denote d_x the total holding time ($\int \mathbb{I}_x(X(t))dt$) in state x and n_{xy} the number of transitions from state x to state y. For the inference about the transition rate matrix generating the stochastic process, we may use the following likelihood results [121]

$$
\mathcal{L}(\omega|Q) = \prod_{\substack{x,y\in\mathcal{X} \\ x\neq y}} (q_{xy})^{n_{xy}} \exp(-q_{xy}d_x), \tag{5.29}
$$

which leads to the Maximum Likelihood Estimator $q_{xy}^{\text{MLE}} = \frac{n_{xy}}{d_x}$.

5.4.4.2 Robust Bayesian inference

We will further be interested in using Equation 5.29 for Bayesian inference. Because the likelihood belongs to the exponential family, it was proposed in [119] to use a product of gamma priors for the Bayesian inference. For a chosen collection of

hyper-parameters $\alpha = \{\alpha_{xy}, x \neq y \in \mathcal{X}\}$ and $\beta = \{\beta_x, x \in \mathcal{X}\}$, the prior density π is

$$\begin{aligned}\pi(Q|\alpha,\beta) &= \prod_{\substack{x,y\in\mathcal{X}\\x\neq y}} (q_{xy})^{\alpha_{xy}-1}\exp(-q_{xy}\beta_x) \\ &= \prod_{\substack{x,y\in\mathcal{X}\\x\neq y}} \text{Gamma}(q_{xy}|\alpha_{xy},\beta_x).\end{aligned} \tag{5.30}$$

Combining the prior, Equation 5.30, with the likelihood, Equation 5.29, we can pose the exact form for the posterior mean,

$$\mathbb{E}[q_{xy}|\alpha,\beta,\omega] = \frac{\alpha_{xy}+n_{xy}}{\beta_x+d_x}. \tag{5.31}$$

To avoid the necessity to choose any particular prior distribution, we may, instead, select a whole set of them. This method is known as a Generalized Bayesian inference [81] and will generally lead to a set of posterior distribution, hence a set of posterior mean estimates and an Imprecise Markov Chain.

In [119], a set of prior distributions based on the imprecise Dirichlet model (IDM) (introduced in [81]) has been proposed. Among the advantages of the IDM model is the property of prior-ignorance. Before observing any samples, the IDM generates vacuous inferences (all events have lower probability 0 and upper probability 1 - this is impossible to achieve with precise priors).

In combination with Equation 5.29 and Equation 5.31, this model leads to a posterior set of rate matrices

$$Q_s := \left\{Q \in \mathscr{Q}, \text{ s.t. } \left(\forall x,y \in \mathcal{X}, x \neq y : q_{xy} = \frac{sA(x,y)+n_{xy}}{d_x}\right), A \in \mathscr{T}\right\}, \tag{5.32}$$

where $\mathscr{T}$ denotes the set of all transition matrices (conditional probabilities of transitioning from one state in $\mathcal{X}$ to another).

As in the IDM model, an analyst have to specify a pseudo-count parameter controlling the updating sensitivity. It is denoted s in Equation 5.32 and is recommended to

be set between 1 and 2. If set to 0, the model would correspond to the maximum likelihood estimate, Equation 5.29.

The derived posterior set of transition rate matrices leads also to a tractable lower transition rate operator, Equation 5.26. For $f \in \mathscr{L}(\mathcal{X})$,

$$[\underline{Q}f](x) = \frac{s}{d_x} \min_{y \in \mathcal{X}} (f(y) - f(x)) + \sum_{y \in \mathcal{X} \setminus \{x\}} \frac{n_{xy}}{d_x} (f(y) - f(x)). \tag{5.33}$$

These results require that every state had been observed, i.e. that $\forall x \in \mathcal{X} : d_x > 0$. This may be overly restrictive requirement for application of the method in realistic scenarios from reliability theory and will be resolved in Section 5.4.6.

5.4.5 Dimension reduction

In this section we will present some methods which help us to decrease the computational complexity arising from the combinatorial explosion of the system components' state space. Since each component is binary and arbitrary combination of component states is permitted, the cardinality of the state-space $\mathcal{X}$ would be 2^N, where N is the number of components in the systems. This itself often makes it intractable to assess reliability for some real systems even without taking into account the dependencies influencing their deterioration or imprecise probability models.

5.4.5.1 Survival signatures

Survival signatures were introduces to facilitate analyses of large heterogeneous systems [37]. They are used to simplify the prescription of the relation between the system state and the component states. Assuming some simplifying relations among the components interactions, exchangeability of their states, we can compress the remaining relevant aspects of the structure function into a lower dimensional summary. If the exchangeability assumption holds, we can still recover exact inference about the system reliability.

For an arbitrary ancillary random variable L, we can decompose the expected value in #eq-rel-rel_definition via the law of total expectation into

$$\mathbb{E}[\varphi(X)] = \mathbb{E}\left\{\mathbb{E}[\varphi(X)|L]\right\}. \tag{5.34}$$

Suppose that we can group the components of the system such that states of the components of the same group are exchangeable. We may choose the ancillary random variable L to be a vector of numbers of functioning components of each group (arbitrarily ordered) and define the system survival signature as $\Phi(l) = \mathbb{E}[\varphi(X)|L = l]$. Due to the exchangeability assumption, the conditional probability $P(X|L)$ is uniform on the subset of $\mathcal{X}$ for which $L = l$. This allows us to compute Φ according to the laws of classical probability as the fraction of functional states,

$$\Phi(l) = \frac{|\{x \in \mathcal{X} : L(x) = l, \varphi(x) = 1\}|}{|\{x \in \mathcal{X} : L(x) = l\}|}, \tag{5.35}$$

where the notation $L(x)$ represents the dependency of L on the actual component states.

If the uncertainty model is given by an additive measure, $\mathbb{E}[\Phi(L)] = \sum_l \Phi(l)P(L = l)$. But since we intend to carry out inferences with imprecise probability models, which are generally non-additive measures, we will rather work directly with the functional formulation. The (lower) survival function at a particular time point t will simply be $\mathbb{E}[\Phi(L(t))]$, where $L(t) := L(X(t))$ and we can use the scheme in Equation 5.20 or Equation 5.28 to compute the requested value.

5.4.5.2 State-space lumping

Lumping serves as a way to decrease sizes of state-spaces of Markov models by grouping some of the states together. But, if the aim is to obtain a Markov Chain again after the lumping, its uses are limited and the existence of a limped HCTMC is not always guaranteed [122]. In general, the lumped process may not remain Markov or homogeneous after the procedure. Generally, it will become a Homogeneous

(generated by a single set of transition rate operators $\mathscr{Q}$ at each time) Imprecise Markov Chain [120].

Let Q be a transition rate matrix of a (precise) HCTMC with state space $\mathcal{X}$. Let $\Lambda : \mathcal{X} \to \hat{\mathcal{X}}$ be a surjective function mapping states from the original state-space on their respective groups in the new state-space. The evolution of the lumped state $\hat{X}(t) = \Lambda(X(t))$ will be induced by the evolution of state X according to the original HCTMC.

Let $f \in \mathscr{L}(\mathcal{X})$ be *lumpable w.r.t.* Λ, meaning that there exists $\hat{f} \in \mathscr{L}(\hat{\mathcal{X}})$ s.t.

$$\forall x \in \mathcal{X} : f(x) = \hat{f}(\Lambda(x)). \tag{5.36}$$

The transition rate operator $\underline{\hat{Q}} : \mathscr{L}(\hat{\mathcal{X}}) \to \mathscr{L}(\hat{\mathcal{X}})$ of the lumped process can be evaluated as

$$[\underline{\hat{Q}}\hat{f}](\hat{x}) = \min_{x \in \Lambda^{-1}(\hat{x})} \left\{ \sum_{\hat{y} \in \hat{\mathcal{X}}} \hat{f}(\hat{y}) \sum_{y \in \Gamma(\hat{y})} Q(x, y) \right\}. \tag{5.37}$$

5.4.5.3 Combining lumping and survival signatures

Since lumping and signatures share similar principles and goals, they can be combined for the purposes of Markov analysis. Lumping will allow us to simplify the joint state-space of the system components and survival signature will provide a method for constructing a lumpable function which we will use for assessing the system reliability.

In order to utilize the survival signatures, we need to assure exchangeability of the states of the grouped components. Clearly, for any Markov model of the component degradation, we may consider each component of its own distinct type. But such a model would lead to no savings of computational resources. Since we intend to perform statistical inference based on observations of independent copies of a system, we will include assumptions in our statistical model to make the inference possible and tractable.

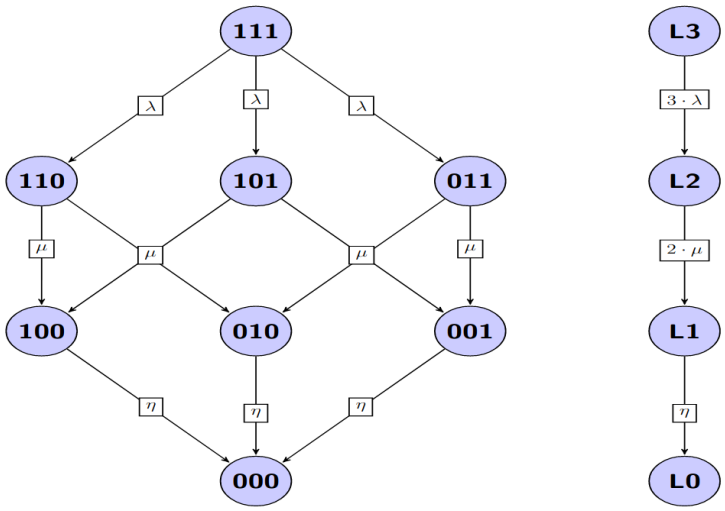


Figure 5.10: On the left: transition diagram of a 3-component load-share system. Edge labels represent the elements q_{xy} of stochastic matrix Q. On the right: transition diagram of the same system after lumping by the number of functioning components.

Let us aggregate the system components into several groups. A simple assumption which will make the combination possible is to assume that the transition rates depend only on the number on functioning components of each group. An example of such Markov model is depicted on Figure 5.10. The example also demonstrates, that this assumption leads to a precise HCTMC after lumping, which we currently need in order to carry out statistical inference as described in Section 5.4.4.1.

This assumption also ensures exchangeability of the component states. If we therefore use $\Lambda(X) = L(X)$, as used in Section 5.4.5.1, as the lumping map, and $\Phi \circ L \in \mathscr{L}(\mathcal{X})$ as the lumping function, we may use the procedures in Equation 5.20 and Equation 5.28 on the lumped process to assess the system reliability ($= \mathbb{E}[\Phi(L(X))]$).

