Estimation of imprecise reliability of systems using random sets and Monte-Carlo resampling procedures

Y. Hou [1] , M. Sallak [2] , and W. Schön [3]

Abstract

This paper is divided into three parts: First, it introduces the use of random sets for reliability assessment of components with rare failure events. The proposed approach is based on the use of operations defined in the random set framework (expectations, confidence intervals, etc.) to obtain upper and lower bounds and confidence intervals of components reliability without assuming any prior distribution about their lifetimes. Then, instead of using failure probabilities calculated directly from each component's observation in order to obtain system reliability, we propose to construct pseudo-system observations directly from components observations in order to obtain the interval system reliability. Finally, the proposed approach is applied on the evaluation of reliability of large-scale systems with very large fault trees and censored reliability data by using Monte-Carlo re-sampling procedure. A comparison with classical probabilistic approaches is also done.

Keywords: Reliability, random set theory, rare failure events, large systems, confidence intervals, Monte-Carlo simulations, re-sampling.

1 INTRODUCTION

The purpose of a reliability assessment is to predict the probability that the system is operating during a specified time interval $[0, t]$ (Barlow and Proschan 1975). The system

[1]PhD student at Sorbonne Universités, Université de Technologie de Compiègne, CNRS, Laboratoire Heudiasyc, UMR 7253, Centre de Recherches de Royallieu CS 60 319, 60 203 Compiègne cedex, France (e-mail: yunhui.hou@utc.fr).

[2]Associate Professor at Sorbonne Universités, Université de Technologie de Compiègne, CNRS, Laboratoire Heudiasyc, UMR 7253, Centre de Recherches de Royallieu CS 60 319, 60 203 Compiègne cedex, France (e-mail: sallakmo@utc.fr).

[3]Professor at Sorbonne Universités, Université de Technologie de Compiègne, CNRS, Laboratoire Heudiasyc, UMR 7253, Centre de Recherches de Royallieu CS 60 319, 60 203 Compiègne cedex, France (e-mail: wschon@utc.fr).

reliability is computed from the system components failure probabilities at a given time t. However, these probabilities can be difficult to be estimated in the presence of insufficient failure data (De Rocquigny 2008). This is particularly the case for highly reliable components with rare failure events (or zero failures). Furthermore, some rare failure events may cause major degradation in systems and should never be neglected during the system reliability computing process. In this case, the components failure probabilities are not only affected by randomness (aleatory uncertainty) but also by imprecision (epistemic uncertainty).

Indeed, during the last years, the dependability and risk assessments community has recognized that there are different sources/types of uncertainties that play an important role in availability and risk evaluation (Winkler 1996; Aven 2011). The most common distinction is to divide uncertainties into aleatory and epistemic uncertainties (Apostolakis 1990; Helton and Burmaster 1996). The former comes from the natural variability of a random event (for example failure or reparation of a component), while the latter represents the lack of knowledge. Aleatory uncertainty is often referred to as irreducible uncertainty because a better understanding of the natural phenomena cannot reduce it. On the contrary, improving our background knowledge can reduce our epistemic uncertainty therefore, we call it reducible uncertainty. These considerations have fostered the use of other alternative theories for taking into account both aleatory and epistemic uncertainties such as imprecise probabilities (Walley 1991), belief functions theory (Dempster 1967; Shafer 1976), possibility theory (Dubois and Prade 1988), random set theory (Matheron 1975), etc.

The rare failure events problem considered in this work is the condition that there is no components failure observed in historical data (or only a very small number of failures). That is what we call components with rare failure events. We then have to raise the natural question: when a component has no failure, how can we consider its reliability parameters? Should we consider that its reliability is equal to one because the empirical estimator gives zero failure probability and zero variance. This is not reasonable, given that such highly reliable components are critical precisely because their failures are rare.

Indeed, the rare failure events problem is mainly caused by the following situations:

- Small sample size: In many systems such as intelligent transportation systems, the sample size of some components is very small. For example, in case when the tested component is a new equipment, there is no previous test data for reliability analysis; some of the components are very expensive or the test cannot be carried out frequently due to some cost considerations. Using Bayesian estimates, we should make some assumptions such as the choice of prior distributions. However, prior distributions may introduce more uncertainty in the reliability estimation results.
- Highly reliable component/system: For highly reliable system or component, the frequency of failure occurrence is very low so that a large amount of tests are needed to obtain one first failure observation, which is not possible due to limited test expense and time.

The difficulty that some research found when estimating the reliability of such systems or components is mainly due to the non distinction between the following two types of uncertainties:

- Aleatory uncertainty which is caused by natural variability of random phenomena whose behavior can be different even under the same condition. This type of uncertainty has been long time studied by probability theory and is often quantified by dispersion measures such as variance.
- Epistemic uncertainty which is caused by lack of information. This type of uncertainty has been represented under framework of uncertainty theories by set-value measures such as belief functions, possibility theory, imprecise probabilities, etc.

In this paper, we propose to estimate the system reliability (systems composed of components with rare failure events) and its associated uncertainty by representing both types of uncertainties under the same framework without additional information (such as prior distributions). Random set theory has been proposed because on one hand it is a general form of random variable in classical probability theory so that several existing tools can be also used in our approach (expected values, confidence intervals, etc.); on the other hand, the probability distribution of a random set is measured by set-valued functions which fits the basic idea of quantifying uncertainty in many uncertainty theories such as belief functions theory and imprecise probabilities.

The random set theory is a mathematical theory which can handle in a unique framework both aleatory and epistemic uncertainties. It is an extension of probability theory to set-valued rather than point-valued maps (Matheron 1975). The random set theory was first applied in statistic (Robbins 1944) and stochastic geometry (Kendall 1974). More recently, the random set theory was also applied in problem inference from incomplete data (Horowitz et al. 2003). The authors considered that the bounds of population parameters can be estimated consistently by replacing the population distribution of the data with the empirical distribution in the functional that give the bounds. In (Tamer 2003), the author considered an example of an incomplete econometric structure which is used by economists to make simplifying assumptions and to avoid multiplicity. A bivariate simultaneous discrete response model which is a stochastic representation of equilibria in a two-person discrete game was studied. In (Tonon 2004), the authors proposed a rock slope stability analysis based on distinct element method and random sets input parameters (rock density and normal stiffness of joints). In (Oberguggenberger and Fellin 2008), the authors proposed another work for the estimation of reliability of structures using random sets theory and geotechnical applications. They particularly used random set models constructed from measurement data by non-parametric methods using inequalities of Tchebycheff type. In (Wang 2010), the author proposed another form of imprecise probabilities based on generalized intervals for reliability assessment of systems. The advantage of using generalized intervals is the fact that they improve the algebraic properties of interval-valued probabilities and simplify the calculus.

In our knowledge, the problem of assessing system reliability using random set expectation and confidence intervals, observations of lifetimes of components (and the number of observed failures), structure function, and pseudo-system observations was never be tackled. This paper aims to overcome the problem of the computation of systems reliability in the presence of rare failure events by proposing the use of random set theory and thus obtaining upper and lower bounds of system and component reliabilities without assuming any prior distribution about failures of these components. Particularly, we will prove that the proposed propagating method based on random set theory is more efficient than imprecise probability approach based on Monte-Carlo simulations. The motivation to choose random set theory is due to the fact that random set theory is a generalization of classical probability theory on random elements(Goodman et al. 2006; Molchanov 2005; Nguyen 2006) which offers the possibility to express both aleatory and epistemic uncertainties. Random set theory is not a unified framework for all uncertainty theories such as belief functions theory, possibility theory, etc. However, belief functions theory can be expressed formally within the framework of random set theory (Nguyen 2006). Moreover, random set theory can be viewed as specific form of imprecise probabilities (Walley 2000). Random set theory has been proposed in our work because on one hand it is a general form of random variable in classical probability theory which means that several existing tools can be also used in our approach (confidence interval, etc.); on the other hand, the probability distribution of a random set is measured by a set-value functions so that it fits the basic idea of quantifying uncertainty in many uncertainty theories.Thus, we are convinced that it will be interesting to introduce the use of such theory in reliability assessment of systems when considering components with highly reliable components. This allows one to use random sets as a bridge between uncertainty measures in decision analysis in the field of dependability.

The rest of this paper is organized as follows. Section 2 reviews the preliminaries about random set theory. Section 3 introduces the proposed method for failure probability estimation of components. Section 4 constructs the propagation method to obtain reliability of systems from components failure events and details the complexity of the proposed approach. Section 5 illustrates a case study to demonstrate the effectiveness of the proposed approach by estimating the reliability of large-scale systems with very large fault trees and censored reliability data. Section 6 concludes the whole paper.

2 RANDOM SET THEORY

In this section, we begin with reviewing basic elements of random set theory. Before introducing the definition of closed random set, we firstly review the definition of random variable on $\mathbb{R}$. Then we introduce the definitions of expectations and confidence intervals in random set theory framework.

2.1 Basic definitions

Let us consider the probability space $(\Omega, \mathcal{A}, P)$, where Ω is the sample space (the set of all possible outcomes), $\mathcal{A}$ is a collection of events (each event is a set containing zero or more outcomes), and P the probability measure which allows the assignment of probabilities to the events such that $P : \mathcal{A} \rightarrow [0, 1]$.

Definition 1. A random variable (real-valued) on the probability space $(\Omega, \mathcal{A}, P)$ is a measurable real map $X : \Omega \rightarrow \mathbb{R}$, where the property of measurability means that for every compact set $K \subset \mathbb{R}$, the set $\{\omega : X(\omega) \in K\}$ is an event in the probability space.

In this paper, we consider only closed sets, i.e. the sets whose complement are open sets.

