Master's thesis

NTNU
Norwegian University of Science and Technology

Ivar Bjerkebæk

The Cross Entropy Algorithm Applied to Monte Carlo Simulation of Power System Reliability

Master's thesis in Teknisk Fysikk
Supervisor: Jon Andreas Støvneng
Co-supervisor: Håkon Toftaker
February 2023

NTNU
Norwegian University of
Science and Technology

Ivar Bjerkebæk

The Cross Entropy Algorithm Applied to Monte Carlo Simulation of Power System Reliability

Masteroppgave i Teknisk Fysikk
Veileder: Jon Andreas Støvneng
Medveileder: Håkon Toftaker
Februar 2023

Norges teknisk-naturvitenskapelige universitet
Fakultet for naturvitenskap
Institutt for fysikk

NTNU
Kunnskap for en bedre verden

Abstract

Modern power systems are immensely complex, and the reliability of electricity supply is governed by rare interruption events which sometimes invoke catastrophic consequences. Proper reliability analysis is crucial both in the operating and planning phase of power systems, and the risk of interruptions must be evaluated with probabilistic models. Monte Carlo simulation is widely used in modern reliability analysis, but due to the rareness of interruptions, a naive Monte Carlo simulation is usually an inefficient approach. This thesis explores how importance sampling can increase the precision of different Monte Carlo models. The research culminates in a novel method that combines the principles of resampling, importance sampling and the cross entropy algorithm. The method is applicable to time-sequential simulations and requires very few model assumptions. When applied to a reliable grid configuration where interruptions occur in about $2 \cdot 10^{-5}$ of the samples, an average improvement in precision of 92.8% of the expected energy not supplied was observed.

Acknowledgements

I would like to give a special thanks to my supervisor Håkon Toftaker, and co-supervisor Iver Bakken Sperstad at SINTEF Energi. Håkon has provided excellent academic guidance, and has suggested new and creative approaches to the problem at points where my progression came to a halt. His engagement, positive attitude and ability to give constructive feedback have been key factors for my motivation throughout the semester. Iver has been an important resource in sharing expert knowledge on power system reliability. He has also provided clear and valuable advice for the process of writing an academic thesis.

Nomenclature

Abbreviations

MCS	Monte Carlo simulation
CMC	Crude Monte Carlo
CE	Cross entropy
NEF	Natural exponential family
PDF	Probability distribution function
ENS	Energy not supplied
EENS	Expected energy not supplied
LR	Likelihood ratio
$\mathrm{P}_{\mathrm{interr}}$	Interrupted power
SAC	System available capacity
IS	Importance sampling
ISD	Importance sampling distribution
SPM	Subspace partition method

Symbols

$f(\cdot)$	Reference PDF
$g(\cdot)$	Importance sampling density
$g^*(\cdot)$	Optimal importance sampling density
$\hat{f}(\cdot)$	Empirical distribution
$\hat{g}(\cdot)$	Importance resampling distribution
$W(\cdot)$	Likelihood ratio
$h(\cdot)$	Target function
$\mathbf{u}$	Reference parameters
$\mathbf{v}$	Importance sampling parameters
$\mathbf{X}^\dagger$	Dagger denotes a resampled random vector

Contents

1 Introduction **5**
- 1.1 Motivation . . . 5
- 1.2 State of the art . . . 5
- 1.3 Research objectives . . . 6
- 1.4 Structure of the document . . . 6

2 Background theory **8**
- 2.1 Power system reliability analysis . . . 8
- 2.2 Monte Carlo integration . . . 10
- 2.3 Variance reduction via importance sampling . . . 12
  - 2.3.1 Remark on the degeneracy of the likelihood ratio . . . 12
- 2.4 The cross entropy algorithm for rare event simulation . . . 14
- 2.5 The bootstrap . . . 17
- 2.6 Sequential Monte Carlo simulation . . . 18
  - 2.6.1 A condition dependent stochastic model for component failure . . . 19
- 2.7 Non-sequential Monte Carlo simulation . . . 20