5.4.6 Forced semi-vacuous model with structural assumptions

The drawback of the methods mentioned in Section 5.4.4.1, the MLE and Equation 5.33, is the necessity to observe all the possible states $x \in \mathcal{X}$. Otherwise the null value of the denominator d_x will cause that the expressions will not be properly

defined. Given the cardinality of the state-space in reliability problems and often a low number of observations, it is highly unlikely that all the states will be observed. Although we have demonstrated some methods to reduce the size of the state-space in the Section 5.4.5, it still remains unlikely to observe all the states even from this smaller state-space in practical applications.

Recall the linearized computation scheme for assessing lower expectations of functionals of stochastic processes, Equation 5.28. If we replace the linear approximation by the original expression it approximates, the general step of the procedure may be viewed as

$$f_i(x) = [\underline{T}_{u_i}^{u_{i+1}} f_{i+1}](x). \tag{5.38}$$

While evaluating f_i for each of the states $x \in \mathcal{X}$ during the step, two situations may occur based on the data from which we draw our inferences about the underlying transition rate matrix. Either state x was observed, in which case we may use the original scheme with $\underline{Q}$ given by Equation 5.33. If state x was not observed, then we know nothing about transitions from x, thus nothing about $\mathbb{E}[X(t) = y|X(s) = x]$ for any y, s, t. But we can reflect this ignorance in the IP models simply by calling for the properties of the transition operator for which $\underline{T}f \geq \inf f$.

But a bare inclusion of vacuous model may lead to excessively imprecise inferences. There is still an additional information we can supply to the model in order to improve our inferences. That is by introducing additional structural assumptions. It is not too preposterous, in many reliability applications, to assume that no more than one failure may occur at any time instance. Exceptions could of course be found if we were to include possibilities of e.g. common cause failures. Nevertheless, even in those cases, the method we are going to introduce might still be applied at least partially.

Assume that for each of the states $x \in \mathcal{X}$, we can rule out some states $W(x) \subset \mathcal{X}$ for which we deem the transition $x \to y \in W(x)$ to be impossible. In the case of non-increasing component state processes, this would be e.g. transitions that would renew some component's functionality. If we judge that only one failure can happen at any time instant, then we could rule out all the transitions which would render two or more components non-functional at the same time. In any case, such an assumption will have an effect on both the lower transition rate operator emerging from the robust Bayesian inference (Equation 5.33) and also on the vacuous prevision.

In the selection of the set of prior distributions for the robust inference, leading to the set of posterior rate matrices in Equation 5.32, we can narrow down the set of admissible transition operators $\mathscr{T}$. If we rule out transitions $W(x)$ for state x, we can reflect this in the set of admissible transition operators by selecting $\mathscr{T}_0(x) := \{A(x,y) \in \mathscr{T} : A(x,z) = 0, \forall z \in W(x)\}$. This would lead to an enhanced posterior transition rate operator

$$\begin{aligned}[\underline{\tilde{Q}}f](x) := \frac{s}{d_x} \min_{y \in \mathcal{X} \setminus W(x)} (f(y) - f(x)) \\ + \sum_{y \in \mathcal{X} \setminus [\{x\} \cup W(x)]} \frac{n_{xy}}{d_x} (f(y) - f(x)).\end{aligned} \tag{5.39}$$

The introduced assumption will also affect the vacuous prevision $\underline{T} \geq \inf f$ simply by taking the infimum only over the set of admissible transitions. Hence $[\underline{\tilde{T}}f](x) \geq \inf_{y \in \mathcal{X} \setminus W(x)} f(y)$.

Therefore, we redefine the step in the linear approximation scheme in Equation 5.28 as

$$[\underline{T}_{u_i}^{u_{i+1}} f](x) \approx \begin{cases} [(I + (u_{i+1} - u_i)\underline{\tilde{Q}})f](x), & d_x > 0, \\ \min_{y \in \mathcal{X} \setminus W(x)} f(y), & d_x = 0. \end{cases} \tag{5.40}$$

Now, the inference is always possible.

5.4.7 Example

We will introduce the methodology on a simple example. We assume the system topology as depicted on Figure 5.11, a system with two types of components. Its survival signature in Table 5.2. We include the assumption about exchangeability of the states of components in the same group, so that we can lump the original state space model according to what was described in Section 5.4.5.3. The structure of the lumped HCTMC, including additional assumptions that 1) component state processes are non-increasing, and 2) only one failure may occur at any time instant, is depicted on Figure 5.12. The arrows represent the only admissible transitions.

We simulate several observations of the system lifetime. The simulated experimental procedure produce evolution of the system components' states from which we can extract d_x and n_{xy} needed for the likelihood function, Equation 5.29. The simulation is terminated by the time the system reaches a failure state, therefore we never observe all the states in the state-space.

The inference of the survival function (Equation 2.5) is computed via the linearized scheme (Equation 5.28) as $\underline{Rel}(t) = \underline{\mathbb{E}}[\Phi(L(t))|L(0) = \mathbf{M}] = \left[\underline{T}^t\Phi\right](\mathbf{M})$. $\mathbf{M}$ represents vector of number of components of each type in the system, hence $L(0) = \mathbf{M}$ implies that all the components are functional at time 0. Analogously it is done for the upper bound.

Resulting inferences of the bounds on the survival function are show on Figure 5.13. The inference seems to properly envelope the ground truth (GT) survival function, exhibit great imprecision if the data is scarce (the first plot with 5 observations), and approaches the ground truth survival function as the number of observations increases.

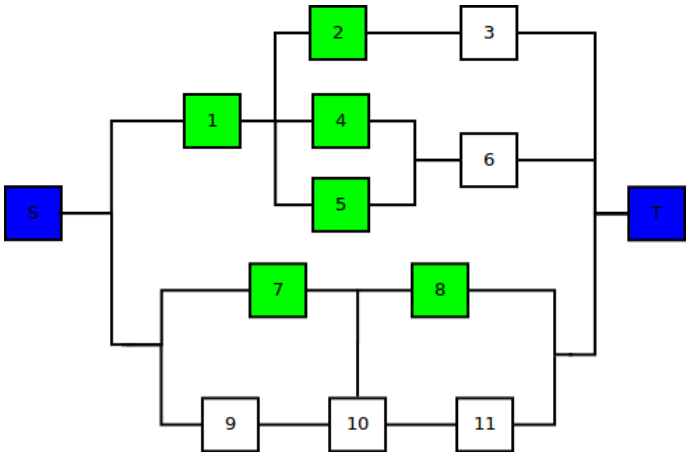


Figure 5.11: Reliability block diagram of the example system. Components of type 1 are represented by the white blocks and components of type 2 are the green ones.

Table 5.2: Survival signature of the example system. Zero contributions are omitted.

l_1	l_2	Φ	l_1	l_2	Φ	l_1	l_2	Φ
3	0	0.1	0	3	0.2	1	5	0.8
4	0	0.4	1	3	0.31	2	5	0.9
5	0	1	2	3	0.47	3	5	1
2	1	0.033	3	3	0.67	4	5	1
3	1	0.166	4	3	0.85	5	5	1
4	1	0.466	5	3	1	0	6	1
5	1	1	0	4	0.4	1	6	1
0	2	0.066	1	4	0.56	2	6	1
1	2	0.1066	2	4	0.74	3	6	1
2	2	0.2	3	4	0.88	4	6	1
3	2	0.3866	4	4	0.9466	5	6	1
4	2	0.66	5	4	1			
5	2	1	0	5	0.6			

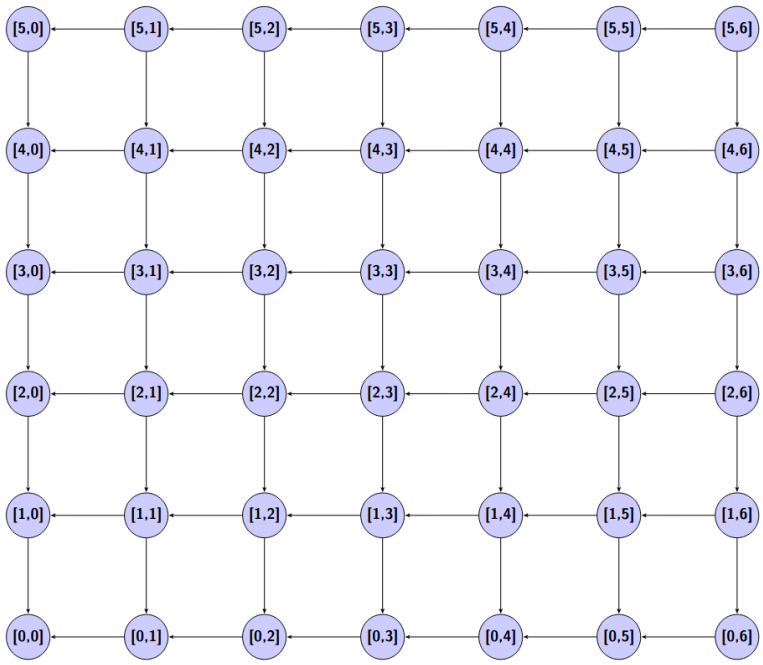


Figure 5.12: Lumped transition diagram of the example system. Only the depicted arrows signify admissible state transitions of the system.

5.4.8 Conclusions

Imprecise Markov Chains present a natural way for conducting Markov analysis in reliability theory. They allow us to robustify our inferences by including multiple plausible models of the systems' dynamics. So far, it seems that they provide the only way to conduct statistical inference in this field because they enable to include vacuous models to substitute for the lack of observations. Imprecise Markov Chains also provide a general lumping framework for dimension reduction for large state-spaces. Combined with the Survival Signature framework, they present a powerful analytic technique for conducting inferences on large - real world systems. Although the method was demonstrated only on a small system, both lumping and survival signatures are being used to enable analyses which would be intractable otherwise due to their high dimensionality. Besides that, Imprecise Probability models allow us to conduct Robust Bayesian inference, which mitigate some of the controversies about selecting prior distributions in Bayesian inference.