Definition 2. (Matheron 1975) A random set on the probability space $(\Omega, \mathcal{A}, P)$ is a measurable map $X : \Omega \rightarrow \mathcal{F}$, where $\mathcal{F}$ is a family of closed subsets of $\mathbb{R}$, and the property of measurability means that for every compact set $K \subset \mathbb{R}$, the set $\{\omega : X(\omega) \cap K \neq \emptyset\}$ is an event in the probability space.

Hence, random sets are random variables whose values are sets. In other words, the theory of random sets includes the classical case of random variables as a special case.

Definition 3. (Matheron 1975) The corresponding probability law (hitting probability) of a random closed set X is defined as $T(K) = P\{X \cap K \neq \emptyset\}$ for all compact sets $K \subset \mathbb{R}$. Hence, we have $T(\emptyset) = 0$ and $T(\Omega) = 1$.

Example 1. Let us consider a random closed set X with a unique element on $\mathbb{R}$: $X = \{Y\}$ where Y is a random variable on $\mathbb{R}$. X is a (singleton) random set. For all compact $K \subset \mathbb{R}$, the hitting probability of X is given by $T(K) = P\{K \cap X \neq \emptyset\} = P\{Y \in K\}$ which is exactly the probability of a random variable.

Example 2. Let us consider a random closed set X defined as $X = \{x \in \mathbb{R} | x \leq Z\} = (-\infty, Z]$ where Z is a random variable. For all compact $K \subset \mathbb{R}$, the hitting probability of X is given by

$$\begin{aligned} T(K) &= P\{K \cap X \neq \emptyset\} \\ &= P\{K \cap (-\infty, Z] \neq \emptyset\} \\ &= P\{\exists x \in K | x \leq Z\} \end{aligned}$$

2.2 Expectation of random sets

This subsection introduces the selection expectation (also called the Aumann expectation) which is the best and most used concept of expectation for random sets.

Definition 4. (Aumann 1965) A random point ξ is said to be a selection of a random set X if $P(\xi \in X) = 1$.

A random set can be approximated by all its selections. A random point/variable is called integrable if its expected value exists. The expectation of a random set is defined by grouping the expected value of all its integrable selections.

Definition 5. (Aumann 1965) The expectation $\mathbb{E}X$ of a random set X on $\mathbb{R}$ is the closure of the family of all expectations for its integrable selections, i.e. $\mathbb{E}X = \{\mathbb{E}\xi | \xi \in \mathcal{T}(X)\}$, where $\mathcal{T}(X)$ is the set of all integrable selections of X.

Example 3. When considering $X = \{Y\}$ as defined in Example 1, there is only one selection for X given by $\mathcal{T}(X) = \{Y\}$, so that $\mathbb{E}X = \{\mathbb{E}Y\}$.

Example 4. When considering X as defined in Example 2. The selections are given by $S = Z - s$, where $s \in \mathbb{R}^+$ is a constant. We have $\mathcal{T}(X) = \{S|S = Z - s, s \in \mathbb{R}^+\}$. Hence, the expectation of X is given by $\mathbb{E}(X) = \{\mathbb{E}S|S \in \mathcal{T}(X)\} = (-\infty, \mathbb{E}(Z)]$.

2.3 Random intervals

Definition 6. (Gil 1992) A random interval X of $\mathbb{R}$ associated with the probability space $(\Omega, \mathcal{A}, P)$ is a random set of $\mathbb{R}$ associated with that probability space such that it may be characterized by means of a 2-dimensional random variable (X^L, X^U). So that $X(\omega) = [X^L(\omega), X^U(\omega)]$, for all $\omega \in \Omega$, and it will be denoted by $X = [X^L, X^U]$.

Let $f(x^L, x^U)$ the joint probability density function of X^L and X^U, the expectation of X is given by

$$\mathbb{E}[X] = [\mathbb{E}(X^L), \mathbb{E}(X^U)] = [\int_{-\infty}^{+\infty} tf^L(t)dt, \int_{-\infty}^{+\infty} tf^U(t)dt] \tag{1}$$

where $f^L(t)$ and $f^U(t)$ are respectively marginal pdf developed from the joint probability density function $f(x^L, x^U)$ (see (Matheron 1975)).

Example 5. Let $U_1, U_2, , U_n$ be i.i.d random variables following uniform distribution $U(0, 1)$. Let $U_{(i)}$ the ith smallest value among $U_1, U_2, ..., U_n$, i.e. $U_{(1)} \leq U_{(2)} \leq ... \leq U_{(n)}$. We aim to compute the expectation of the random set $X = [U_{(k)}, U_{(k+1)}]$, $(k = 1, ..., n-1)$. According to the theorem given in (David and Nagaraja 2003), the probability distribution of the kth order statistic $U_{(k)}$ is a Beta distribution with parameters k and $n-k+1$, i.e. $U_{(k)} \sim Beta(k, n-k+1)$. Then, using (1) the expectation of $X = [U_{(k)}, U_{(k+1)}]$ is given by $\mathbb{E}X = [\mathbb{E}U_{(k)}, \mathbb{E}U_{(k+1)}] = [\frac{k}{n+1}, \frac{k+1}{n+1}]$.

2.4 Confidence interval

Consider a random variable X whose distribution depends on a parameter $\theta = (\theta_1, ..., \theta_m)$, $\theta \in \Theta \subseteq \mathbb{R}^m$ where Θ is the parameter space. Let $\varphi(\theta) \in \mathbb{R}$ a parameter of the distribution of X which depends on θ.

Given a certain number of observations of X, the random set $\mathcal{C} = [c_1, c_2] \subset \mathbb{R}$ which contains the true value parameter $\varphi(\theta)$ is called a confidence set for $\varphi(\theta)$ at level $1-\alpha$ ($\alpha \in [0,1]$) if for all possible value of θ: $P\{\varphi(\theta) \in [c_1, c_2]\} = 1-\alpha$. Notice that here we use point estimate of $\varphi(\theta)$ to construct the confidence interval.

Now we extend the use of confidence interval to apply on interval estimates of $\varphi(\theta)$. Let a random interval $[a^L, a^U]$ with known cumulative distribution functions F_L and F_U independent from all parameters θ such that $P(a^L \leq \varphi(\theta) \leq a^U) = 1$. The confidence interval of $\varphi(\theta)$, $\mathcal{C}' = [u^L, u^U]$ at level $1-\alpha$ is given by

$$[\hat{u}^L, \hat{u}^U] = [F_L^{-1}(\alpha/2), F_U^{-1}(1-\alpha/2)] \tag{2}$$

where $F_L^{-1}(\alpha)$ and $F_U^{-1}(\alpha)$ are respectively the α-quantile functions of F_L and F_U (see (Matheron 1975)).

3 ESTIMATION OF FAILURE PROBABILITY OF COMPONENTS

In this section, we will explain how to obtain upper and lower expected values and confidence intervals of components failure probabilities from observation data. Note that the state of a component will be represented by a random variable $X \in \{0,1\}$, where 1 and 0 denote respectively the working and failed states of the component. Thus, the failure probability of a component at time t given some observations O will be represented by a random variable P, i.e. $P = P\{X = 0|O\}$.

3.1 Expected values

Let us consider an observation pool $O = \{x_1, x_2, , x_n\}$ which contains n samples of a random variable $X \in \{0,1\}$. The realization x_i is equal to 0 if the component is down at the ith observation and 1 otherwise. We assume that we have observed k failures (k is the number of $x_i = 0$) in the n observations of O.

Theorem 1. Order statistics of the uniform distribution (David and Nagaraja 2003)

Let $U_1, U_2, ..., U_n$ be an i.i.d. sample from the uniform distribution $U(0,1)$. Let $U_{(k)}$ be the kth order statistic from this sample. Then, the probability distribution of $U_{(k)}$ is a Beta distribution with parameters k and $n-k+1$. The expected value of $U_{(k)}$ is

$$E[U_{(k)}] = \frac{k}{n+1}$$

Let $P^L = U_{(k)}$ and $P^U = U_{(k+1)}$ be respectively the k and $k+1$ order statistics of the variables U_i. In this case, we have exactly k variables less than $P = P\{X=0|O\}$ because we have observed k failures, i.e.

$$P^L \leq P \leq P^U$$

Thus, according to Theorem 1, we can write

$$P^L = U_{(k)} \sim Beta(k, n+1-k)$$
$$P^U = U_{(k+1)} \sim Beta(k+1, n-k)$$

and the expected value of P is then given by

$$E[P] \in [E[P^L], E[P^U]] = [\frac{k}{n+1}, \frac{k+1}{n+1}] \quad (3)$$

Let $p = P\{X=0\}$. According to central limit theorem, as the sample number $n \to \infty$,

$$k \to \lfloor np \rfloor$$

where $\lfloor np \rfloor$ denotes the integer part of np such that $\lfloor np \rfloor \leq np \leq \lfloor np \rfloor + 1$. Then as $n \to \infty$,

$$P^L \sim Beta(\lfloor np \rfloor, n+1-\lfloor np \rfloor) \qquad P^U \sim Beta(\lfloor np \rfloor + 1, n - \lfloor np \rfloor)$$

$$\mathbb{E}(P^L) = \frac{k}{n+1} = \frac{\lfloor np \rfloor}{n+1} \qquad \mathbb{E}(P^U) = \frac{k+1}{n+1} = \frac{\lfloor np \rfloor + 1}{n+1}$$

The epistemic uncertainty of P associated with size-limited observations O is quantified by $|\mathbb{E}(P^U) - \mathbb{E}(P^L)| = \frac{1}{n+1}$. The epistemic uncertainty converges to 0 as the number of available observations n converges to infinity.