3 Constructing a reference case for model verification **22**

4 Long-term reliability analysis **24**
- 4.1 Verification of the sequential Monte Carlo model . . . 24
- 4.2 Long term EENS prediction with the sequential model . . . 25

5 Non-sequential Monte Carlo Simulation and Importance Sampling **28**
- 5.1 Using a predefined importance sampling density . . . 28
- 5.2 Using the cross entropy algorithm . . . 29

6 Sequential Monte Carlo Simulation and Importance Sampling **32**
- 6.1 Two different likelihood ratios . . . 32
- 6.2 Using the steady-state likelihood . . . 33
- 6.3 Using the trajectory likelihood ratio with censoring . . . 34

7 Resampling **38**
- 7.1 Resampling from the empirical distribution . . . 38
- 7.2 Importance resampling . . . 39
  - 7.2.1 The cross entropy solution for a finite support discrete distribution . . . 39
  - 7.2.2 Estimating optimal probability weights for importance resampling ENS . . . 41
  - 7.2.3 The subspace partition method . . . 43
  - 7.2.4 The cross entropy solution for the subspace partition method . . . 44
  - 7.2.5 The subspace partition method for importance resampling ENS . . . 46
- 7.3 Application to the Foros-Istad model . . . 49
- 7.4 Remarks on extended application of the subspace partition method . . . 51

8 Concluding remarks and suggestions for further research **52**

References **53**

1 Introduction

1.1 Motivation

The work presented in this Master's thesis is part of the research project VulPro: Risk and vulnerability prognosis for power system development and asset management. The project is headed by SINTEF Energy Research in collaboration with Statnett, Landsnet and NVE, and is co-funded by the Norwegian Research Council. The general objective of VulPro is to develop new methods for long-term prediction of power system reliability, providing system planners with a better decision basis for asset mangagement and system development. One of the main topics of the project is to incorporate existing risk models for individual power system components into reliability of supply analysis of the power system. Such models should take into account the technical condition of components, how the condition is affected by aging and maintenance, and finally how these factors influence the probability of failure. When the complexity of the stochastic processes governing the system state increases, Monte Carlo simulation (MCS) remains a powerful and general tool for estimating reliability indices. The fundamental challenge in power system reliability evaluation is to analyse rare interruption events that have severe consequences, and capturing such rare events by Monte Carlo simulation can require an enormous number of samples. Therefore, development of successful variance reduction techniques which increase the accuracy of Monte Carlo estimates will be of great value.

1.2 State of the art

Power system reliability is evaluated using either analytical methods, Monte Carlo simulation, or a hybrid of the two [1]. In this context, there are two main approaches to Monte Carlo simulation: Sequential simulation and non-sequential simulation. In the sequential approach, the system state is simulated chronologically, while non-sequential models sample system states from a stationary distribution. As VulPro focuses on accounting for the technical condition of components, and how it evolves with time, sequential simulation is the main interest in this thesis. However, the stationary representation is simpler, and therefore a non-sequential model was useful in the first stage of the work to familiarize with established variance reduction techniques.

The main and most effective variance reduction techniques are importance sampling and conditional Monte Carlo, and both methods can dramatically improve the precision of Monte Carlo estimates. There is a variety of other variance reduction techniques, but these usually provide moderate improvements [20]. The scope of this thesis is limited to importance sampling. Importance sampling is performed by sampling from a probability distribution which is different from the true distribution of the system, and each sample is weighted by the likelihood ratio to obtain an unbiased estimator. There exists a theoretically optimal importance sampling distribution (ISD) which gives identically zero variance, and variance reduction is achieved when the ISD is close to the optimal distribution in some sense. The major challenge of importance sampling is to find an adequate ISD, and many different methods have been proposed to achieve this, see e.g [21, 20] for some examples. Two closely related importance sampling techniques called *variance minimization* and the *cross entropy method*, first described in [19] and [10], are especially relevant for power system reliability analysis. Both methods constrain the ISD to have the same functional form as the reference distribution, and aim to find the optimal distribution parameters. The variance minimization method was developed first, and the cross entropy (CE) algorithm can be regarded as an advancement of variance minimization that

usually results in more well structured optimization problems.