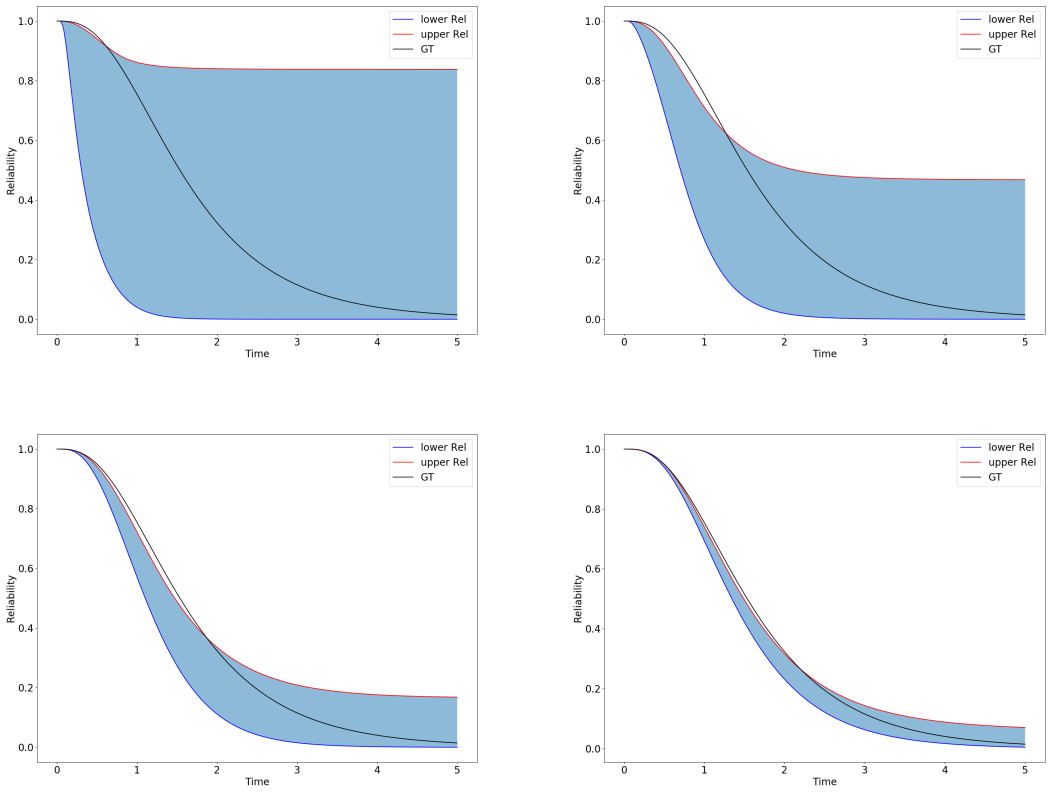


Figure 5.13: Inferred bounds for the survival function of the example system for n=5,15,50, and 150 observations (depicted respectively from top to bottom).

5.5 Simultaneous sampling from a set of distributions

The following section is an exerpt of a conference contribution submitted to UQOP2021, [123].

We propose an extension to Markov Chain Monte Carlo methods for inferences in the imprecise probability framework. The algorithm is based on simultaneous sampling from all the Markov chains targeting the distributions in the credal set. The algorithm constructs a chain of random sets, which can be used for conservative estimation of lower and upper expected values of derived random variables. Tight bounds on the set of estimators arising from the set of admitted stochastic models can be obtained when the credal set is finite for general models. Conservative bounds can be obtained for some classes of models also when the credal set is uncountable.

Computational complexity for uncountable credal sets is not bounded and heuristic fixes need to be implemented.

5.5.1 Introduction

Computations with sets of probability distributions can be helpful for sensitivity analysis in stochastic inference [78] or when imprecise models are used as an uncertainty model [81]. We will refer to this set of probability distributions as a credal set $\mathcal{M}$. These probability distributions define a set of random variables on some common measurable space $(\mathbb{X}, \mathcal{B})$ and treated as a set of mappings $X : \Omega \to \mathcal{X}$, such that $P(dX) =$. Our aim will be to evaluate or estimate expected values of function(s) $f : \mathcal{X} \to \mathbb{R}$ over the set of distributions in $\mathcal{M}$. For coherent imprecise models, that is those which adhere to some rationality constraints [112], the set of expected values over this set is convex. This enables us to limit our focus solely on lower and upper bounds over this set. We will denote these as lower and upper expected values defined in Equation 5.41.

$$\underline{\mathbb{E}} f := \inf_{P \in \mathcal{M}} \int f dP \qquad \text{and} \qquad \overline{\mathbb{E}} f := \sup_{P \in \mathcal{M}} \int f dP. \tag{5.41}$$

Coherency of the stochastic model will also ensure that these bounds will be conjugated such that $\overline{\mathbb{E}} f = -\underline{\mathbb{E}}(-f)$.

Computing expected values of stochastic models is generally an intractable problem. Common practice is to estimate them, most often with Monte Carlo methods [74]. Monte Carlo methods simulate random sampling procedures and use statistical techniques to estimate population parameters - here the theoretical expected values.

For imprecise models, an extension of Importance Sampling Monte Carlo can be used [124, 125]. Importance Sampling constructs a set of samples from an ancillary distribution P_Y and estimates expected values based on theoretical results for changing measures. Under some mild conditions, the expected value can be obtained as

$\mathbb{E}f(X) = \int f(x)P_X(dx) = \int f(y)\frac{dP_X}{dP_Y}(y)P_Y(dy)$, where $\frac{dP_X}{dP_Y}$ is the Radon-Nikodym derivative. The extension for imprecise models is achieved by constructing common independent samples from P_Y and estimating the lower expected value by subsequent minimization of $\mathbb{E}f(X)$ through varying $\frac{dP_X}{dP_Y}$ for $P_X \in \mathcal{M}$.

We will show how another technique, Markov Chain Monte Carlo, can be extended for inferences with imprecise probability models.

5.5.2 Markov chain Monte Carlo

Markov Chain Monte Carlo (MCMC) [126] is a class of methods which are used when we cannot sample from the targeted distribution directly. It constructs and subsequently samples a Markov chain $(X_i)_{i=1}^{\infty}$ whose stationary distribution is the target distribution. These samples will not be independent like in the case of Importance Sampling, but, still, the Law of Large Numbers and a variant of Central Limit Theorem hold, so it can be used to estimate expected values of interest. The chain is characterised by an initial distribution P_0 for X_0 and a transition kernel such that $X_i \sim K(X_{i-1}, .)$. The transition operator is constructed so that the targeted distribution is its invariant, i.e. $P = KP$. The underlying idea is that, individually, each of the samples can be viewed as a sample from the target distribution. Nevertheless, because P_0 is generally different from the target distribution, the chain would only asymptotically converge towards the target distribution.

Still, it can be shown, that under some mild conditions, an expected value of $f : \mathcal{X} \to \mathbb{R}$ can be estimated from a sample of finite trajectory of $(X_i)_k^{k+n}$ by

$$\hat{\mathbb{E}}f = \frac{1}{n}\sum_{i=k}^{k+n} f(X_i) \tag{5.42}$$

for k, n both large enough.

In this section, we will adhere to the common notation and represent samples of random variables by respective lower-case letters. We aim to assess $\mathbb{E}_P f$. MCMC

algorithms generally proceed as follows. Let P be the target distribution and K_P a Markov Chain transition kernel with stationary distribution P. The initial position x_0 is sampled from some distribution P_0 and the sample trajectory, $(x_0, \ldots, x_i)$, is sequentially extended by sampling X_{i+1} from $K_P(x_i, .)$.

A variant of Markov Chain Monte Carlo is the Metropolis-Hastings procedure [126] which constructs K_P with an aid of an ancillary kernel Q. Let prefix $d\cdot$ represent the densities of respective probability measures. New position, x_{k+1}, is sampled as follows:

1. Sample from $X' \sim Q(x_k, .)$;
2. Set $x_{k+1} = \begin{cases} x' \text{ with probability } a_P(x_k, x') \\ x_k \text{ with probability } 1 - a_P(x_k, x') \end{cases}$,

where $a_P(x, x') := \min(\{1, \frac{dP(x')dQ(x',x)}{dP(x)dQ(x,x')}\})$.

These will further be denoted simply by lower-case letters, i.e. $q(x, .) := dQ(x, .)$.

5.5.3 Simultaneous sampling

The proposed extension of MCMC for imprecise probability models lies in sampling a set of chains targeting all the distributions in some credal set $\mathcal{M}$. We will therefore first discuss how to sample from some set of distributions simultaneously. Let $\Gamma_P : \mathcal{A} \to \mathcal{X}$ be a mapping which transforms some ancillary random variable A with known distribution P_A into a random variable with distribution P. Then $\Gamma_P(A) \overset{d}{=} X \sim P$, which means that for functions $f : \mathcal{X} \to \mathbb{R}$, we can use the ancillary random variable to compute the expected values $\mathbb{E}_{P_A} f(\Gamma_P(A)) = \mathbb{E}_P f(X)$.

As an example, let P be an exponential distribution with rate λ. We can choose the ancillary random variable to be $A \sim \text{ Unif}((0, 1))$ and $\Gamma_P : [0, 1] \to \mathbb{R} : \Gamma_P(A) = \frac{-log(1-A)}{\lambda}$. This Γ_P, which is exactly the quantile function of the respective exponential variable, satisfy $\Gamma_P(A) \sim \text{Exp}(\lambda)$. This procedure is commonly reffered to as the

inverse-transform method [74].

However, in order to extend the inverse-transform sampling to sets of distributions, we define $\Gamma_{\mathcal{M}} : \mathcal{A} \to 2^{\mathcal{X}}$ as a mapping which takes the union of all $\Gamma_P; P \in \mathcal{M}$. Hence $\Gamma_{\mathcal{M}}(A)$ is a non-empty random set such that, by construction, $\forall P \in \mathcal{M}, \Gamma_P(A) \in \Gamma_{\mathcal{M}}(A)$ a.s. For integrable functions of interest, $f : \mathcal{X} \to \mathbb{R}$, it will therefore also hold that $\forall P \in \mathcal{M} : f(\Gamma_P(A)) \in f(\Gamma_{\mathcal{M}}(A))$. We will define the expected value of f over $\mathcal{M}$ by the Aumann integral [87, 127]

$$\mathbb{E}_{\mathcal{M}} f := \mathbb{E} f(\Gamma_{\mathcal{M}}(A)) = \{\mathbb{E} f(X) : X \in S(\Gamma_{\mathcal{M}}(A))\}, \tag{5.43}$$

where $S(\Gamma_{\mathcal{M}}(A))$ is the set of all integrable selectors of $\Gamma_{\mathcal{M}}(A)$ [87], i.e. all random variables X, such that $X \in \Gamma_{\mathcal{M}}(A)$ almost surely.

Clearly, by construction, $\mathcal{M} \subset S(\Gamma_{\mathcal{M}}(A))$, so

$$\forall P \in \mathcal{M} : \inf \mathbb{E}_{\mathcal{M}} f \leq \mathbb{E}_P f, \tag{5.44}$$

so we can use $\inf \mathbb{E} f(\Gamma_{\mathcal{M}}(A))$ as a lower bound for $\inf_{P \in \mathcal{M}} \mathbb{E} f(X) = \underline{\mathbb{E}} f$.