3.2 Confidence intervals

Since $P^L \sim Beta(k, n+1-k)$ and $P^U \sim Beta(k+1, n-k)$, using Eq. 2, the confidence interval of P at level $1-\alpha$ is given by

$$[u^L, u^U] = [F_L^{-1}(\alpha/2), F_U^{-1}(1-\alpha/2)] \tag{4}$$

Where $F_L^{-1}(\alpha/2)$ and $F_U^{-1}(1-\alpha/2)$ denote respectively the α quantile of the inversion function of $Beta(k, n+1-k)$ and the inverse function of $Beta(k+1, n-k)$.

Then, we can write

$$[u^L, u^U] = [I^{-1}_{\alpha/2}(k, n+1-k), I^{-1}_{1-\alpha/2}(k+1, n-k)] \tag{5}$$

where $I_x(a, b)$ is the regularized incomplete beta function which represents the inverse function of the beta distribution.

where $I_x(a, b)$ is the regularized incomplete beta function which represents the inverse function of the beta distribution.

The confidence interval describes the most probably field for P which takes account of both aleatory and epistemic uncertainty, so that the difference between u^U and u^L becomes smaller but does not converge to 0 as n converges to infinity (the aleatory uncertainty is not reducible).

Example 6. When considering a component X, we have observed 20 trials among which

k failures ($x_i = 0$) are detected. Using (8) and (5), we compute the expected values and confidence intervals of the component failure probability $P = P(X = 0|O)$ as a function of k. The obtained results are shown in Fig. 1. The value $|\mathbb{E}(P^U) - \mathbb{E}(P^L)|$ is constant as the epistemic uncertainty only depends on the amount of information contained in O (the number n of observations). However, the value $u^U - u^L$ has smaller value when k is near 0 or 1 because the aleatory variation of P is smaller with k is near 0 or 1.

3.3 Expected values with uncertain data part

In some reliability studies, the exact time to failure of some test components is unknown. In this case, the data are called censored. In this section, we consider singly-censored data which are more common in controlled studies. That means that all the test components operate for the same amount of time. Components functioning at the end of the study are considered censored data. Failed components are considered exact failures. Thus, we consider an observation pool O of the random variable X with precise value, and the fact that there are also n' observations $O' = \{x'_1, x'_2, ..., x'_{n'}\}$ whose value is uncertain because of censored time. We aim to provide some useful results related to these cases.

Since $X \in \{0, 1\}$, without any assumption, the probability $P = P\{X = 0|O, O'\}$

$$P\{P \in [P\{X = 0|O^L\}, P\{X = 0|O^U\}]\} = 1$$

$$O^L = \{x_1, x_2, ..., x_n, \underbrace{0, \ldots, 0}_{n'}\} \qquad O^U = \{x_1, x_2, ..., x_n, \underbrace{1, \ldots, 1}_{n'}\}$$

Then P can be presented in form of a random interval $[P^L, P^U]$: P^L is the lower random variable bound of $P\{X = 0|O^L\}$ and P^U is the upper random variable bound of $P\{X =$

$0|O^U\}$ so that

$$P^L = U_{(k)} \sim Beta(k, n + n' + 1 - k) \tag{6}$$

$$P^U = U_{(k+1)} \sim Beta(k + 1 + n', n - k) \tag{7}$$

3.4 Case of dependent samples

Let us consider the case when the samples composed of the combinations of components lifetime observations are not independent because we use sampling with replacement for some components, and the considered sample size is much larger than the smallest components observation size.

The samples become not i.i.d, and the proposed Theorem 1 (David and Nagaraja 2003) cannot be directly applied to obtain expectations of upper and lower random variables (order statistics) of the system unreliability $P_{s,t} = P\{Y_t = 0|\tilde{O}_{W_s}\}$.

In general, for samples not necessarily independent or identically distributed, Arnold and Groeneveld (Arnold and Groeneveld 1979) obtained an upper bound on the expected value of the order statistics. Moreover, Aven (Aven 1985) have also presented an alternative upper bound on the expected value of the highest order statistic. More specifically, in (Gascuel and Caraux 1992) the authors proposed a formula for computing general bounds of the expected values of order statistics, and particular bounds when considering variables which are dependent and follow some usual distributions such as uniform distribution, exponential distribution and normal distribution.

Let us consider an observation pool $O = \{x_1, x_2, ..., x_n\}$ which contains n samples of a random variable $X \in \{0, 1\}$. The realization x_i is equal to 0 if the component is down at the *ith* observation and 1 otherwise. We assume that we have observed k failures (k is the number of $x_i = 0$) in the n observations of O.

Theorem 2. Order statistics of the uniform distribution for dependent samples(Gascuel and Caraux 1992)

Let $U_1, U_2, ..., U_n$ be an i.d. dependent sample from the uniform distribution $U(0,1)$. The upper and lower bounds $E[P^L]$ and $E[P^U]$ of the expectation of the kth order statistics of the variables U_k are given by

$$E[P^L] = \frac{k}{2n} \qquad E[P^U] = 1 - \frac{n-k+1}{2n}$$

In this case, we have exactly k variables less than $P = P\{X = 0|O\}$ because we have observed k failures, i.e.

$$P^L \leq P \leq P^U$$

Thus, according to Theorem 2, the expected value of P is then given by

$$E[P] \in [E[P^L], E[P^U]] = [\frac{k}{2n}, 1 - \frac{n-k+1}{2n}] \tag{8}$$

As we can see the lower bound of P for dependent samples is always lower than the lower bound of P for i.i.d. samples ($\frac{k}{2n} \leq \frac{k}{n+1}$ for all $n \geq 1$) . Moreover, the upper bound of P for dependent samples is greater than the upper bound of P for i.i.d. samples if the size of n is high or the difference between n and k is important ($1 - \frac{n-k+1}{2n} - \frac{k+1}{n+1} = n^2 - 2n - kn + k - 1$).

We can conclude that, in general, when we consider dependency of samples, we obtain a more conservative intervals of system reliability.

4 ESTIMATION OF SYSTEM RELIABILITY

In this section, we will explain how to obtain system reliability estimates from component observations data by constructing pseudo system observations and using the parameter estimation method presented in the previous section. The following key assumptions are taken into account:

- System and components are allowed to take only two possible states: either working, or failed.

- Component failures are s-independent. Failure of one component does not impact the failures of the other components.
- The structure function is coherent. That is, improvement of component states cannot damage the system.
- The components are not repairable.

4.1 Basic definitions

The state of a system is determined by the states of all its components using a structure function $\varphi : \{0,1\}^d \rightarrow \{0,1\}$ so that

$$\varphi(X_{1,t}, ..., X_{d,t}) = \begin{cases} 1 & \text{if the system is working at } t \\ 0 & \text{if the system fails before } t \end{cases}$$

where $X_{1,t}, ..., X_{d,t}$ are random variables representing states of each system component $C_1, ..., C_d$ at time t determined by the lifetime $W_1, ..., W_d$, i.e. for each $i = 1, ..., d$

$$X_{i,t} = \mathbf{1}_{\{W_i \geq t\}}$$

The binary random variable Y_t representing the state of a system composed of d independent components at instant t

$$Y_t = \varphi(X_{1,t}, ..., X_{d,t}) = \varphi(\mathbf{1}_{\{W_1 > t\}}, ..., \mathbf{1}_{\{W_d > t\}})$$

Let W_s the lifetime of the system given by

$$W_s = \varphi_T(\mathbf{W}) = \varphi_T(W_1, ..., W_d)$$

where $\varphi_T : (\mathbb{R}^+)^d \rightarrow \mathbb{R}^+$, for example, for parallel systems it takes the maximal lifetime; for serial systems it takes the minimal lifetime. Hence, we also have

$$Y_t = \mathbf{1}_{\{\varphi_T(\mathbf{W})>t\}} = \mathbf{1}_{\{\varphi_T(W_1,...,W_d)>t\}}$$

The system reliability $R_{s,t}$ can be given by

$$\begin{aligned} R_{s,t} &= P(Y_t = 1) \\ &= P(W_s > t) \\ &= P(\varphi_T(W_1, ..., W_d) > t) \end{aligned}$$

We also recall that the system structure function φ is computed from minimal paths P_i (minimal sets of components such that if all the components in the set are working, the system will be in a working state) or minimal cuts C_i (minimal sets of components such that if all the components in the set fail, the system will be in a failed state) as follows:

$$\varphi_X(X_{1,t}, ..., X_{d,t}) = 1 - \prod_{j=1}^{k}(1 - \prod_{i\in P_j} X_{i,y}) = \prod_{j=1}^{k'}(1 - \prod_{i\in C_j}(1 - X_{i,t}))$$

where k and k' denote respectively the number of minimal paths and cuts.

The system unreliability can also be expressed according to the failure probabilities of its components by the following formula

$$U_{s,t} = \bar{R}(\mathbf{q}_t) = \mathbb{P}(\varphi(X_{1,t}, ..., X_{d,t}) = 0) = \sum_{\mathbf{x}:\varphi(\mathbf{x})=0} \prod_{i=1}^{d} q_i^{x_i}(1 - q_i)^{1-x_i} \tag{9}$$

where $\mathbf{q}_t = (q_{1,t}, ..., q_{d,t}) = (P(X_{1,t} = 0), ..., \mathbb{P}(X_{d,t} = 0)) = (P(W_1 \leq t), ..., P(W_d \leq t))$ for $t \in \mathbb{R}^+$.

4.2 Pseudo observation construction

Let $O_{W_i} = \{w_{1,i}, ..., w_{n_i,i} \in \mathbb{R}^+\}$ be an observation pool which contains n_i i.i.d lifetime observations of component C_i, and suppose that $n_1 \leq n_2 \leq ... \leq n_d$ (we consider a system with d components).