The cross entropy algorithm has especially been applied to non-sequential models with impressivr results, see e.g [22, 7]. Application to sequential models is less studied, but reference [6] should be mentioned as a good example which also entails an educational comparison to non-sequential methods. To the best of the author's knowledge, the CE method has only been applied to sequential simulations where the system state follows a Markov process. A Markov process is memoryless by definition, and is therefore unable to express the temporal evolution of component condition. The sequential MC model which is currently under development at SINTEF [26] takes into account aging and maintenance of transformers. These factors lead to time dependent component failure rates, and the system state follows a semi-Markov process. Application of the cross entropy algorithm to a general sequential MC simulation with time dependent failure rates will require invention of new methods.

1.3 Research objectives

Case study [26] reveals that SINTEF's Monte Carlo prototype suffers from serious issues regarding precision when applied to a moderately large system which is representative for a Nordic transmission grid [24]. Naturally, the precision can be improved by using a larger sample size, but since precision is proportional to the inverse square root of sample size, this is not an efficient approach. The author's ambition is to use importance sampling, preferably in combination with the cross entropy algorithm, to achieve variance reduction on SINTEF's Monte Carlo model. Regarding the time limitation and formal requirements of a master's thesis in applied physics at NTNU, the following research question and way points were defined:

Q: Can importance sampling reduce the variance of reliability index estimates for a sequential Monte Carlo simulation with time dependent component failure rates?

WP1: Implement a non-sequential MC model as a test bed to familiarize with importance sampling techniques.

WP2: Test different importance sampling techniques on a simplified version of the existing sequential model developed at SINTEF.

WP3: With the gained knowledge from WP1-2, develop a novel method to reduce the variance of the general sequential model with time dependent failure rates.

The scope of the thesis is further limited to only run simulations on a small 4-bus test network in order to maintain manageable run time during model development. The main focus is on the probabilistic part of reliability analysis, and thus existing software [14] is used without modification to determine the system's response to contingencies.

1.4 Structure of the document

The organization of the thesis is intended to clearly differentiate between the parts of the text which concern literature review or backround theory, and the author's own contributions. The Python code[1] for the project is based on an existing model from SINTEF which has been modified and extended. Details on the implementation are mostly avoided in the thesis. Chapter 2 describes backgroud theory covering the most fundamental concepts

[1]The current version of the code can be found on Bitbucket repository: `stash.code.sintef.no`, project folder SAMREL/vulpro, Branch: SAMREL 214.

of power system reliability analysis, Monte Carlo integration and importance sampling. Chapter 3 and 4 recapitulates some findings from the author's summer internship at SINTEF that are directly relevant to the thesis, these chapters describe the test system and illustrates long term reliability prediction with the Monte Carlo prototype. A more comprehensive report on the results from the summer internship can be found in reference [3]. The remainder of the results and discussion can be divided into 3 parts, each part concerns one of the way points defined above. Chapter 5 is is connected to WP1 where the cross entropy algorithm is applied to a non-sequential MC model. Chapter 6 covers WP2 and discusses importance sampling applied to a simplified sequential simulation where the system state follows a Markov process. Chapter 7 constitutes WP3 and explores the combined methods of resampling and importance sampling to achieve variance reduction on the general sequential model. And as the main result of the research, a novel method which was named the *subspace partition method* (SPM) is proposed. Finally, section 8 contains concluding remarks and suggestions for further research.

2 Background theory

2.1 Power system reliability analysis

The modern electrical power system is extremely complex, and its function is critical to society. Therefore, adequate models for assessing the reliability of power supply are necessary both in the planning and operation phase. The power systems ability of supplying energy to customers can be viewed from two different perspectives. Firstly, the power system must be engineered such that it can supply its customers under normal operation. In addition, the size and complexity of the system makes it necessary to analyse the ability to supply energy in situations outside the range of normal operation. The latter is the scope of power system reliability analysis and its main tool is probability theory. System states that are outside of normal operation can be caused by unexpected peaks in load, failures of grid components or events limiting the power generation, and the frequency and duration of such events can be described by stochastic models. A detailed reliability model for the entire power system would be impractical both in terms of interpretability and computational cost. Therefore the scope of analysis should be limited to specific parts of the power system based on their function. A useful hierarchical division is introduced in [1]. The classification consists of three levels and is sketched in figure 1. The ranking of the hierarchy is based on the direction of power flow. The generation facilities are on top, the transmission grid is in the middle and the distribution grid is on the bottom. The analysis in this paper concerns hierarchical level II: generation of power, and its distribution to delivery points in the transmission grid, in the literature this is often called composite system reliability analysis. The rest of this section is based on a report on the OPAL-methodology for reliability analysis developed at SINTEF [14].