Now, in the case of independent sampling, empirical approximation of $\Gamma_{\mathcal{M}}(A)$ and Equation 5.44 could be used to compute estimates on the lower bound for expected values. Operations with sets are realized via Minkowski operators. Convergence is assured by the strong law of large numbers [128]. But we shall construct dependent samples in our MCMC scheme.

5.5.4 Markov chain Monte Carlo for imprecise models

Here we describe how trajectories of several Markov Chains can be constructed simultaneously. The algorithm is based on an extension of the Metropolis-Hastings procedure described in Section 5.5.2. A general propagation step can be rewritten with an aid of ancillary random variables which will later serve a similar purpose as in Section 5.5.3. Given the last position of the chain targeting distribution P is x_k:

1. Sample ancillary $U_Q \sim \text{Unif}((0,1))$.
2. Construct a proposal step $x' = \Gamma_{Q(x_k,.)}(u_Q)$.
3. Sample ancillary $U_A \sim \text{Unif}((0,1))$.
4. Set $x_{k+1} = \begin{cases} x' & \text{if } u_A < a_P(x_k, x') \\ x_k & \text{otherwise} \end{cases}$

where $a_P(x, x') = \min(\{1, \frac{dP(x')dQ(x',x)}{dP(x)dQ(x,x')}\})$ is the Metropolis-Hastings acceptance probability and Γ_P, together with the respective U, represents a mapping, such that $\Gamma_P(U) \sim P$ as in Section 5.5.3.

This whole operation can be summarized such that $X_{k+1} = \Gamma_{K_P(X_k,.)}(U_Q, U_A)$ and applied recursively to describe the whole Markov chain via ancillary variables. Denoting U_0 an ancillary variable for sampling the initial position X_0,

$$X_0 = \Gamma_{P_0}(U_0) \tag{5.45}$$

$$X_k = \Gamma_{K_P(X_{k-1},.)}(U_{Qk}, U_{Ak}). \tag{5.46}$$

We will further use superscript ^{P} to specify the chain target distribution, thus denoting X_k^P the k-th position of the Metropolis-Hastings chain targeting P.

The ancillary variables can now be utilized for simultaneous sampling. Let $\mathcal{M}$ be a credal set, a set of probability distributions on some common space $\mathcal{X}$. Define $\mathbb{X}_k = \{\Gamma_{Pk}(U_{Qk}, U_{Ak}), P \in \mathcal{M}\}$ a set of k-th positions of respective Markov Chains targeting distributions $P \in \mathcal{M}$. Since $\mathbb{X}$ contains each individual sample, $\inf \mathbb{X}_k \leq \Gamma_{Pk}(U_{Qk}, U_{Ak}) = X_k^P$,

$$\begin{aligned} \frac{1}{n}\sum_{i=k}^{k+n} \inf f(\mathbb{X}_k) &\leq \inf_{P\in\mathcal{M}} \frac{1}{n}\sum_{i=k}^{k+n} f(\Gamma_{Pk}(U_{Qk}, U_{Ak})) \\ &= \inf_{P\in\mathcal{M}} \frac{1}{n}\sum_{i=k}^{k+n} f(X_k^P) = \inf_{P\in\mathcal{M}} \hat{\mathbb{E}}_P f. \end{aligned} \tag{5.47}$$

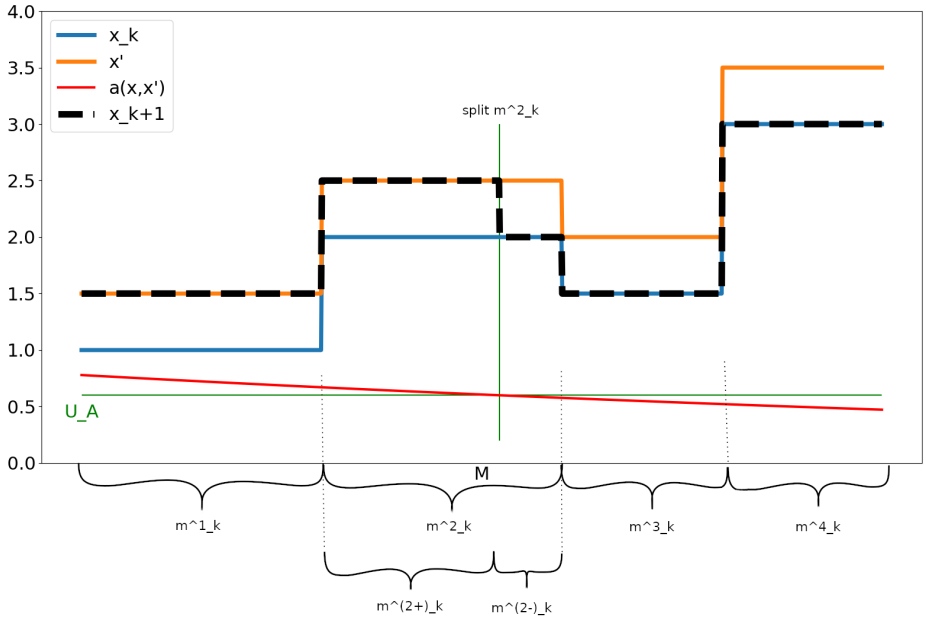


Figure 5.14: An iteration of iMCMC. A new step x' (orange) is proposed based on current positions of all the chains (blue) in $\mathcal{M}$, acceptance criterion $a(x, x')$ (red) is computed and chains are propagated (black dashed) based on common random variable U (green horizontal line). One of the sheaves, M_2, is split during this step.

Similarly as in Section 5.5.3, we can use the sampled set-process as a lower estimate of the lower expected value of interest.

The procedure can be visualised as follows. First, represent $\mathcal{M}$ by $M \subset \mathbb{R}^d$, which is trivial if $\mathcal{M}$ is a set of parametrized distributions. (Example: $M = [1, 2] \subset \mathbb{R}$ for representing a set of exponential distributions, $\mathcal{M}$, with rate parameters $\lambda \in M$.) In order to keep an analogue to the Markov property for the constructed chain of sets, we define random elements $\mathbb{Y} : M \to \mathbb{X}$. These are random mappings representing relation between individual distributions $P \in \mathcal{M}$ (aka M) and random samples, X^P, related to them. In the notation introduced above, $\mathbb{X}_k = \mathbb{Y}_k(M)$ is the collection of samples from all the distributions in $\mathcal{M}$. Figure 5.14 depicts a sample of $\mathbb{Y}_k$ (blue line), proposed moves for each of the chains (orange line), evaluated acceptance criteria (red line below), and new positions of the chains $\mathbb{Y}_{k+1}$ (dashed black line). In the example, some chains accepted their proposals and some rejected.

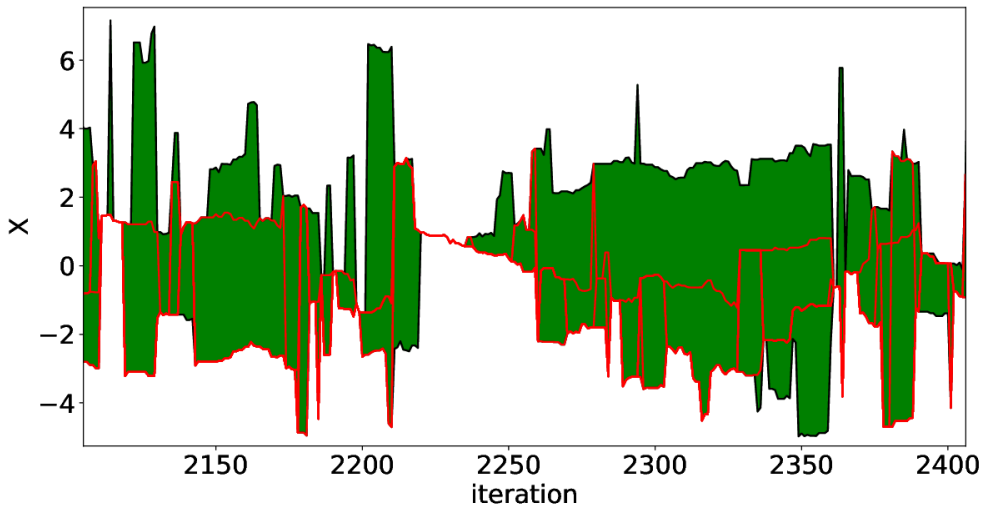


Figure 5.15: An example trajectory of a set of Markov chains. Depicted in red are 5 individual chains. Green area represents convex closure of sets of positions of 1000 chains.

An example of evolution of the set of chains is depicted in Figure 5.15. In this example, we construct chains for an imprecise normal distribution with means $\mu \in [-2, 2]$. Therefore $M = [-2, 2]$. The set-valued estimator, Equation 5.47 - left-hand side, gives lower bound on $\underline{\mathbb{E}}\hat{X}$ of $-2.4 < 2$.

5.5.5 Practical implementation

In this section we will describe how we can practically simultaneously propagate even uncountably many chains. We achieve this by propagating sets of pairs (m, x), where $m \subset \mathcal{M}$ and $x \in \mathcal{X}$. These pairs represent that each of the Markov Chains targeting $P \in m$ is, at the given step, at the same position.

Denote $MC_k = \{(m_k^1, x_k^1), (m_k^2, x_k^2), \ldots\}$ as the set of these pairs at step k. In the introduced notation, $\mathbb{Y}_k(y) = x \iff \exists$ pair $(m, x) \in MC_k : y \in m$.

At each step, we require that $\bigcup_i m_k^i = \mathcal{M}$ in order to propagate chains for all $P \in \mathcal{M}$. Also, omitting the details on boundaries, we require that, for each k, each $P \in \mathcal{M}$ lies in exactly one of m_k^i. For technical reasons, we will assume that the distributions in $\mathcal{M}$ have non-empty intersection of their supports and that all chains start from a

common initial point; thus $MC_0 = \{(\mathcal{M}, x_0)\}$. This requirement can be bypassed. We choose a common proposal kernel Q for all $P \in \mathcal{M}$. The MCMC procedure described in Section 5.5.4 will proceed similarly only jointly for each of the pairs in MC_k. At each step, for each of the pairs, we will construct proposal position x'_i by sampling from $Q(x^i_k, .)$. This will be common to all $P \in m^i_k$ because of the choice of common proposal kernel Q and common ancillary variable U_Q, and will depend solely on the last position of the set of chains, x^i_k. The ancillary decision variable U_A is again common to all chains, so one of three situations may occur. Either

1. $\forall P \in m^i_k : u_A < a_P(x^i_k, x'_i)$,
2. $\forall P \in m^i_k : u_A \geq a_P(x^i_k, x'_i)$,
3. $\exists P_1, P_2 \in m^i_k : u_A < a_{P_1}(x^i_k, x'_i) \wedge u_A \geq a_{P_2}(x^i_k, x'_i)$.