Instead of using failure probabilities calculated from each component's observation pool to propagate to the system level in order to compute the system unreliability $U_{s,t}$, our proposed idea is to build directly pseudo-observations of W_s by randomly picking one observation from each component observation pool O_{W_i}, building a system composed of those observation combinations and obtaining a sample of W_s through function φ_T.

It is easy to think $O_{\mathbf{W}}$, including all possible combinations of elements from $O_{W_1}, ..., O_{W_d}$ with $n_C = |O_{\mathbf{W}}| = \prod_{l=1}^{d} n_i$ elements, would be a choice for samples to build our pseudo-observation pool. But the observations combinations in $O_{\mathbf{W}}$ can not be i.i.d. because some component observations may be used repeatedly. However, we will show that some subsets of $O_{\mathbf{W}}$, $\tilde{O}_{\mathbf{W}}$, have all their $\tilde{n}_C$ elements (approximately) i.i.d.

Consider that each combination $\mathbf{w}_i = (w_{i,1}, ..., w_{i,d})$ from the chosen subset $\tilde{O}_{\mathbf{W}} \subset O_{\mathbf{W}}$ with $\tilde{n}_C = min(n_1, ..., n_d) = n_1$, is obtained by randomly picking a sample $w_{i,j}$ from each O_{W_j}, $1 \leq j \leq d$, without replacement. There is no sample from each observation pool O_{W_j} used more than once when constructing $\tilde{O}_{\mathbf{W}}$; $n_j - n_1$ samples from O_{W_j},$1 \leq j \leq d$ are not used.

Let $\tilde{n}_C = max(n_1, ..., n_d) = n_d$. For each combination $\mathbf{w}_i = (w_{i,1}, ..., w_{i,d})$ from $\tilde{n}_C$, $w_{i,d}$ is given by randomly picking a sample from O_{W_d} without replacement; $w_{i,j}$, $1 \leq j < d$, is given by randomly picking a sample from O_{W_j} with replacement because there are not enough observations. Although all the other components have less element observations than C_d, the number of repeated samples picked into $\tilde{O}_{\mathbf{W}}$ is limited when the difference in sample size $n_d - n_j$, $1 \leq j < d$, is not very large.

For $n_1 < \tilde{n}_C < n_d$. For $\tilde{O}_{\mathbf{W}}$ having at least n_{med} observations where n_{med} denotes the middle value in the list of possibles values of $n : \{n_1, n_2, ..., n_d\}$, each $w_{i,j}$ is picked randomly without replacement from O_{W_j}; for $\tilde{O}_{\mathbf{W}}$ having less than n_{med} observations, each $w_{i,j}$ is picked randomly with replacement from O_{W_j}. It is more likely to be an identically and independently distributed observations than the precedent case with $\tilde{n}_C = max(n_1, ..., n_d)$, as less samples are repeatedly used and treated as new observations.

In summary, the size of $\tilde{O}_{\mathbf{W}}$, $\tilde{n}_C$, is fixed between $min(n_1, ..., n_d)$ and $max(n_1, ..., n_d)$. For component whose sample size is smaller than $\tilde{n}_C$, component samples are re-sampled with replacement; while for those whose sample size is larger than $\tilde{n}_C$, component level re-sampling is without replacement.

After all, the pseudo observations of W_s, $\tilde{O}_{W_s} = \{\tilde{w}_{s,1}, ..., \tilde{w}_{s,n_s}\}$ is obtained by applying the function φ_T on each component from $\tilde{O}_{\mathbf{W}}$, i.e.

$$\tilde{O}_{W_s} = \{\tilde{w}_{s,j} : \tilde{w}_{s,j} = \varphi_T(\tilde{\mathbf{w}}_j), \mathbf{w}_j \in \tilde{O}_{\mathbf{W}}\}$$

where $n_s = \tilde{n}_C$.

Since all observations in $\tilde{O}_{\mathbf{W}}$ are i.i.d, the pseudo observations in $\tilde{O}_{W_s}$ of W_s are also seen as (approximately) independent and identically distributed ones.

4.3 Estimation of system failure probability/distribution

Let random variable $P_{s,t}$ denote the system unreliability given all components observations, i.e.

$$\begin{aligned} P_{s,t} &= P\{Y_t = 0 | O_{W_1}, ..., O_{W_d}\} \\ &= P\{Y_t = 0 | \tilde{O}_{\mathbf{W}}\} \\ &= P\{Y_t = 0 | \tilde{O}_{W_s}\} \end{aligned}$$

where $\tilde{O}_{\mathbf{W}}$ is a randomly re-sampled observation pool deviated from $O_{W_1}, ..., O_{W_d}$ with $n_s = \tilde{n}_C$ elements according to the construction procedure mentioned in previous section; $\tilde{O}_{W_s}$ is obtained by applying the function $\varphi_T(\mathbf{w})$ on each element in $\tilde{O}_{\mathbf{W}}$. Here given $\tilde{O}_{\mathbf{W}}$, for a fixed instant of time t, we can apply the exactly same estimation method of failure probability as the one at component level by introducing the random interval $[P^L_{s,t}, P^U_{s,t}]$ where

$$P^L_{s,t} = U_{(k_{s,t})} \sim Beta(k_{s,t}, n_s + 1 - k_{s,t}); P^U_{s,t} = U_{(k_{s,t}+1)} \sim Beta(k_{s,t}+1, n_s - k_{s,t})$$

The expected value of $U_{S,t}$ is then given by

$$\mathbb{E}(U_{s,t}) = 1 - \mathbb{E}(R_{s,t}) = \mathbb{E}[P_{s,t}] \in [\mathbb{E}(P^L_{s,t}), \mathbb{E}(P^U_{s,t}))] = [\frac{k_{s,t}}{n_s+1}, \frac{k_{s,t}+1}{n_s+1}] \qquad (10)$$

Using (2), the confidence interval at level $1-\alpha$ for P is given by

$$[u^L, u^U] = [I^{-1}_{\alpha/2}(k_{s,t}, n_s + 1 - k_{s,t}), I^{-1}_{1-\alpha/2}(k_{s,t}+1, n_s - k_{s,t})] \qquad (11)$$

The advantage of this method is that the probability distribution at system level is very complicated if we propagate the reliability distribution of each component using the structure function. For example, if all components follows exponential distribution, only a serial system has exponential distributed reliability. Our method uses directly empirical distribution functions which always exist and can present all possible distributions.

4.4 Estimation of $k_{s,t}$

The pseudo observation pools $\tilde{O}_{W_s}$ are randomly drawn from $O_{W_1}, ..., O_{W_d}$, we need to estimate the number of system failure $k_{s,t}$ among n_s trials from $\tilde{O}_C(\tilde{O}_{W_s})$.

MLE estimation We firstly estimate empirically failure probability $q_{i,t} = P\{W_i \leq t | O_{W_i}\}$ for each system component

$$\hat{q}_{i,t} = \frac{\sum_{j=1}^{n_i} \mathbf{1}_{w_{i,j} \leq t}}{n_i} \qquad (12)$$

and then calculate the empirical system unreliability by propagating component level measures to system level, i.e.

$$\hat{U}_{s,t} = \bar{R}(\hat{\mathbf{q}}_t) \qquad (13)$$

where $\hat{\mathbf{q}}_t = (\hat{q}_{i,t}, ..., \hat{q}_{d,t})$. This method estimates $k_{s,t}$ by $\hat{k}_{s,t} = \hat{U}_{s,t} n_s$.

Algorithm 1 Estimate $k_{s,t}$ with MLE

Require: Component observations O_{W_1}, ..., O_{W_d}; instant t
  for $i = 1$ to d **do**
    Estimate component unreliability at instant t: $\hat{q}_{i,t} = \dfrac{\sum_{j=1}^{n_i} \mathbf{1}_{\{w_{i,j} \leq t\}}}{n_i}$
  end for
  Calculate the probability of system failure according to components unreliability: $\hat{U}_{s,t} = \bar{R}(\hat{\mathbf{q}}_t)$
  Estimate the value of $k_{s,t}$: $\hat{k}_{s,t} = \hat{U}_{s,t} n_s$

Monte Carlo re-sampling/simulation For large systems, the propagation of probabilistic measures are very complicated. Here we use Monte Carlo method to simulate the construction of the pseudo system level lifetime observation pool $\tilde{O}_{\mathbf{W}} = \{w_{s,1}, ..., w_{s,n_s}\}$ with $n_s = \tilde{n}_C$ trials inside and estimate empirically the number of failures having occurred before given instant t using the definition equation

$$\hat{k}_{s,t} = \sum_{j=1}^{n_s} \mathbf{1}_{\{w_{s,j} \leq t\}}$$

with N simulation executions.