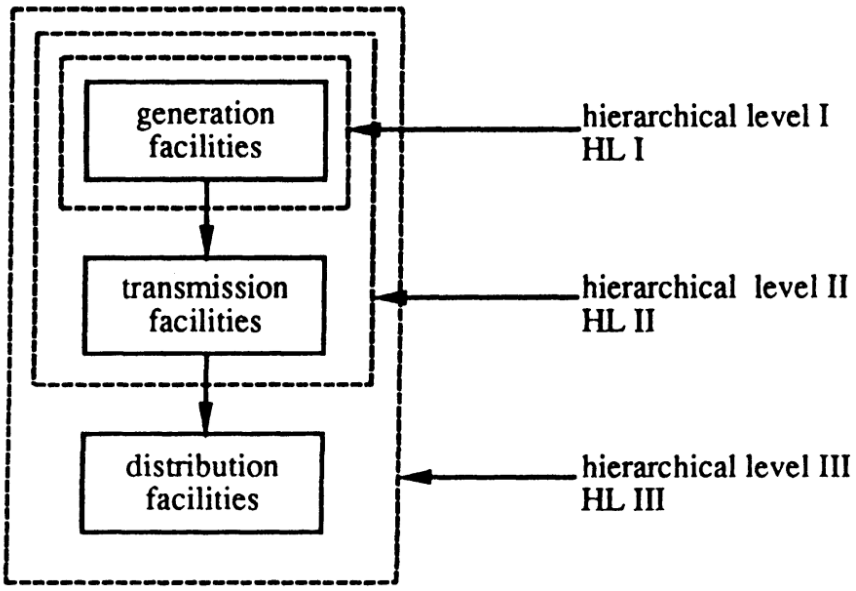


Figure 1: Hierarchical levels of power systems. Figure copied from [1].

The starting point for the analysis are data structures that specify the topology, specifications and constraints of all components in the system together with a set of *operating states*. An operating state can be defined in a simplified manner as a specific combination of load and generation in the system that has a physical solution and lies within the

security constraints. Given an operation state, one can analyse different *contingencies*, or *contingency states* and determine whether these lead to interruptions in power supply at delivery points. A contingency state is defined as a situation in which one or multiple components in the power system are in a failure state. Given an operating state and a contingency state one of four possible outcomes are possible:

1. The system is still within operational limits.
2. The contingency leaves the system in a state outside operational limits, but corrective measures are sufficient to bring the system back to normal operation without load shedding.
3. The contingency violates the operational limits of the system, and corrective actions are insufficient. Load shedding must be performed at one or multiple delivery points to bring the system back within operation limits.
4. Blackout. There is no stable physical solution to the optimal power flow problem.

The analysis used to determine the amount of load shedding and corrective measures rests on an optimal power flow algorithm, and is not within the scope of this thesis.

The contingency state can be represented as a time dependent random vector

$$\mathbf{X}(\mathbf{t}) = (X_1(t), \ldots, X_n(t)), \qquad X_i \in \{0, 1\}, \tag{2.1}$$

where each variable represents the state of a single component in the system and $X_i = 0$ means that the component is in operation, while $X_i = 1$ represents failure. In the most general case, the sojourn times a component spends in either state before the next transition are random variables with arbitrary distributions. E.g when the component enters state $i \in \{0, 1\}$, it remains there for a random amount of time $\tau_i \sim f_i(t)$ before the next transition. This stochastic process can be classified a semi-Markov process [18]. It is common and convenient to assume that the sojourn times are exponential random variables, where $\tau_0 \sim \text{Exp}(1/\lambda)$ and $\tau_1 \sim \text{Exp}(1/\mu)$, and λ and μ are the failure rate and repair rate. In this case, the component state follows a continuous time Markov Chain, also called a Markov process [18]. The Markov process is memoryless, and its stationary probabilities P_1 and P_0 are