In cases 1,2, the whole pair (m^i_k, x^i_{k+1}) will be included in MC_{k+1}. In the latter case, we will instead include two new pairs into MC_{k+1}, (m^{i+}_k, x'_i) and (m^{i-}_k, x^i_k), based on which $P \in m^i_k$ accept or reject the proposal. We refer to the latter situation as branching.

The method is indicated in Figure 5.14. In the figure, we can see 4 distinct pairs (m^i_k, x^i_k), common proposals x'^i for each of the pairs, and a split occurring for m^2_k. MC_{k+1} therefore equals $\{(m^1_k, x'_1), (m^{2+}_k, x'_2), (m^{2-}_k, x^2_k), (m^3_k, x^3_k), (m^4_k, x^4_k)\}$.

Splitting of credal sets is generally an intractable task since we would need not only to check uncountable many conditions, but also find a way how to represent an arbitrary partition of m^i_k. Nevertheless, in certain cases, this can be done efficiently.

Assume that the credal set $\mathcal{M}$ is a set of distributions from the same exponential family. Their densities can therefore be represented in natural form as

$$p(x|\eta) = h(x)g(\eta)\exp\left[\eta T(x)\right]. \tag{5.48}$$

Inserting Equation 5.48 into the criterion for accepting proposal x', we obtain

$$
\begin{aligned}
u_A < a_\eta(x, x') &= \frac{h(x')g(\eta)\exp\left[\eta T(x')\right] q(x', x)}{h(x)g(\eta)\exp\left[\eta T(x)\right] q(x, x')} \\
&= \frac{h(x')q(x', x)}{h(x)q(x, x')}\exp\left[\eta(T(x') - T(x))\right] \qquad (5.49) \\
\log(u_A) - \log\left(\frac{h(x')q(x', x)}{h(x)q(x, x')}\right) &< \eta(T(x') - T(x)),
\end{aligned}
$$

which is a linear condition if we represent $\mathcal{M}$ as $M \subset \mathbb{R}$ in the space of natural parameters.

A similar simplification occurs also for joint distributions of independent random variables. A special case are hierarchical models if the imprecision is in variables without parents like Bayesian inference with imprecise prior distributions. For inference on model parameters θ in the light of observation x with imprecise hyperparameters η, the posterior is:

$$
p_\eta(\theta|x) = \frac{L(x|\theta)p_0(\theta|\eta)}{\int L(x|\theta)p_0(\theta|\eta)d\theta}. \qquad (5.50)
$$

Plugging Equation 5.50 into the acceptance ratio for proposal τ:

$$
\begin{aligned}
a(\theta, \tau) &= \frac{p_\eta(\tau|x)q(\tau, \theta)}{p_\eta(\theta|x)q(\theta, \tau)} \\
&= \frac{q(\tau, \theta)}{q(\theta, \tau)}\frac{L(x|\tau)}{L(x|\theta)}\frac{\frac{p_0(\tau|\eta)}{\int L(x|\tau)p_0(\tau|\eta)d\tau}}{\frac{p_0(\theta|\eta)}{\int L(x|\theta)p_0(\theta|\eta)d\theta}} \qquad (5.51) \\
&= \frac{q(\tau, \theta)}{q(\theta, \tau)}\frac{L(x|\tau)}{L(x|\theta)}\frac{p_0(\tau|\eta)}{p_0(\theta|\eta)} \\
&\propto \exp\left[\eta\left(T(\tau) - T(\theta)\right)\right].
\end{aligned}
$$

Another advantage of the finite implementation is that $\mathbb{X}_k$ will remain finite. This enables us to evaluate $\inf f(\mathbb{X}_k)$ through comparison of all elements instead of solving an intractable optimization problem.

5.5.6 Representation through a set of inequalities

Although we may represent the splitting conditions by hyper-planes for a wide class of models, tracking the branches could still be intractable. In order to facilitate this

task, we will introduce how to numerically represent credal sets as sets of inequalities. This poses some restrictions on $\mathcal{M}$ since it also needs to be represented in this way.

Assume that the imprecise model allows linear splitting as described in Section 5.5.5. Denote $\mathbb{C}(\mathcal{M})$ a set of linear inequality constraints $c_j^T\eta > b_j$, such that $\mathcal{M} = \{\eta : C\eta \succ b\}$, where $\succ$ represents point-wise vector dominance. We rewrite Equation 5.49 as $b(u_a, x, x') < s(x, x')^T\eta$ for the sake of compact representation. At each branching occurrence, the newly created sets m_k^+, m_k^- can be represented as new sets of inequalities, such that

$$\begin{aligned}\mathbb{C}(m_k^+) &= \mathbb{C}(m_k) \cup \{s(x_k, x_k')^T\eta > b(u_A, x_k, x_k')\},\\ \mathbb{C}(m_k^-) &= \mathbb{C}(m_k) \cup \{-s(x_k, x_k')^T\eta > -b(u_A, x_k, x_k')\},\end{aligned} \tag{5.52}$$

while omitting the boundary cases again. Branching occurs if $\exists\eta_1, \eta_2 : s(x, x')^T\eta_1 > b(u_A, x, x') \wedge s(x, x')^T\eta_2 < b(u_A, x, x')$. This can be checked by solving two linear programs. Define

$$\begin{aligned}\underline{b} &:= \min_{\eta\in m} s(x, x')^T\eta,\\ \overline{b} &:= \max_{\eta\in m} s(x, x')^T\eta = -\min_{\eta\in m} -s(x, x')^T\eta.\end{aligned} \tag{5.53}$$

Condition $\eta \in m$ represents the set of linear inequality constraints $\mathbb{C}(m)$. Clearly, $\underline{b}$ and $\overline{b}$ represent achievable bounds on $b = s(x, x')^T\eta$ so branching occurs iff $b \in (\underline{b}, \overline{b})$. If $b < \underline{b}$, then all $\eta \in m$ accept the proposal x'. If $b > \overline{b}$, then all $\eta \in m$ reject it.

This procedure would add a new constraint at each branching occasion. This might eventually make some of them redundant. Whether a constraint is redundant and can be removed from $\mathbb{C}(m)$ can be checked every couple of iterations in order to decrease the size of involved linear programs. This can be done for each constraint $c_j^T\eta > b_j$ in $\mathbb{C}(m)$ by either:

- Solving $\min_{\eta\in m} c_j^T\eta$. If the solution is greater than b_j, then the constraint is redundant.

- Solving a dual problem to $\min_{\eta \in m} c_j^T \eta$. If constraint j is inactive, it is redundant.

5.5.7 Credal set merging

Branching emerging from the above mentioned procedures would produce an ever increasing number of partitions of $\mathcal{M}$ and become intractable for construction of long chains. In this section, we will show how to partially counter this tendency through special choice of the proposal kernel Q.

$$q_m(x'|x) = (1-\alpha)q_{rw}(x'|x) + \alpha q(x'). \tag{5.54}$$

If the proposal kernel is independent on the last position of the chain it enables proposing the same position even for deviated chains with positive probability that all will accept this proposal. On this occasion, previously branched chains would coalesce. This tendency can also be achieved by introducing the independent proposal distribution as a component of a proposal mixture (Equation 5.54) together with the standard random walk kernel.

Occasional merging of the split branches allows us to decrease the computational demands of the procedure. If it is allowed, we need to adjust our implementation in order to keep track of the individual chains in order to estimate $\underline{\mathbb{E}} f$ by the minimum of MCMC estimates over $\mathcal{M}$ as on the right-hand side of Equation 5.47. In the uncountable case, an expensive post-hoc analysis is required in order to untangle all possible trajectories of the chains and obtain theoretically tighter bounds than from the set representation on the left-hand side of Equation 5.47. Random set induced bounds on $\underline{\mathbb{E}} f$ are therefore used in practice.

Nevertheless, the branching tendency for uncountable credal sets still leads to exponential growth of credal set partitions. Although the merging is frequent, it cannot be guaranteed that the number of partitions will remain within limits imposed by our hardware. A pragmatic solution to this problem is to limit ourselves

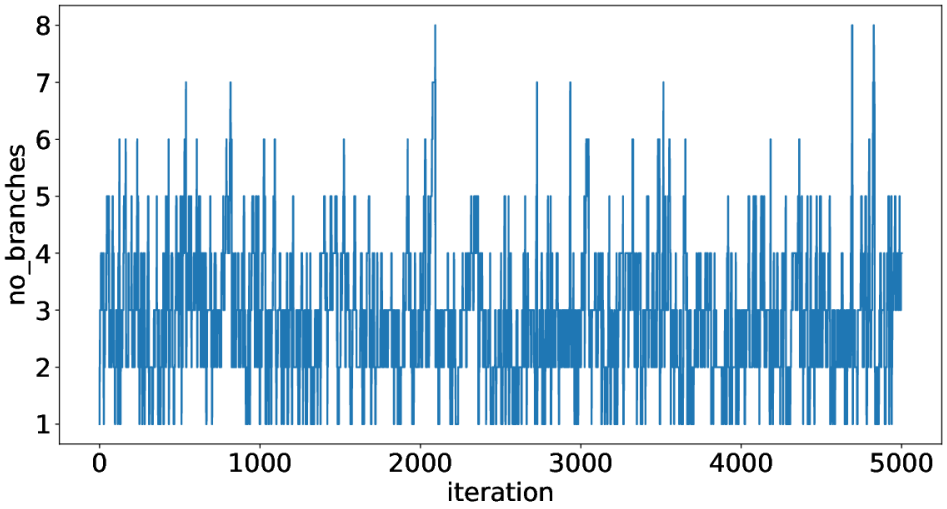


Figure 5.16: An example of evolution of the number of credal set partitions for $|\mathcal{M}| = 1000$.

to calculations with finite credal sets. For an infinite one, it is possible to create its finite subset and use the resulting estimators to estimate lower expectation over the original set. A combination of a dense finite subset of the original credal set with the above mentioned methods can counter both the computational costs of tending to large number of chains individually and the pathological excessive branching. Evolution of the number of branches for $|\mathcal{M}| = 1000$ for a simple model of imprecise normal distribution is depicted in Figure 5.16. During the evolution, the active amount of branches remains a fraction of the original credal set size.

5.5.8 Discussion

We have introduced several steps towards Markov Chain Monte Carlo methods for imprecise probabilities based on an extension of Metropolis-Hastings algorithm. The procedure can be used to obtain conservative bounds on extremes of MCMC estimators arising from Markov Chains targeting individual distributions in the credal set. For a class of problems, it was shown in Section 5.5.5 how to design a numerically tractable representations of the set of individual chains even for uncountable credal sets.