4.5 Case with empty/censored component observation pool(s)

Consider that there are n'_i samples $O'_{W_i} = \{w'_{i,1}, ..., w'_{i,n'_i}\}$ censored with maximal observation time t_i^{max} for a component C_i. The only information available is

$$w'_{i,j} \in [t_i^{max}, +\infty), \qquad j = 1, ..., n'_i, \qquad i = 1, ..., d$$

In order to merge the precise observation part and censored observations, the upper and lower observation pools are built without additional assumption as follows

$$O_{W_i}^L = \{w_{i,1}, ..., w_{i,n_i}, \underbrace{t_i^{max}, \ldots, t_i^{max}}_{n'_i}\} \qquad O_{W_i}^U = \{w_{i,1}, ..., w_{i,n_i}, \underbrace{+\infty, \ldots, +\infty}_{n'_i}\}$$

Algorithm 2 Estimate $k_{s,t}$ with Monte Carlo re-sampling/simulation

Require: Component observations O_{W_1}, ..., O_{W_d}; number of simulation trials N; instant t
  for $j = 1$ to N **do**
    for $i = 1$ to d **do**
      if $n_i < n_s$ **then**
        Select randomly with replacement n_s samples from O_{W_i} : $\tilde{w}_{i,1}, ..., \tilde{w}_{i,n_s}$
      else
        Select randomly without replacement n_s samples from O_{W_i} : $\tilde{w}_{i,1}, ..., \tilde{w}_{i,n_s}$
      end if
    end for
    for $m = 1$ to n_s **do**
      Construct pseudo system lifetime observations:
      $\tilde{w}_{s,m} = \varphi_T(\tilde{w}_{1,m}, ..., \tilde{w}_{d,m})$
    end for
    $k_{s,t,j} = \sum_{i=1}^{n_s} \mathbf{1}_{\{\tilde{w}_{s,i}<t\}}$;
  end for
  Aggregate the sampled value of $k_{s,t}$ by average: $\hat{k}_{s,t} = \dfrac{\sum_{j=1}^{N} k_{s,t,j}}{N}$

Then $P_{s,t}$ can be presented in form of a random interval $[P^L_{s,t}, P^U_{s,t}]$: $P^L_{s,t}$ is the lower random variable bound of $P\{W_s \leq t | O^L_{W_1}, ..., O^L_{W_d}\}$ and P^U is the upper random variable bound of $P\{W_s \leq t | O^U_{W_1}, ..., O^U_{W_d}\}$ so that

$$P^L_{s,t} = U_{(k)} \sim Beta(k^L_{s,t}, n_s + 1 - k^L_{s,t}) \qquad P^U_{s,t} = U_{(k+1)} \sim Beta(k^U_{s,t} + 1, n_s - k^U_{s,t}) \qquad (14)$$

where $k^L_{s,t}$ and $k^U_{s,t}$ are numbers of system failures observed before instant t in pseudo observations $\tilde{O}^L_{\mathbf{W}}$ and $\tilde{O}^U_{\mathbf{W}}$ which are deviated respectively from $O^L_{W_1}, ..., O^L_{W_d}$ and $O^U_{W_1}, ..., O^U_{W_d}$. Both $\tilde{O}^L_{\mathbf{W}}$ and $\tilde{O}^L_{\mathbf{W}}$ are constructed by random re-sampling. Then algorithms estimating $\hat{k}^L_{s,t}$ and $\hat{k}^U_{s,t}$ are given in Algorithms 3 and 4.

Using (2), the confidence interval at level $1 - \alpha$ for $P_{s,t}$ is given by

$$[u^L, u^U] = [I^{-1}_{\alpha/2}(k^L_{s,t}, n_s + 1 - k^L_{s,t}), I^{-1}_{1-\alpha/2}(k^U_{s,t} + 1, n_s - k^U_{s,t})]$$

Algorithm 3 Estimate $k_{s,t}$ with MLE

Require: Component observations O_{W_1}, ..., O_{W_d}; instant t
  for $i = 1$ to d **do**
    if $t_i^{max} < t$ **then**
      Estimate component unreliability at instant t: $\hat{q}_{i,t}^L = \hat{q}_{i,t}^U = \dfrac{\sum_{j=1}^{n_i} \mathbf{1}_{\{w_{i,j} \le t\}}}{n_i + n_i'}$
    else
      Estimate component unreliability at instant t:
      $\hat{q}_{i,t}^L = \dfrac{\sum_{j=1}^{n_i} \mathbf{1}_{\{w_{i,j} \le t\}}}{n_i + n_i'}$; $\hat{q}_{i,t}^U = \dfrac{\sum_{j=1}^{n_i} \mathbf{1}_{\{w_{i,j} \le t\}} + n_i'}{n_i + n_i'}$
    end if
  end for
  Calculate the probability of system failure according to components unreliability: $\hat{U}_{s,t}^L = \bar{R}(\hat{\mathbf{q}}_t^L)$; $\hat{U}_{s,t}^U = \bar{R}(\hat{\mathbf{q}}_t^U)$
  Estimate the value of $k_{s,t}$: $\hat{k}_{s,t}^L = \hat{U}_{s,t}^L n_s$; $\hat{k}_{s,t}^U = \hat{U}_{s,t}^U n_s$

Algorithm 4 Estimate $k_{s,t}$ with Monte Carlo re-sampling/simulation

Require: Component observations O_{W_1}, ..., O_{W_d}; number of simulation trials N; instant t
  for $j = 1$ to N **do**
    for $i = 1$ to d **do**
      Composite the upper and lower observation pools of component C_i with $O_{W_i} = \{w_{i,1}, ..., w_{i,n_i}\}$
      $O_{W_i}^L = \{w_{i,1}, ..., w_{i,n_i}, \underbrace{t_i^{max}, \ldots, t_i^{max}}_{n_i'}\}$; $O_{W_i}^U = \{w_{i,1}, ..., w_{i,n_i}, \underbrace{+\infty, \ldots, +\infty}_{n_i'}\}$
      if $n_i + n_i' < n_s$ **then**
        Select randomly with replacement n_s samples from $O_{W_i}^L$ and $O_{W_i}^L$ respectively: $\{\tilde{w}_{i,1}^L, ..., \tilde{w}_{i,n_s}^L\}$; $\{\tilde{w}_{i,1}^U, ..., \tilde{w}_{i,n_s}^U\}$
      else
        Select randomly without replacement n_s samples from $O_{W_i}^L$ and $O_{W_i}^L$ respectively: $\{\tilde{w}_{i,1}^L, ..., \tilde{w}_{i,n_s}^L\}$; $\{\tilde{w}_{i,1}^U, ..., \tilde{w}_{i,n_s}^U\}$
      end if
    end for
    for $m = 1$ to n_s **do**
      Construct upper and lower pseudo system lifetime observations:
      $\tilde{w}_{s,m}^L = \varphi_T(\tilde{w}_{1,m}^L, ..., \tilde{w}_{d,m}^L)$; $\tilde{w}_{s,m}^U = \varphi_T(\tilde{w}_{1,m}^U, ..., \tilde{w}_{d,m}^U)$
    end for
    Calculate the uppper and lower $k_{s,t,j}$ estimates: $k_{s,t,j}^L = \sum_{i=1}^{n_s} \mathbf{1}_{\{\tilde{w}_{s,i}^L < t\}}$; $k_{s,t,j}^U = \sum_{i=1}^{n_s} \mathbf{1}_{\{\tilde{w}_{s,i}^U < t\}}$
  end for
  Aggregate the sampled value of $k_{s,t}$ by average: $\hat{k}_{s,t}^L = \dfrac{\sum_{j=1}^{N} k_{s,t,j}^L}{N}$ $\hat{k}_{s,t}^U = \dfrac{\sum_{j=1}^{N} k_{s,t,j}^U}{N}$

4.6 Complexity of algorithms

In order to compare the complexity of our proposed methods with classical probabilistic assessment(PA) methods, we need to focus on the estimation of $k_{s,t}$. The first proposed algorithm with MLE uses the classical assessment function $\bar{R}(\mathbf{q})$ (9) to estimate $k_{s,t}$. This function requires the set of minimal cut-sets whose building algorithms as well as the function's inclusion-exclusion calculation have at least exponential complexity. In case of large scale systems, without truncation on cut-sets length or ignorance of some components, it is very difficult. While the second proposed algorithm with Monte Carlo simulation uses the function $\varphi_T(\mathbf{w})$ which is easier to build. For example, for a fault tree composed of N_G gates and d components, the complexity of function $\varphi_T(\mathbf{w})$ is in order of $O(N_G.d)$. Besides, once the pseudo observations are obtained, there rests only light calculations for different instant of time t.

By avoiding the NP-difficult MCS building procedure, the second proposed algorithm reduces efficiently the execution cost (time/complexity) for estimating the system unreliability as well as its confidence interval.

5 NUMERICAL APPLICATION

5.1 Parallel/serial systems

Consider a system composed of three independent components, C_1, C_2 and C_3, whose lifetime observations are given in Table 1 (Fig. 2) with different possible structures shown in Fig. 3.

Using Algorithm 1 and Algorithm 2 with $N = 1000$, $n_s = 4$ and $1 - \alpha = 0.95$, we obtained results with negligible differences. In this case, the empirical reliability estimates P_{MLE}, $P_{Jeffreys}$, and $P_{Uniform}$ obtained by probabilistic structure function $R(\hat{\mathbf{q}}_t)$(9) with $\hat{\mathbf{q}}_t$ given by (12) are always bounded by our upper and lower estimates. A system fails when the last working component in one of the minimal cut-set stops working. In our obtained stair-step graphs(Fig. 4), the instants where the estimation and the CI values drop correspond to one of component lifetime observations in O_{W_i}(Table 1): 21 , 28, 30, 30, 36, 55, 56, 60, 73, 78, 84. The reliability of parallel systems mostly depends on the most reliable component which is C_3 in our case. As shown in Fig. 4a, the result reliability measures do not start decreasing until the instant $t = 56$, the first failure point observation of component C_3. The reliability of serial systems depends on the least reliable component which is C_2 in our case; once one of the components fails the system fails. The curves stop decreasing on function of time after the instant $t = 56$ where the last samples of C_2 fails(Fig. 4b). All components in the two previous examples have equal importance. The magnitude of each decrease is only propositional to the number of samples corresponding to this instant and the weight of such samples in their observation pool($1/n_i$). The two following systems combined parallel and serial structures so that the instants where system reliability estimates start and stop decreasing can be traced using the conclusion obtained in the previous examples. Meanwhile, the component's importance becomes different comparing between components and varies on function of time and component reliability. In Structure 3, the decrease magnitudes caused by samples of C_1 are different on function of time. The shapes of obtained curves are similar to the one of C_1 reliability function. According to all available observations C_3 is much more reliability than the parallel structure composed of the other two component, which indicates that C_1 is critical to the system reliability given this structure and observations. In Structure 4, the shape of all estimates especially P_{MLE} is almost the same as the reliability function of C_3. The serial part according to observations that we have fails much faster than C_3 which is on parallel. In this specific case, C_3 is critical to the system reliability, which confirms what we observed from our results. Moreover, we propose a comparison with approaches based on a uniform prior ($\beta(1,1)$) and Jeffreys' prior ($\beta(1/2,1/2)$) . We choose the uniform prior because the results can be quite different from the MLE if both the number of observations and the number of failures are small. We choose the Jeffreys' prior because it is invariant under re-parametrizations (the inference do not depend on how a model is parameterized). However, whereas Jeffreys' prior is widely accepted in single parameter models, its use in multi-parameters models is more controversial. As we can see our proposed approach is more conservative than the approaches based on different prior.