$$P_1 = \frac{\mu^{-1}}{\mu^{-1} + \lambda^{-1}}, \tag{2.2}$$

$$P_0 = \frac{\lambda^{-1}}{\mu^{-1} + \lambda^{-1}}. \tag{2.3}$$

The state-diagram of the process is sketched in figure 2, and a component which undergoes this specific Markov process will for simplicity be referred to as a *Markovian component* in later chapters.

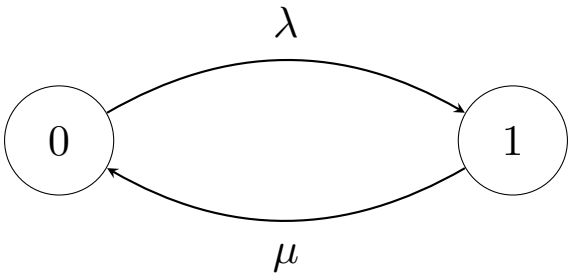


Figure 2: A two state continuous-time Markov chain for a power system component.

Based on the distribution of the contingency vector, one can obtain quantitative measures of the systems reliability in terms of statistical expectations, these are called *reliability indices*. The two basic reliability indices are the loss of load frequency ($LOLF$) and loss of load duration ($LOLD$) where loss of load means partial or total interruption of power supply. Other reliability indices can be obtained from these. This thesis focuses mainly on the expected energy not supplied ($EENS$), which is as the name suggests the expected value of energy not supplied (ENS). ENS is defined as the time integral of interrupted power (P_{interr}) over a given time period T

$$ENS = \int_0^T P_{\text{interr}}[t, \mathbf{X}(t)]dt, \tag{2.4}$$

where P_{interr} is the difference between the system's available capacity (SAC) and the load L. If there is excessive capacity, P_{interr} is defined to be 0

$$P_{\text{interr}} = \max\{0, L - SAC\}. \tag{2.5}$$

Reliability indices can be computed using analytical methods or estimated by Monte Carlo simulation. The OPAL method is an example of an analytical model which systematically analyses a specified list of contingency states by a so called *contingency enumeration* approach [2]. Analytical methods are typically fast, but require certain model assumptions. Monte Carlo simulation does not, in principle, put any restriction on the stochastic model, but can be computationally expensive.

2.2 Monte Carlo integration

Monte Carlo integration is a technique for approximating integrals by sampling random numbers. This section will focus on integrals in terms of expected values, but many problems which are not of probabilistic nature can be rewritten as expectation values, making Monte Carlo integration a very general tool. Let $\mathbf{X}$ be a random vector with probability density function (PDF) $f(\mathbf{x})$.

$$\mathbf{X} \sim f(\mathbf{x}). \tag{2.6}$$

Suppose one wants to compute the expected value θ of some *target function* $h(\mathbf{X})$

$$\theta \equiv \mathbb{E}[h(\mathbf{X})] = \int h(\mathbf{x}) \cdot f(\mathbf{x})\, d^n x. \tag{2.7}$$

θ can be approximated by sampling N independent random vectors from f and calculating the sample average

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^{N} h(\mathbf{X_i}). \tag{2.8}$$

This estimator has two important properties: It is unbiased, and its standard deviation scales as $1/\sqrt{N}$. These results follow directly from the basic properties of the expected value and variance operators. The first can be proved using the linearity of the expected value operator

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} h(\mathbf{X_i})\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[h(\mathbf{X_i})] = \theta. \tag{2.9}$$

The second property can be proved as follows

$$\mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N} h(\mathbf{X_i})\right] = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}[h(\mathbf{X_i})] = \frac{1}{N}\mathrm{Var}[h(\mathbf{X})]. \tag{2.10}$$

Additionally, it follows from the *central limit theorem* that the distribution of $\hat{\theta}$ tends towards the normal distribution for large N. A recapitulation of the central limit theorem without proof will follow.