Nevertheless, we judge that the methodology is still not ready for practical use.

For uncountable credal sets, the exact procedure leads to high demand on computational resources through excessive branching of the chains. Although this tendency can be limited to some extent, as described in Section 5.5.7, it cannot be guaranteed that these requirements will be bounded. A heuristic approach was proposed in Section 5.5.7 based on finite approximation of infinite credal sets. Preliminary results show that this could limit the computational demands at the cost of introducing additional error to the estimates.

No theoretical analysis of convergence was provided. As shown in Section 5.5.4, the procedure practically constructs a Markov Chain $(\mathbb{Y}_k)$ of mappings $\mathcal{M} \to \mathcal{X}$ with unspecified structure. Properties of associated random sets $\mathbb{X}_k = \mathbb{Y}_k(\mathcal{M})$ might be more accessible.

5.6 Simultaneous sampling for uniform distribution

In this section we aim to explore further the imprecise Markov Chain Monte Carlo method proposed in Section 5.5. The method was presented on a general inference problem with a set of target distribution. This section will investigate its properties on a simple model.

5.6.1 Nested set of uniform distribution

The target set of models is a nested set of uniform distributions. This provides both simple enough problem for some theoretical investigation. Simultaneously, it provides possible applications for models which can be modelled as inverse sampling procedures or for creating predictive random sets for inferential models from Section 3.4.5.

For simplicity, we consider a one-sided nested set of uniform distributions $\mathcal{M} = \{Unif(0, m) : m \in [0, 1]\}$.

If we construct a random set from a seed variable $A \sim Unif(0, 1)$ by

$$R = \{A \max P : P \in \mathcal{M}\}, \tag{5.55}$$

where $\max P$ represents the maximum value the uniform distribution $P = Unif(0, m)$ can obtain (thus $\max P = m$), we obtain exactly predictive random set from Section 3.4.5.

5.6.2 Transition kernel for set of uniform distributions

Considering each distribution $P \in \mathcal{M}$ individually, the acceptance ratio $a(x, x')$ is simple to evaluate since the uniform distribution has constant density on its support.

$$a(x, x') = \begin{cases} 1; & x' \leq \max P \\ 0; & x' > \max P \end{cases} \tag{5.56}$$

Corollary 5.6.1 (Persistent zero.)**.** *Since* $\mathcal{M} \ni Unif(0, 0) \equiv 0$*,*

then 0 *will be present in each collection of samples* $\mathbb{X}_k = \mathbb{Y}_k(M)$*.*

Corollary 5.6.2 (Explicit starting sample.)**.** *Since* $\mathcal{M} \ni Unif(0, 0) \equiv 0$*,*

then $\{0\} = \bigcap_{P \in \mathcal{M}} supp(P) = \mathbb{X}_0$*.*

The transition from $\mathbb{X}_k$ to $\mathbb{X}_{k+1}$ is depicted on Figure 5.17. Algorithmically, given a proposal $A \sim Unif(0, 1)$,

$$\mathbb{X}_{k+1} = \{x \in \mathbb{X}_k : x \leq A\} \cup \{A\}. \tag{5.57}$$

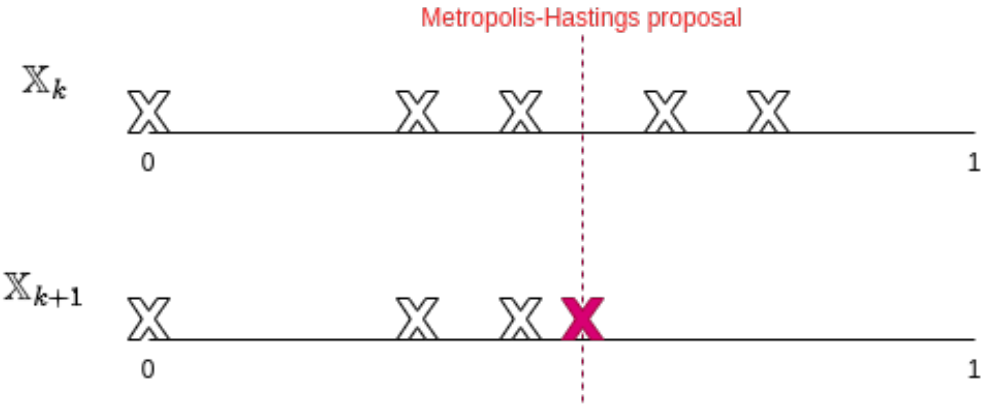


Figure 5.17: Transition of the simultaneous sampling of nested uniform distributions. The "X" mark individual elements of $\mathbb{X}$

5.6.3 Asymptotic properties

The constructed process is a random set process. To explore its properties, we investigate the properties of individual random sets (from Section 3.3.3).

First trivial observation is that the contour function $\gamma(x) = \delta(x)$. This is because $0 \in \mathbb{X}_k, \forall k$ and the probability of convering x is zero since we sample individual points.

Next, consider general (for simplicity) closed set $E \subset [0, 1]$ and the hitting probability $T_k(E) = P(\mathbb{X}_k \cap E \neq \emptyset)$. Considering Metropolis-Hastings proposal $A \sim Unif(0, 1)$, we can express $T_{k+1}(E)$ using law of total probability considering cases:

1. $\mathbb{X}_k \cap E \neq \emptyset$ and $A \geq \min(E)$,
2. $\mathbb{X}_k \cap E = \emptyset$ and $A \in E$.

Therefore

$$\begin{aligned} T_{k+1}(E) &= T_k(E) \cdot P(A > \min(E)) + (1 - T_k(E)) \cdot P(A \in E) \\ &= T_k(E) \cdot (1 - \min(E)) + (1 - T_k(E)) \cdot (\max(E) - \min(E)). \end{aligned} \tag{5.58}$$

If there exist a stable distribution such that $T_k(E) = T_{k+1}(E)$, then

$$\begin{aligned} T(E) &= T(E) - T(E)\min(E) + \max(E) - \min(E) - T(E)\max(E) + T(E)\min(E) \\ &= T(E)(1 - \max(E)) + \max(E) - \min(E) \end{aligned} \tag{5.59}$$

Therefore

$$T(E) = \frac{\max(E) - \min(E)}{\max(E)}, \tag{5.60}$$

with a special edge case of $T(\{0\}) = 1$ due to Corollary 5.6.1.

Theorem 5.6.1 (Fixed point exists.)**.** *For fixed $E \subset [0, 1]$, there exists $T(E)$ such that $T(E) = T_k(E) = T_{k+1}(E)$.*

Proof. Since the mapping $[0, 1] \ni T_k(E) \to T_{k+1}(E) \in [0, 1]$ in Equation 5.58 is continuous surjection on a compact closed set, there exists at least one fixed point according to Brouwer fixed point theorem. □

5.6.4 Numerical comparison

In Table 5.3 we compare the hitting probabilities of investigated sampling methodologies. On this particular example, it can be seen that the proposed simultaneous sampling method has at least the same hitting probability than simple uniform sampling. Nevertheless, if we compare it with a naive construction of a valid predictive random set R from Equation 5.55, the hitting probabilities are lower. For $|E| \to 0$, the hitting probability tends to 0.

Table 5.3: Comparison of probability of hitting E with random uniform sample A, stable distribution of the proposed procedure $T(E)$ and valid predictive random set R.

E	$P(A \in E)$	$T(E)$	$P(R \cap E \neq \emptyset)$
[0,1]	1	1	1
[1/2,1]	1/2	1/2	1/2
[1/4,1]	3/4	3/4	3/4
[1/4, 1/2]	1/4	1/2	3/4
[1/2,3/4]	1/4	1/3	1/2

5.6.5 Conclusions

In this section, we have demonstrated a simple investigation into properties of simultaneous sampling methodology introduced in Section 5.5 on nested set of uniform distributions. The investigated model is simple enough to calculate the evolution of hitting probabilities and even find a stable hitting distribution for specific class of events (closed subsets of $[0, 1]$).

The constructed random set process still consist of individual samples of measure 0. This itself disqualifies the method from possibility of being used as a sampling technique for constructing predictive random sets for inferential models framework (Section 3.4.5) since the coverage probability $\gamma(x) = 0, \forall x \in [0, 1]$, which is does not stochastically dominate uniform distribution.

Nevertheless, the methodology still poses beneficial qualities, namely construction of singular point samples at each step, which makes it useful for practical numerical applications. As shown in Table 5.3, the method provides sligtly higher hitting probabilities for particular events, which may lead to better frequentist coverage properties than simple single-point sampling with minimal additional computational effort.

Chapter 6

Conclusions

The thesis presented multiple results from investigation of application of imprecise probabilities to system reliability engineering. Emphasis was put on using survival signature models (Section 2.3) and their versatility in system analysis (Section 2.3.5). Computation of system signatures is still a challenging task, but recent results already allow to compute them for large real-world systems (Section 2.3.6).

Further investigations were focused on application of imprecise probability methods for statistical inference of system reliablity. Reliable and robust statistical inference of system component failure laws is necessary to provide confidence in the system reliability assessment itself. Apart from revision of generally used statistical methods like maximum likelihood estimation and Bayesian inference, fiducial methods based on direct inversion of the structural uncertainty model and methods based on imprecise probabilities were investigated (Section 3.4). These promise to provide additional benefits in terms of robustness and inferential validity of the constructed models (Section 3.4.5), but at the cost of increased computational complexity, which was attempted to bypass by the proposed sampling algorithm (Section 5.5 and Section 5.6).

In conclusion, survival signatures provide effective representation of complex systems.

Using survival signatures transfers combinatorial complexity of system reliability assessment to a preprocessing phase and allow efficient computations afterwards. The signatures preserve the monotone structure of coherent systems, which facilitates analyses of systems subjected to severe uncertainty. Imprecise probabilities show as a versatile tool for uncertainty modeling including problems of statistical inference. Imprecision allows us to avoid including additional assumptions into statistical models and lead to well-calibrated statistical methods like fiducial inference, nonparametric predictive inference or inferential models. This family of fiducial methods allow us to construct Bayes-like posterior distributional estimates for the parameters of interest without the need of artificially specifying prior distributions needed for Bayesian inference. At the moment, the methods are often computationally and analytically more demanding which limits their practical applications to specific cases.

Survival signatures have been demonstrated as a promising methodology for system reliability analysis with uncertainty modelled using imprecise probability theory.

Bibliography

[1] Krpelik, D., Huang, X., Aslett, L. J. M. and Coolen, F. P. A. (2018). Reliability assessment of phased mission systems subjected to epistemic uncertainty and optimisation of the phase ordering.Glasgow.