5.2 Application on large systems (very large fault trees)

In order to illustrate that our proposed approach can be applied efficiently to reliability studies of large systems, we consider the fault trees BAOBAB1 and BAOBAB2 (which are constituted of real-life fault trees using various sources) from Aralia benchmark[1]. The coherent fault trees BAOBAB1 and BAOBAB2 contain respectively 61 and 32 components and both have more than 4000 minimal cutsets (Table 2) such that the calculation of the system reliability using probabilistic assessment by propagation from component level is very expensive.

The reference method is based on assumption that the lifetime of each component C_i, W_i follows an exponential distribution with constant failure rate μ_i estimated by maximum likelihood estimator(MLE): $\hat{\lambda}_i = \frac{n_i}{\sum_{j=1}^{n_i} w_{j,i}}$. The component unreliability at instant t is estimated by $\hat{q}_{i,t} = 1 - exp(-\hat{\lambda}_i.t)$. Then the reference method provides system reliability estimate using Monte Carlo simulation with exponential hypothesis and failure rate estimation mentioned above for each component.

In both examples, the component lifetime sample datasets(Table 3, Table 5) include censored parts with $n_i = 0$ which means all n'_i observations for a certain component C_i have value larger than the maximal observation time t_i^{max}(Table 4, Table 6). For the reference method, since all instants t where the component reliability will be calculated is smaller than t_i^{max}, we have $\hat{\lambda}_i = 0$ and $\hat{q}_{i,t} = 0$.

We obtained reliability estimated value and confidence interval at level and $1 - \alpha = 0.95$ using Algorithm 4 with $N = 1000$ and $n_s = 4$ as well as reference method with 10000 simulations for both fault trees shown in Fig. 5 and Fig. 6.

As mentioned in previous examples, the instants where the results values drop correspond to one of component lifetime observations in O_{W_i}. The reference method result is close

[1]Fault trees distributed by Antoine Rauzy which can be found at: http://www.itu.dk/research/cla/externals/clib/Aralia.zip

to the lower estimated value $\mathbb{E}P_{s,t}^{L}$ is always inside the confidence interval. The graphs presenting the upper and lower estimates and confidence interval bounds for system reliability have stair-step shapes while the reference methods gives smooth curves(MC Exp MLE). It is due to the fact that our method reflects exactly the information giving by available observations without any additional assumption comparing with the reference method which can be seen as infinitely precise as the curve is plotted. However, the exponential hypothesis on component lifetime distribution is not always confirmed, and the estimation of failure rate has very large variance when the size of observation pools is very limited(less than 30 in our cases). Our method has advantages on these points especially using Monte Carlo re-sampling.

6 CONCLUSION

The random set theory is a mathematical theory which can handle in a unique framework both aleatory and epistemic uncertainties. When the components are highly reliable, the random set theory and the construction of pseudo-system observations can be used to estimate upper and lower bounds of system reliability.

It was illustrated that random set theory combined with Monte-Carlo re-sampling procedures is very practical in dealing with epistemic uncertainty due to the fact that we have only few or zero failure events of components. Based on our approach, we can estimate the system reliability within a range in the form of confidence interval. The reliability bounds were obtained after a reasonable number of simulations. We also proposed more efficient algorithms to adopt the framework of random set theory on large-scale systems by simulating the re-sampling procedure. The numerical results obtained for different system configurations confirm the solid theoretical foundation of the proposed approach.

To summarize, our method presents two main advantages:

- The robustness regarding epistemic uncertainty: Indeed our method reflects more adequately the epistemic uncertainty comparing to classical probabilistic approach.

Moreover, in our approach the epistemic uncertainty only depends on the amount observation contained in the observation pool which proves its robustness regarding epistemic uncertainty.

- The efficiency comparing to traditional methods based on minimal links or cuts: Since we use a re-sampling method, system reliability can be derived without the need to evaluate the minimal cut sets as intermediate results. This has major implications for improving the efficiency of system reliability estimation.

Future works will focus on discussing how to determine the parameters of our proposed algorithms (n_s and N) as well as quantifying the error caused by the re-sampling step. We are also interested in applying our method on multi-state system reliability estimation and component importance evaluation. More specifically, The different system multi-state probabilities and system performance indices will be presented in form of random set bounded by random variables following Dirichlet distribution or Beta distribution so that interval form expected values and confidence intervals of system performance indices can be obtained.

ACKNOWLEDGMENT

This work was funded and carried out by the Collegium UTC CNRS INSIS, Région Picardie, and the French National Research Agency, through the project ANR-13-JS03-0007 RECIF.

8 References

Alvarez, D. A. and Hurtado, J. E. (2014). “An efficient method for the estimation of structural reliability intervals with random sets, dependence modeling and uncertain inputs.” *Computers & Structures*, 142(0), 54 – 63.

Apostolakis, G. E. (1990). “The concept of probability in safety assessments of technological systems.” *Science*, 250, 1359–1364.

Arnold, B. C. and Groeneveld, R. A. (1979). “Bounds on expectations of linear systematic statistics based on dependent samples.” *Mathematics of Operations Research*, 4, 441–447.

Aumann, R. (1965). "Intervals of set-valued functions." *Math. Analysis and Appl.*, 12, 1–12.

Aven, T. (1985). "Upper (lower) bounds on the mean of the maximum (minimum) of a number of random variables.." *Journal of Applied Probability*, 22, 723–728.

Aven, T. (2011). "Interpretations of alternative uncertainty representations in a reliability and risk analysis context." *Reliability Engineering and System Safety*, 96, 353–360.

Bae, H.-R., Tonon, F., Grandhi, R., and Pettitx, C. (2006). "Using random set theory to calculate reliability bounds for a wing structure." *Structure and Infrastructure Engineering*, 2, 191–200.

Barlow, R. and Proschan, F. (1975). *Statistical theory of reliability and life testing: probability models*. Holt Rinehart and Winston.

David, H. A. and Nagaraja, H. N. (2003). *Order Statistics (3rd Edition)*.

De Rocquigny, E. (2008). *Uncertainty in Industrial Practice: A Guide to Quantitative Uncertainty Management*. Wiley-Blackwell.

Dempster, A. (1967). "Upper and lower probabilities induced by a multivalued mapping." *Annals of Mathematical Statistics*, 38, 325–339.

Dubois, D. and Prade, H. (1988). *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York.

Gascuel, O. and Caraux, G. (1992). "Bounds on expectations of order statistics via extremal dependences." *Statistics and Probability Letters*, 15, 143–148.

Gil, M. A. (1992). "A note on the connection between fuzzy numbers and random intervals." *Statistics & Probability Letters*, 13(4), 311–319.

Goodman, I., Mahler, R., and Nguyen, H. (2006). *Mathematics of Data Fusion*. Kluwer Academic Publishers.

Helton, J. and Burmaster, D. (1996). "Guest editorial: treatment of aleatory and epistemic uncertainty in performance assessments for complex systems.." *Reliability Engineering and System Safety*, 54, 91–94.

Horowitz, J. L., Manski, C. F., Ponomareva, M., and Stoye, J. (2003). "Computation of

Bounds on Population Parameters When the Data Are Incomplete." *Reliable Computing*, 9(6), 419–440.

Kendall, D. G. (1974). *Foundations of a theory of random sets*. J. Wiley, New York, Chapter Stochastic Geometry.

Matheron, G. (1975). *Random Sets and Integral Geometry*. J.Wiley, New York.

Molchanov, I. (2005). *Theory of Random Sets*. Springer, New York.

Nguyen, H. (2006). *An Introduction to Random Sets*. Chapman and Hall/CRC Press.

Oberguggenberger, M. and Fellin, W. (2008). "Reliability bounds through random sets: Non-parametric methods and geotechnical applications." *Computers & Structures*, 86(10), 1093 – 1101 Uncertainty in Structural Analysis - Their Effect on Robustness, Sensitivity and Design.

Robbins, H. E. (1944). "On the measure of a random set." *I. Ann. Math. Statist.*, 15, 70–74.

Shafer, G. (1976). *A mathematical Theory of Evidence*. Princeton University Press, New Jersey.

Tamer, E. (2003). "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria." *Review of Economic Studies*, 70(1), 147–165.

Tonon, F. (2004). "Using random set theory to propagate epistemic uncertainty through a mechanical system." *Reliability Engineering and System Safety*, 85, 169–181.

Walley, P. (1991). *Statistical reasoning with imprecise Probabilities*. Chapman and Hall, New York.