Theorem. (The central limit theorem [18]). Suppose the random variables X_i, $i = 1, \ldots, N$ are independent and identically distributed with mean μ and variance σ^2. Construct a new random variable

$$Z = \frac{\sum_{i=1}^{N}(X_i - \mu)}{\sigma\sqrt{N}}. \tag{2.11}$$

The central limit theorem states that Z is standard normal distributed, $Z \sim n(0, 1)$ when N tends to infinity irrespective of the distribution of X_i.

It follows from the central limit theorem that for large N

$$\hat{\theta} \sim n(\theta, \sigma_h^2/N) \tag{2.12}$$

Where $\sigma_h^2 \equiv \mathrm{Var}[h(\mathbf{X})]$.

Monte Carlo integration can be useful in a number of different cases, and it is sometimes the only viable estimation method. Some general situations where Monte Carlo integration is convenient will be listed in the following, and for this discussion it will be useful to repeat the general form of the integral that is estimated

$$\theta = \int h(\mathbf{x}) \cdot f(\mathbf{x})\, d^n x.$$

Monte Carlo integration can be useful when:

1. The function $h(\mathbf{X})$ is complicated to evaluate making integration difficult.
2. The integral has high dimension. The volume of the integration domain scales exponentially with the dimension n, and the number of target function evaluations in other numerical integration methods will typically also scale exponentially. This is known as the curse of dimensionality. However, the uncertainty of the Monte Carlo estimate scales as $1/\sqrt{N}$ independent of dimension.
3. The probability distribution function $f(\mathbf{x})$ is complicated or unknown. In some situations the samples $\mathbf{X}_i$ are generated from a stochastic process that can be simulated, but it might be difficult to determine an analytical expression for the PDF.
4. $f(\mathbf{x})$ is known up to a constant factor, that is $f(\mathbf{x}) = c \cdot g(\mathbf{x})$ where c is unknown. In this case there exists methods for simulating a Markov chain that has f as its stationary distribution. These methods are known as Markov chain Monte Carlo methods, and the Metropolis-Hastings algorithm is the most famous one.

2.3 Variance reduction via importance sampling

The convergence rate of $1/\sqrt{N}$ is both the main strength and the main weakness of Monte Carlo integration. This convergence rate can be relied upon no matter how complicated the integrand or integration domain should be. On the other hand, a convergence rate of $1/\sqrt{N}$ can be painstakingly slow if the constant factor in the uncertainty is large. As an example, reducing the uncertainty by a factor of 10 requires 100 times the number of samples. Fortunately, there are methods for reducing the uncertainty in Monte Carlo estimates, these are called *variance reduction techniques*. This section will describe an often used variance reduction technique called *importance sampling* (IS) [18, 20].

Recall the general problem of Monte Carlo integration as described in the previous section which is to estimate the expected value

$$\theta = \mathbb{E}[h(\mathbf{X})] = \int h(\mathbf{x}) \cdot f(\mathbf{x})\, d^n x.$$

The estimator $\hat{\theta}$ proposed in equation (2.8) will from now on be referred to as the *crude* Monte Carlo estimate

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^{N} h(\mathbf{X_i}).$$

When using importance sampling, samples $\mathbf{X}_i$ are drawn from a new distribution g different from f, $\mathbf{X}_i \sim g$. And each sample is weighted to obtain an unbiased estimate of θ. In order to illustrate how to obtain the correct weights it is instructive to rewrite (2.7) as

$$\theta = \int h(\mathbf{x}) \cdot \frac{f(\mathbf{x})}{g(\mathbf{x})} \cdot g(\mathbf{x})\, d^n x = \mathbb{E}_g \left[h(\mathbf{X}) \cdot \frac{f(\mathbf{X})}{g(\mathbf{X})} \right], \tag{2.13}$$

where $\mathbb{E}_g$ denotes the expected value under g. Note that the support of $f(\mathbf{x}) \cdot h(\mathbf{x})$ must be contained in the support of $g(\mathbf{x})$ for equation (2.13) to hold. Now one can construct a new unbiased estimator of θ

$$\tilde{\theta} = \frac{1}{N} \sum_{i=1}^{N} h(\mathbf{X_i}) \cdot \frac{f(\mathbf{X_i})}{g(\mathbf{X_i})}, \tag{2.14}$$

where the samples are weighted by the likelihood ratio $W(\mathbf{X}) \equiv \frac{f(\mathbf{X})}{g(\mathbf{X})}$. For a good choice of importance sampling distribution (ISD), the variance of $\tilde{\theta}$ can be much lower than that of the crude estimate.