[2] Krpelik, D., Coolen, F. P. A. and Aslett, L. J. M. (2019). A decomposition approach for computation of survival signatures of heterogeneous systems with subsystems with shared components.

[3] Krpelik, D., Coolen, F. P. A. and Aslett, L. J. M. (2018). Imprecise probability inference on masked multicomponent system. 133–40.

[4] Krpelik, D., Coolen, F. P. A. and Aslett, L. J. M. (2019). On robust markov analysis for reliability assessment of complex systems using imprecise markov chains.

[5] Krpelik, D., Aslett, L. J. M. and Coolen, F. P. A. (2021). Simultaneous sampling for robust markov chain monte carlo inference. In *Advances in uncertainty quantification and optimization under uncertainty with aerospace applications* (M. Vasile and D. Quagliarella, ed) pp 173–83. Springer International Publishing, Cham.

[6] Vasile, M. (2021). *Optimization under uncertainty with applications to aerospace engineering*. M. Vasile, ed Springer International Publishing AG.

[7] Jardine, A. K. S. (1973). *Maintenance, replacement, and reliability*. Wiley.

[8] Osaki, S. (2002). *Stochastic models in reliability and maintenance*. Springer.

[9] Lewis, E. E. (1995). *Introduction to reliability engineering*. Wiley.

[10] Barlow, R. E. (1996). *Mathematical theory of reliability*. SIAM, Philadelphia, MS.

[11] Singpurwalla, N. D. (2006). *Reliability and risk: A bayesian perspective*. Wiley-Blackwell, Hoboken, NJ.

[12] Todinov, M. (2015). *Reliability and risk models: Setting reliability requirements*. John Wiley & Sons, Nashville, TN.

[13] Casella, G. and Berger, R. L. (2010). *Statistical inference*. Thomson Press (India) Ltd.

[14] Ruijters, E. and Stoelinga, M. (2015). Fault tree analysis: A survey of the state-of-the-art in modeling, analysis and tools. *Computer Science Review* **15-16** 29–62.

[15] Lee, W. S., Grosh, D. L., Tillman, F. A. and Lie, C. H. (1985). Fault tree analysis, methods, and applications: A review. *IEEE Transactions on Reliability* **R-34** 194–203.

[16] Haasl, D. F., Roberts, N. H., Vesely, W. E. and Goldberg, F. F. (1981). *Fault tree handbook*.

[17] O'Connor, P. D. T. and Kleyner, A. (2012). *Practical reliability engineering*. Wiley.

[18] Bryant. (1986). Graph-based algorithms for boolean function manipulation. *IEEE Transactions on Computers* **C-35** 677–91.

[19] Brace, K. S., Rudell, R. L. and Bryant, R. E. (1990). Efficient implementation of a BDD package. In *Conference proceedings on 27th ACM/IEEE design automation conference - DAC '90* DAC '90. ACM Press.

[20] Rauzy, A. (1993). New algorithms for fault trees analysis. *Reliability Engineering & System Safety* **40** 203–11.

[21] Andrews, J. D. and Dunnett, S. J. (2000). Event-tree analysis using binary decision diagrams. *IEEE Transactions on Reliability* **49** 230–8.

[22] Hardy, G., Lucet, C. and Limnios, N. (2007). K-terminal network reliability measures with binary decision diagrams. *IEEE Transactions on Reliability* **56** 506–15.

[23] Bjorkman, K. (2013). Solving dynamic flowgraph methodology models using binary decision diagrams. *Reliability Engineering & System Safety* **111** 206–16.

[24] Xing, L. (2008). An efficient binary-decision-diagram-based approach for network reliability and sensitivity analysis. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* **38** 105–15.

[25] Heckerman, D. (2008). A tutorial on learning with bayesian networks. In *Innovations in bayesian networks: Theory and applications* (D. E. Holmes and L. C. Jain, ed) pp 33–82. Springer Berlin Heidelberg, Berlin, Heidelberg.

[26] Kennet, R. S., Ruggeri, F. and Faltin, F. W. (2018). *Analytic methods in systems and software testing*. John Wiley; Sons (Verlag).

[27] Bessiere, P., Mazer, E., Ahuactzin, J. and Mekhnacha, K. (2013). *Bayesian programming*. CRC Press.

[28] Esary, J. D. and Ziehms, H. (1975). Reliability analysis of phased missions. *Naval Postgraduate School (U.S.)*.

[29] Band, R. A. L. and Andrews, J. D. (2004). Phased mission modelling using fault tree analysis. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering* **218** 83–91.

[30] Andrews, J. D., Prescott, D. R. and Remenyte-Prescott, R. (2008). A systems reliability approach to decision making in autonomous multi-platform systems operating a phased mission. In *2008 annual reliability and maintainability symposium* pp 8–14. IEEE.

[31] Andrews, J. D., Poole, J. and Chen, W.-H. (2013). Fast mission reliability prediction for unmanned aerial vehicles. *Reliability Engineering & System Safety* **120** 3–9.

[32] Coolen, F. P. A. and Coolen-Maturi, T. (2021). The survival signature for quantifying system reliability: An introductory overview from practical perspective. In *Reliability engineering and computational intelligence* pp 23–37. Springer International Publishing.

[33] Samaniego, F. J. (2007). *System signatures and their applications in engineering reliability*. Springer US.

[34] Patelli, E., Feng, G., Coolen, F. P. A. and Coolen-Maturi, T. (2017). Simulation methods for system reliability using the survival signature. *Reliability Engineering & System Safety* **167** 327–37.

[35] Samaniego, F. J. (1985). On closure of the IFR class under formation of coherent systems. *IEEE Transactions on Reliability* **R-34** 69–72.

[36] Boland, P. J. (2001). Signatures of indirect majority systems. *Journal of Applied Probability* **38** 597–603.

[37] Coolen, F. P. A. and Coolen-Maturi, T. (2012). Generalizing the signature to systems with multiple types of components. In *Complex systems and dependability* (W. Zamojski, J. Mazurkiewicz, J. Sugier, T. Walkowiak and J. Kacprzyk, ed) pp 115–30. Berlin, Heidelberg.

[38] Coolen, F. P. A. and Coolen-Maturi, T. (2015). Modelling uncertain aspects of system dependability with survival signatures. 19–34.

[39] Rusnak, P., Zaitseva, E., Coolen, F. P. A., Kvassay, M. and Levashenko, V. (2023). Logic differential calculus for reliability analysis based on survival signature. *IEEE Trans. Dependable Secur. Comput.* **20** 1529–40.

[40] Rusnak, P. (2021). Algorithms for calculation of logical derivatives for survival signature and their analysis. In *2021 international conference on information and digital technologies (IDT)* pp 116–22.

[41] Behrensdorf, J., Brandt, S., Broggi, M. and Beer, M. (2018). Efficient approximation of the survival signature for large networks. In *Proceedings of the 6th international symposium on reliability engineering and risk management* 6ISRERM 2018. Research Publishing Services.

[42] Coolen, F. P., Coolen-Maturi, T. and Al-nefaiee, A. H. (2014). Nonparametric predictive inference for system reliability using the survival signature. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* **228** 437–48.

[43] Aslett, L. J. M., Coolen, F. P. A. and Wilson, S. P. (2014). Bayesian inference for reliability of systems and networks using the survival signature. *Risk Analysis* **35** 1640–51.

[44] Walter, G., Aslett, L. J. M. and Coolen, F. P. A. (2017). Bayesian nonparametric system reliability using sets of priors. *International Journal of Approximate Reasoning* **80** 67–88.

[45] Feng, G., Reed, S., Patelli, E., Beer, M. and Coolen, F. P. A. (2017). Efficient reliability and uncertainty assessment on lifeline networks using the survival signature. In *UNCECOMP 2017 - proceedings of the 2nd international conference on uncertainty quantification in computational sciences and engineering* (G. Stefanou, M. Papadrakakis and V. Papadopoulos, ed) pp 90–9.

[46] Coolen, F. P. A. and Coolen-Maturi, T. (2015). Predictive inference for system reliability after common-cause component failures. *Reliability Engineering & System Safety* **135** 27–33.

[47] Mi, J., Beer, M., Li, Y.-F., Broggi, M. and Cheng, Y. (2020). Reliability and importance analysis of uncertain system with common cause failures based on survival signature. *Reliability Engineering & System Safety* **201** 106988.

[48] George-Williams, H., Feng, G., Coolen, F. P., Beer, M. and Patelli, E. (2019). Extending the survival signature paradigm to complex systems with non-repairable dependent failures. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* **233** 505–19.

[49] Rusnak, P., Zaitseva, E., Levashenko, V., Bolvashenkov, I. and Kammermann, J. (2024). Importance analysis of a system based on survival signature by structural importance measures. *Reliability Engineering & System Safety* **243** 109814.

[50] Feng, G., Patelli, E., Beer, M. and Coolen, F. P. A. (2016). Imprecise system reliability and component importance based on survival signature. *Reliability Engineering & System Safety* **150** 116–25.

[51] Eryilmaz, S., Coolen, F. P. A. and Coolen-Maturi, T. (2018). Marginal and joint reliability importance based on survival signature. *Reliability Engineering & System Safety* **172** 118–28.

[52] Huang, X., Aslett, L. J. M. and Coolen, F. P. A. (2019). Reliability analysis of general phased mission systems with a new survival signature. *Reliability Engineering & System Safety* **189** 416–22.

[53] Eryilmaz, S., Coolen, F. P. A. and Coolen-Maturi, T. (2018). Mean residual life of coherent systems consisting of multiple types of dependent components. *Naval Research Logistics (NRL)* **65** 86–97.

[54] Memari, M., Zarezadeh, S. and Asadi, M. (2021). Optimal preventive maintenance for reparable networks. *Applied Stochastic Models in Business and Industry* **37** 1017–41.

[55] Pakdaman, Z., Ahmadi, J. and Doostparast, M. (2017). Signature-based approach for stress-strength systems. *Statistical Papers* **60** 1631–47.

[56] Liu, Y., Shi, Y., Bai, X. and Liu, B. (2018). Dynamic stress–strength reliability estimation of system with survival signature. *Statistical Theory and Related Fields* **2** 181–95.

[57] Eryilmaz, S. and Tuncel, A. (2016). Generalizing the survival signature to unrepairable homogeneous multi-state systems. *Naval Research Logistics (NRL)* **63** 593–9.

[58] Liu, Y., Shi, Y., Bai, X. and Liu, B. (2018). Stress–strength reliability analysis of system with multiple types of components using survival signature. *Journal of Computational and Applied Mathematics* **342** 375–98.