Walley, P. (2000). "Towards a unified theory of imprecise probability." *International Journal of Approximate Reasoning*, 24, 125–148.

Wang, Y. (2010). "Imprecise probabilities based on generalised intervals for system reliability assessment." *International Journal of Reliability and Safety*, 4, 319–342.

Winkler, R. L. (1996). "Uncertainty in probabilistic risk assessment." *Reliability Engineering & System Safety*, 54(2-3), 127–132.

List of Tables

C_i	n_i	O_{W_i}
1	5	78 73 36 55 28
2	4	30 21 56 30
3	3	60 84 56

Table 1: Parallel/serial systems: Observation for three system components

Fault tree	BAOBAB1	BAOBAB2
Component number	61	32
Number of minimal cutsets (max length=4)	72	395
Number of minimal cutsets (max length=5)	472	1025
Number of minimal cutsets (max length=6)	2684	4805
Number of minimal cutsets (max length=7)	17432	4805
Number of minimal cutsets (max length=8)	25892	4805
Total number of minimal cutsets	46188	4805

Table 2: Complexity information about studied fault trees

C_i	n_i	O_{W_i}	C_i	n_i	O_{W_i}	C_i	n_i	O_{W_i}
1	0		23	3	35,42,2	45	3	54,66,46
2	1	20	24	0		46	4	3,30,7,145
3	0		25	4	93,111,117,23	47	2	18,10
4	5	24,12,33,134,25	26	0		48	2	68,46
5	4	4,28,3,39	27	2	73,35	49	3	124,17,9
6	5	20,26,41,30,24	28	4	138,5,48,30	50	2	160,34
7	0		29	2	73,94	51	4	38,9,67,12
8	0		30	0		52	1	227
9	3	88,42,8	31	4	41,78,15,57	53	4	31,62,155,144
10	2	33,25	32	4	26,152,1,250	54	3	1,9,2
11	5	16,16,19,51,3	33	2	145,6	55	0	
12	3	117,89,32	34	1	61	56	0	
13	0		35	2	7,59	57	4	3,31,4,5
14	1	123665	36	0		58	0	
15	2	3200,1964	37	2	25,91	59	1	26
16	4	3986,466,1513,685	38	1	107	60	4	27,19,16,24
17	2	184,253	39	3	9,5,4	61	5	24,1,14,14,53
18	3	12,389,29	40	3	24,170,61			
19	2	819,158	41	5	1,4,13,12,27			
20	3	318,84,388	42	1	60			
21	5	125,25,25,9,74	43	4	143,5,35,79			
22	5	102,60,105,28,6	44	3	42,8,9			

Table 3: BAOBAB1: Component C_i, $i = 1, ..., d = 62$

C_i	n'_i	t_i^{max}
1	2	1000
3	3	2000
7	3	2000
8	2	2000
13	2	1000
24	3	2000
26	3	2000
30	2	2000
36	2	1000
55	3	2000
56	3	2000
58	3	2000

Table 4: BAOBAB1: Component censored observations

C_i	n_i	O_{W_i}	C_i	n_i	O_{W_i}	C_i	n_i	O_{W_i}
1	1	223	12	1	528	23	2	67,19
2	2	28,253	13	2	269,767	24	1	16
3	3	2,23,20	14	1	122	25	2	3,15
4	4	20,28,18,99	15	3	105,638,1462	26	5	29,8,10,28,75
5	5	167,293,71,90,114	16	5	953,2570,1436,1670,3397	27	0	
6	5	492,97,1287,704,620	17	5	708,1282,1133,3052,34	28	2	14,17
7	5	2068,35,1015,123,27	18	1	19	29	5	6,32,5,23,2
8	4	422,126,254,619	19	1	70	30	0	
9	4	338,375,128,1823	20	0		31	3	10,1,3
10	5	1960,1504,1005,412,469	21	4	17,3,9,1	32	0	
11	2	208,161	22	1	54			

Table 5: BAOBAB2: Component C_i, $i = 1, ..., d = 32$

C_i	n'_i	t_i^{max}
20	2	1000
27	3	2000
30	3	2000
32	2	2000

Table 6: BAOBAB2: Component censored observations

List of Figures

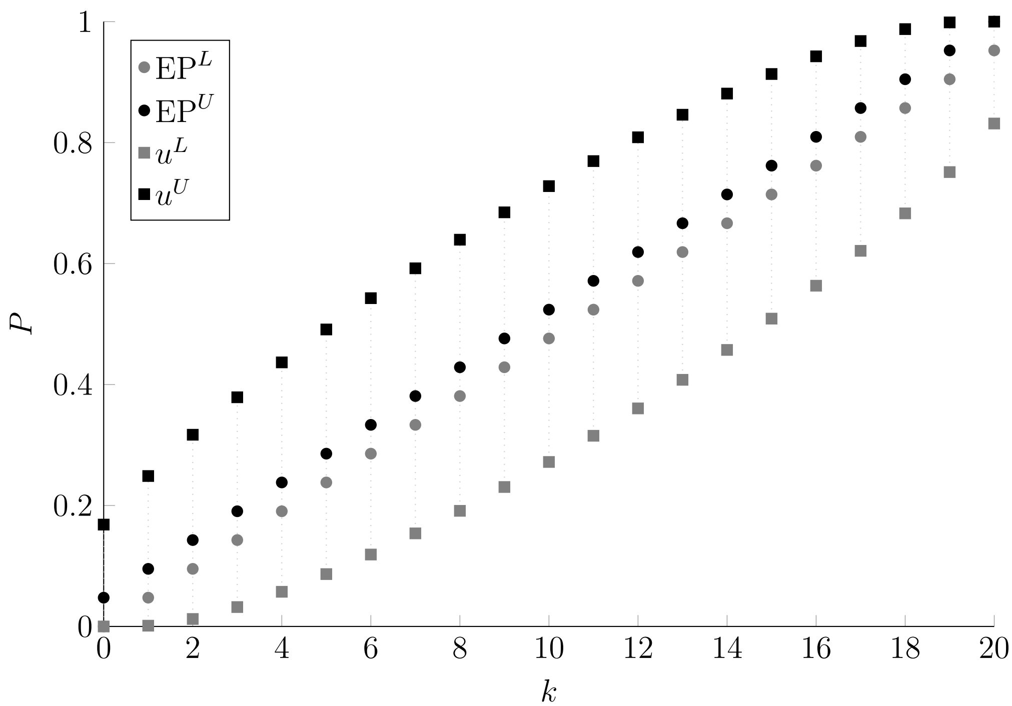


Figure 1: Expected values and confidence intervals of the component failure probability ($1-\alpha=0.95$, $n=20$, $k=0:20$)

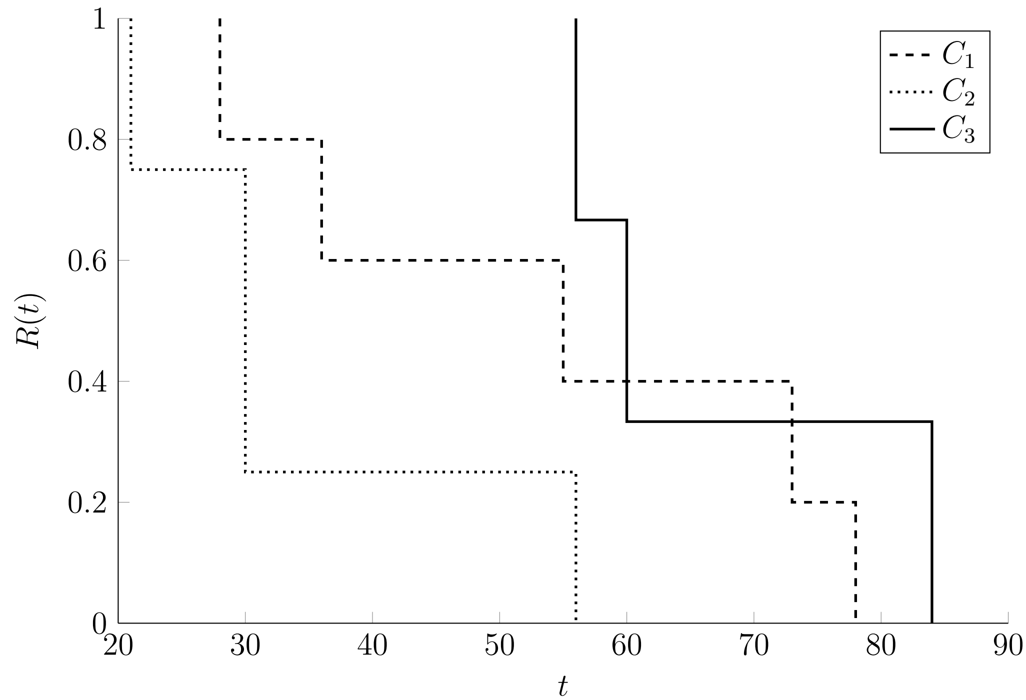
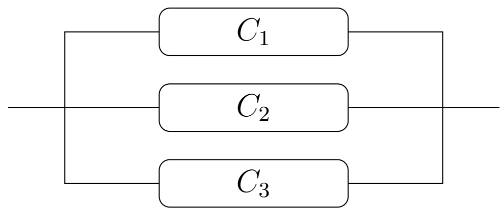
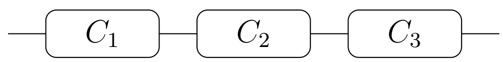


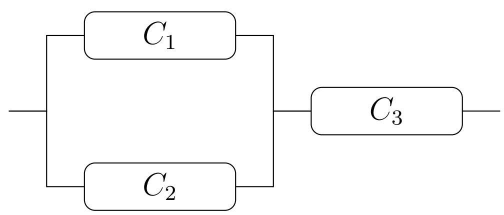
Figure 2: Parallel/serial systems: Empirical reliability for three system components: C_1 C_2 and C_3