2.3.1 Remark on the degeneracy of the likelihood ratio

The main challenge of importance sampling is to find a suitable ISD that reduces the variance compared to the crude estimator, and the variance of the importance sampling estimator can be very sensitive to the choice of ISD. This is connected to the fact that the distribution of W under g easily becomes very skewed, and this is especially true if the dimension of $\mathbf{X}$ is high. Following reference [20], the problem of extreme skewness of the likelihood ratio and high variance of the IS estimator will be refered to as *degeneracy*. This may confuse some readers since the term degeneracy in mathematics usually refers to the case where an object reduces to a simpler form, but in this context degeneration is synonymous to corruption or collapse. This subsection will first describe some characteristics of a successful ISD, and finally the degeneration of the likelihood ratio will be illustrated through a toy example.

The variance of the IS estimator is

$$\mathrm{Var}[\tilde{\theta}] = \frac{1}{N}\left(\mathbb{E}_g\left[h^2(\mathbf{X})\cdot\frac{f^2(\mathbf{X})}{g^2(\mathbf{X})}\right] - \theta^2\right). \tag{2.15}$$

In fact, there exists an optimal ISD g^* that gives zero variance of the IS estimator. The optimal ISD is given by

$$g^*(\mathbf{x}) = \frac{h(\mathbf{x})\cdot f(\mathbf{x})}{\theta}, \tag{2.16}$$

which is easily verified by inserting the expression for g^* into equation (2.15). Successful importance sampling is achieved when the ISD is similar, or close to the optimal distribution in some sense. However, sampling directly from g^* is problematic since this requires that θ is known in advance, and calculating θ is the very heart of the problem. In addition to being close to the optimal distribution, some other favourable characteristics of a suitable ISD can be identified by examining equation (2.15). Naturally, the variance of the IS estimator should be finite and this is equivalent to the requirement that

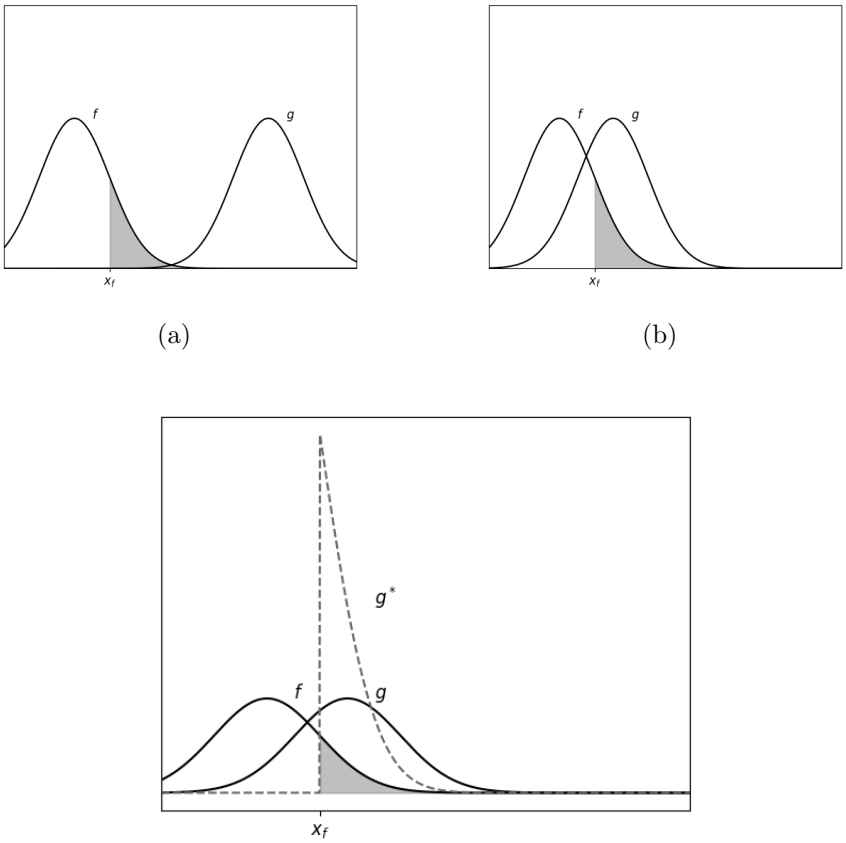
$$\mathbb{E}_g\left[h^2(\mathbf{X})\cdot\frac{f^2(\mathbf{X})}{g^2(\mathbf{X})}\right] < \infty. \tag{2.17}$$