[59] Yi, H., Cui, L. and Balakrishnan, N. (2021). Computation of survival signatures for multi-state consecutive-k systems. *Reliability Engineering & System Safety* **208** 107429.

[60] Qin, J. and Coolen, F. P. A. (2022). Survival signature for reliability evaluation of a multi-state system with multi-state components. *Reliability Engineering & System Safety* **218** 108129.

[61] Aslett, L. J. M. (2015). ReliabilityTheory: Tools for structural reliability analysis.Available at https://CRAN.R-project.org/package=ReliabilityTheory.

[62] Krpelík, D. (2018). Signature calculator.Available at https://github.com/dandaKrpelik/signatureCalculator.

[63] Da, G., Zheng, B. and Hu, T. (2012). On computing signatures of coherent systems. *Journal of Multivariate Analysis* **103** 142–50.

[64] Da, G., Xu, M. and Chan, P. S. (2018). An efficient algorithm for computing the signatures of systems with exchangeable components and applications. *IISE Transactions* **50** 584–95.

[65] Reed, S. (2017). An efficient algorithm for exact computation of system and survival signatures using binary decision diagrams. *Reliability Engineering & System Safety* **165** 257–67.

[66] Reed, S., Löfstrand, M. and Andrews, J. (2019). An efficient algorithm for computing exact system and survival signatures of k-terminal network reliability. *Reliability Engineering & System Safety* **185** 429–39.

[67] Behrensdorf, J., Regenhardt, T.-E., Broggi, M. and Beer, M. (2021). Numerically efficient computation of the survival signature for the reliability analysis of large networks. *Reliability Engineering & System Safety* **216** 107935.

[68] Maio, F. P. D., Pettorossi, C. and Zio, E. (2023). Entropy-driven monte carlo simulation method for approximating the survival signature of complex infrastructures. *Reliability Engineering & System Safety* **231** 108982.

[69] Maio, F. P. D., Pettorossi, C. and Zio, E. (2023). Ensemble of artificial neural networks for approximating the survival signature of critical infrastructures. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*.

[70] Balch, M. S., Martin, R. and Ferson, S. (2019). Satellite conjunction analysis and the false confidence theorem. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **475** 20180565.

[71] Kolmogorov, A. N. (1956). *Foundations of the theory of probability*. Chelsea Pub. Co.

[72] Whittle, P. (1992). *Probability via expectation*. Springer, New York, NY.

[73] Finetti, B. de. (2017). *Theory of probability: A critical introductory treatment*. A. Machi and A. Smith, ed John Wiley & Sons, Nashville, TN.

[74] Kroese, D. P., Taimre, T. and Botev, Z. I. (2011). *Handbook of monte carlo methods*. Wiley.

[75] Troffaes, M. C. M. and Cooman, G. de. (2014). *Lower previsions*. Wiley.

[76] Savage, L. J. (2012). *The foundations of statistics*. Dover Publications.

[77] Boole, G. (1854). *An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities*.

[78] Berger, J. O., Moreno, E., Pericchi, L. R., Bayarri, M. J., Bernardo, J. M., Cano, J. A., De la Horra, J., Martín, J., Ríos-Insúa, D., Betrò, B., Dasgupta, A., Gustafson, P., Wasserman, L., Kadane, J. B., Srinivasan, C., Lavine, M., O'Hagan, A., Polasek, W., Robert, C. P., Goutis, C., Ruggeri, F., Salinetti, G. and Sivaganesan, S. (1994). An overview of robust bayesian analysis. *Test* **3** 5–124.

[79] Choquet, G. (1954). Theory of capacities. *Annales de l'institut Fourier* **5** 131–295.

[80] Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics* **38** 325–39.

[81] Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. Taylor & Francis.

[82] Weichselberger, K. (2000). The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning* **24** 149–70.

[83] Walley, P. (2000). Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning* **24** 125–48.

[84] Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D. S. and Sentz, K. (2003). *Constructing probability boxes and dempster-shafer structures*. Sandia National Laboratories.

[85] Coolen, F. P. A. (1994). *Statistical modeling of expert opinions using imprecise probabilities*. Eindhoven University of Technology.

[86] Dempster, A. P. (2008). The dempster–shafer calculus for statisticians. *International Journal of Approximate Reasoning* **48** 365–77.

[87] Nguyen, H. T. (2006). *An introduction to random sets*. Chapman; Hall/CRC.

[88] Matheron, G. (1975). *Random sets and integral geometry*. John Wiley & Sons Inc.

[89] Molchanov, I. (2005). *Theory of random sets*. Springer-Verlag London.

[90] Martin, R. and Liu, C. (2015). *Inferential models: Reasoning with uncertainty*. Chapman; Hall/CRC.

[91] Oberguggenberger, M. and Fellin, W. (2008). Reliability bounds through random sets: Non-parametric methods and geotechnical applications. *Computers & Structures* **86** 1093–101.

[92] Fetz, T. and Oberguggenberger, M. (2004). Propagation of uncertainty through multivariate functions in the framework of sets of probability measures. *Reliability Engineering & System Safety* **85** 73–87.

[93] Saw, J. G., Yang, M. C. K. and Mo, T. C. (1984). Chebyshev inequality with estimated mean and variance. *The American Statistician* **38** 130–2.

[94] Alvarez, D. A. (2006). On the calculation of the bounds of probability of events using infinite random sets. *International Journal of Approximate Reasoning* **43** 241–67.

[95] Balch, M. S. (2010). *Methods for rigorous uncertainty quantification with application to a mars atmosphere model*. PhD thesis, Virginia Polytechnic Institute; State University, Blacksburg, Virginia.

[96] Walley, P. and Fine, T. L. (1982). Towards a frequentist theory of upper and lower probability. *The Annals of Statistics* **10** 741–61.

[97] Fisher, R. A. (1930). Inverse probability. *Mathematical Proceedings of the Cambridge Philosophical Society* **26** 528–35.

[98] Fraser, D. A. S. (1968). *The structure of inference*. John Wiley & Sons.

[99] Xie, M. and Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review. *International Statistical Review* **81** 3–39.

[100] Balch, M. S. (2012). Mathematical foundations for a theory of confidence structures. *International Journal of Approximate Reasoning* **53** 1003–19.

[101] Ferson, S., Balch, M., Sentz, K. and Siegrist, J. (2013). Computing with confidence. *8th International Symposium on Imprecise Probability: Theories and Applications*.

[102] Dempster, A. P. (1966). New methods for reasoning towards posterior distributions based on sample data. *The Annals of Mathematical Statistics* **37** 355–74.

[103] Hannig, J., Iyer, H. K., Lai, R. C. S. and Lee, T. C. M. (2016). Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association* **111** 1346–61.

[104] Coolen, F. P. A. (1998). Low structure imprecise predictive inference for bayes problem. *Statistics & Probability Letters* **36** 349–57.

[105] Hill, B. M. (1968). Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association* **63** 677.

[106] Augustin, T. and Coolen, F. P. A. (2004). Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference* **124** 251–72.

[107] Coolen, F. P. A., Coolen-Schrijner, P. and Yan, K. J. (2002). Nonparametric predictive inference in reliability. *Reliability Engineering & System Safety* **78** 185–93.

[108] Coolen, F. P. A. and Yan, K. J. (2004). Nonparametric predictive inference with right-censored data. *Journal of Statistical Planning and Inference* **126** 25–54.

[109] Janurová, K. and Briš, R. (2014). A nonparametric approach to medical survival data: Uncertainty in the context of risk in mortality analysis. *Reliability Engineering & System Safety* **125** 145–52.

[110] Martin, R. (2019). False confidence, non-additive beliefs, and valid statistical inference. *International Journal of Approximate Reasoning* **113** 39–73.

[111] Da, G., Chan, P. S. and Xu, M. (2018). On the signature of complex system: A decomposed approach. *European Journal of Operational Research* **265** 1115–23.

[112] Augustin, T., Coolen, F. P., De Cooman, G. and Troffaes, M. C. (2014). *Introduction to imprecise probabilities*. Wiley.

[113] Aslett, L. J. M. (2012). *MCMC for inference on phase-type and masked system lifetime models*. PhD thesis, Trinity College Dublin.

[114] Beer, M., Ferson, S. and Kreinovich, V. (2013). Imprecise probabilities in engineering analyses. *Mechanical Systems and Signal Processing* **37** 4–29.

[115] Ferson, S., Kreinovich, V., Hajagos, J., Oberkampf, W. and Ginzburg, L. (2007). *Experimental uncertainty estimation and statistics for data having interval uncertainty.* Office of Scientific; Technical Information (OSTI).

[116] Norris, J. R. (1997). *Markov chains.* Cambridge University Press.

[117] Krak, T., Bock, J. D. and Siebes, A. (2017). *Imprecise continuous-time markov chains.*

[118] Erreygers, A. and Bock, J. D. (2017). Imprecise continuous-time markov chains: Efficient computational methods with guaranteed error bounds. In *PMLR: Proceedings of machine learning research.*

[119] Krak, T., Erreygers, A. and Bock, J. D. (2019). An imprecise probabilistic estimator for the transition rate matrix of a continuous-time markov chain. In *Uncertainty modelling in data science.*

[120] Erreygers, A. and Bock, J. D. (2018). *Computing inferences for large-scale continuous-time markov chains by combining lumping with imprecision.*

[121] Inamura, Y. (2006). *Estimating continuous time transition matrices from discretely observed data.* Bank of Japan.

[122] Ball, F. and Yeo, G. F. (1993). Lumpability and marginalisability for continuous-time markov chains. *Journal of Applied Probability* **30** 518–28.

[123] Krpelik, D., Aslett, L. J. M. and Coolen, F. P. A. (2021). Simultaneous sampling for robust markov chain monte carlo inference. In *Advances in uncertainty quantification and optimization under uncertainty with aerospace applications* (M. Vasile and D. Quagliarella, ed) pp 173–83. Springer International Publishing, Cham.

[124] O'Neill, B. (2009). Importance sampling for bayesian sensitivity analysis. *International Journal of Approximate Reasoning* **50** 270–8.

[125] Troffaes, M. C. M. (2018). Imprecise monte carlo simulation and iterative importance sampling for the estimation of lower previsions. *International Journal of Approximate Reasoning* **101** 31–48.

[126] Brooks, S., Gelman, A., Jones, G. and Meng, X.-L. (2011). *Handbook of markov chain monte carlo*. Chapman; Hall/CRC.

[127] Aumann, R. J. (1965). Integrals of set-valued functions. *Journal of Mathematical Analysis and Applications* **12** 1–12.

[128] Artstein, Z. and Vitale, R. A. (1975). A strong law of large numbers for random compact sets. *The Annals of Probability* **3** 879–82.