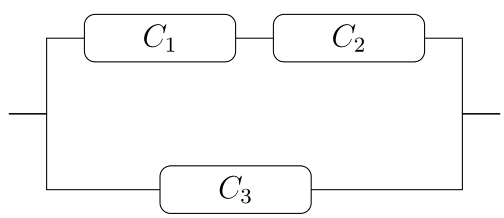
(a) Structure 1: Parallel system



(b) Structure 2: Serial system

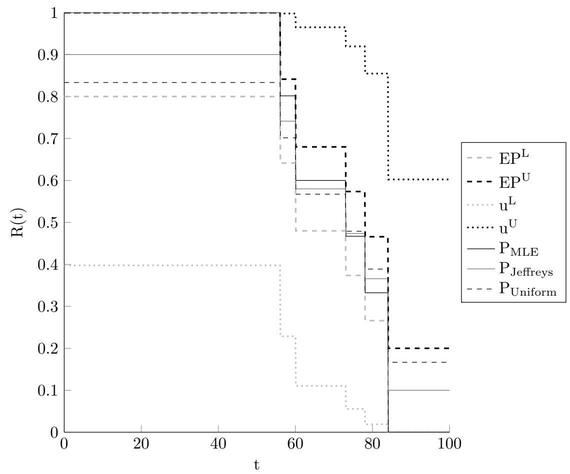


(c) Structure 3: First two component in parallel and the third in series with them

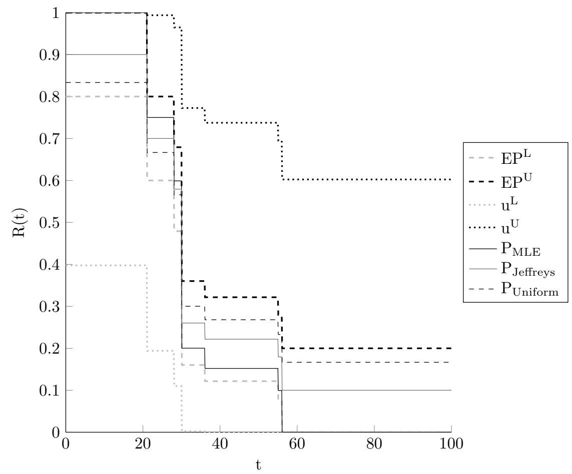


(d) Structure 4: First two component in series and the third in parallel with them

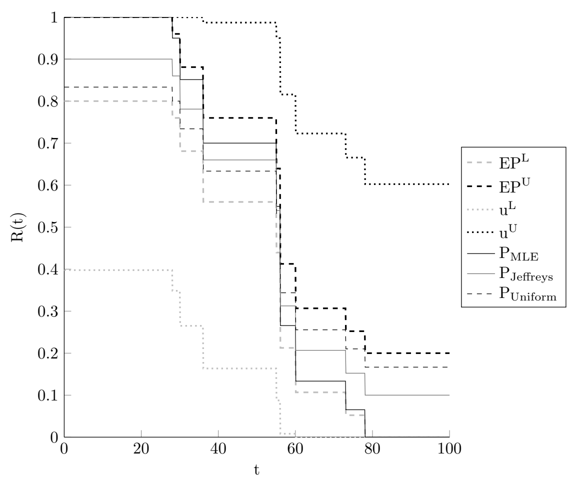
Figure 3: Parallel/serial systems: System structure of three-component system



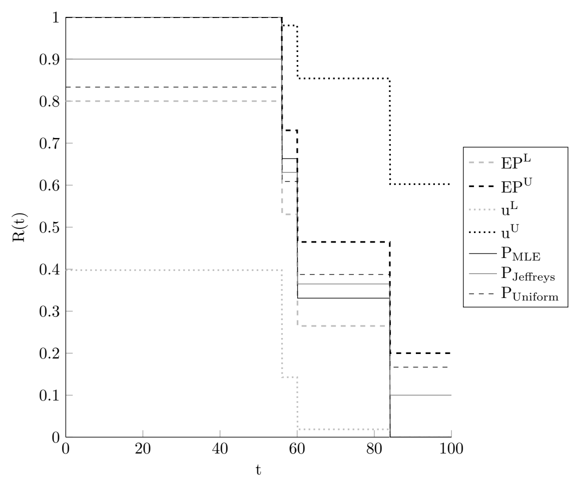
(a) Structure 1: Parallel system



(b) Structure 2: Serial system



(c) Structure 3: First two component in parallel and the third in series with them



(d) Structure 4: First two component in series and the third in parallel with them

Figure 4: Parallel/serial systems: System reliability $R_{s,t}$ with $n_s = 4$

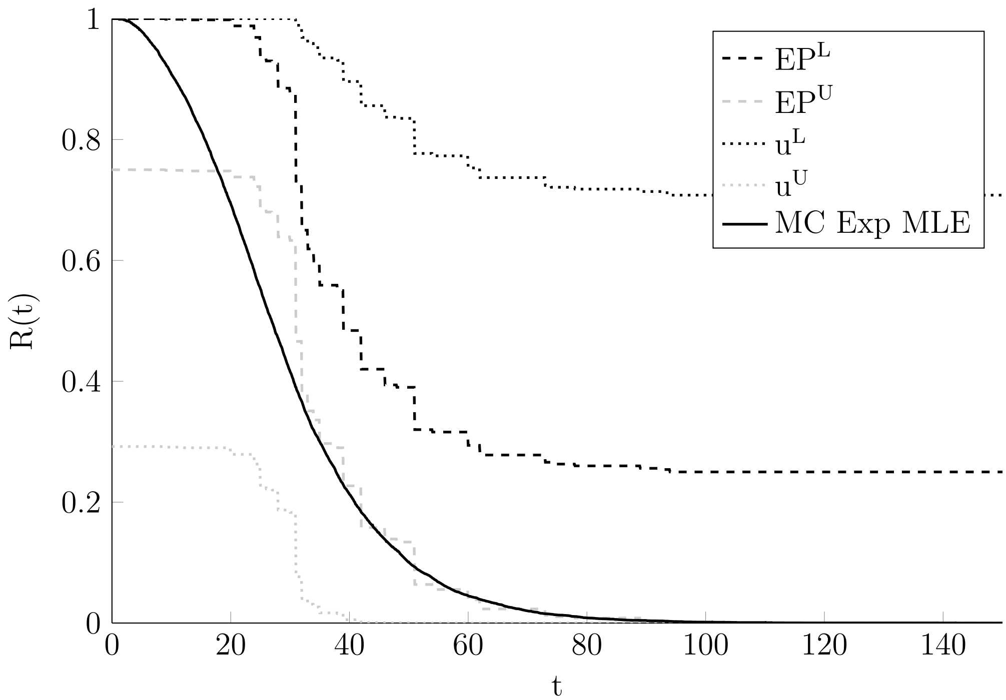


Figure 5: Fault tree BAOBAB1: System reliability $R_{s,t}$ with $N = 1000$ $n_s = 3$ $N_{MC} = 10000$

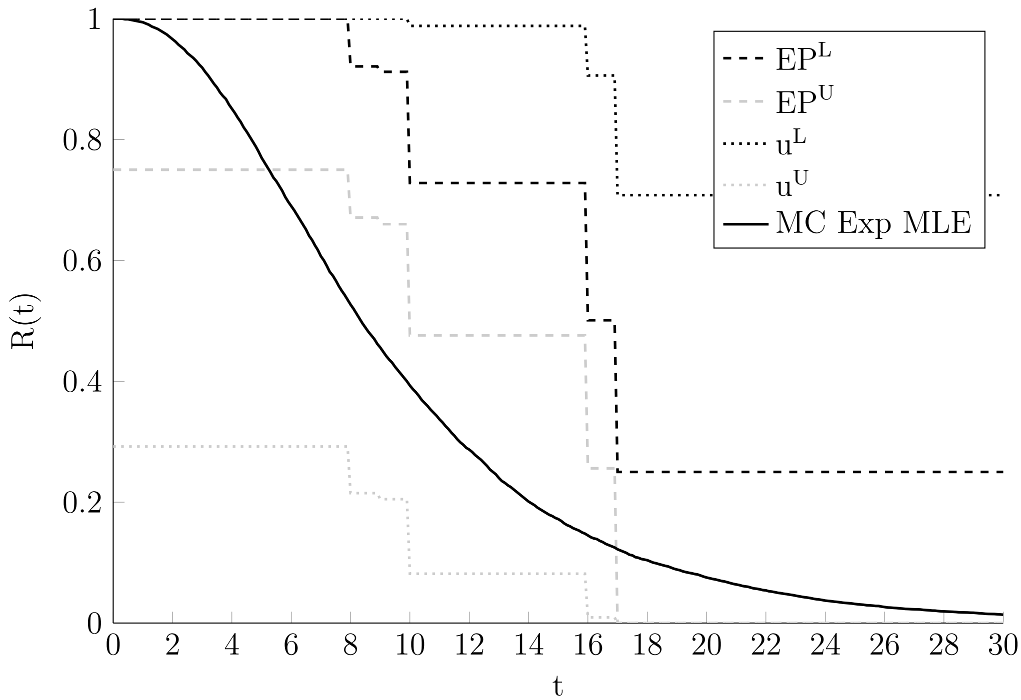


Figure 6: Fault tree BAOBAB2: System reliability $R_{s,t}$ with $N = 1000$ $n_s = 3$ $N_{MC} = 10000$