This can be achieved by asserting that the likelihood ratio $f(\mathbf{X})/g(\mathbf{X})$ is bounded. Additionally g should not have a lighter tail than f in the region where h has support. Having a lighter tail means that the likelihood of g goes towards zero faster than f when moving further away from the bulk of probability mass.

Example. (Degeneration of the likelihood ratio.) Consider a system which can be described by a single random variable X which has a Gaussian distribution with mean μ and variance σ^2. The system is in a failure state when X is bigger than some value x_f, and is functioning otherwise. Suppose one wants to estimate the probability of failure $P_f \equiv P\{X > x_f\} = \mathbb{E}[I\{X > x_f\}]$ by importance sampling. The importance sampling estimator is

$$\tilde{P}_f = \frac{1}{N}\sum_{i=1}^{N} I\{X > x_f\}\cdot\frac{f(X)}{g(X)}.$$

Figure 3 illustrates different choices of ISD for the above problem. Plot 3a shows an example of an importance sampling density that would lead to degeneracy. The bulk of the probability mass lies in the failure domain $x > x_f$, thus most samples will represent failure states. If P_f is very low, the failure states are the "important samples", and therefore one might be tempted to think that the ISD in figure 3a would lead to successful importance sampling. However, the distribution of W under g is extremely skewed. By visual inspection on can see that $f/g << 1$ under the bulk of g, so the observed likelihood ratios will usually be very close to zero. At the same time $\mathbb{E}_g[W(\mathbf{X})] = 1$ by definition, therefore the likelihood ratio must take on extremely large values with a low probability. This is degeneracy in a nutshell, and for high dimensional problems the skewness of W is amplified. Figure 3b shows a good choice of ISD. The bulk of the probability mass is in the "important" failure region and the distribution of the likelihood ratio is only moderately skewed. Figure 3 shows the same ISD as in (b) together with the theoretically optimal ISD. Note that g and g^* are somewhat similar in the sense that they have most of their mass centered in the same region. For this simple example, the variance of the IS estimator was found by numerical integration of equation (2.15), and the ISD in figure 3a gives more than 2000 times higher variance than the ISD in (b).



(a) (b)

(c)

Figure 3: (a): An ISD that leads to degeneracy of the likelihood ratio. (b): A good ISD. (c): A good ISD and the optimal ISD.

2.4 The cross entropy algorithm for rare event simulation

As discussed in the previous section, successful importance sampling is achieved when the ISD is close to the theoretically optimal ISD, g^*. The *cross entropy algorithm* [10, 4, 20] presents an automatic method to find an ISD that is as close as possible to the optimal ISD under certain constraints, and it is especially effective for *rare event simulation*. The rationale behind the algorithm, and the main results will be presented in this section.

First the specific meaning of rare event simulation should be clarified. Suppose one wants to estimate the probability

$$l \equiv P\{h(\mathbf{X}) \geq \gamma\}. \tag{2.18}$$

If l is very small, say 10^{-5} or less, then rare events are defined as the set

$$\{\mathbf{X}|h(\mathbf{x}) \geq \gamma\}. \tag{2.19}$$

(2.18) can also be written as an expectation value

$$P\{h(\mathbf{X}) \geq \gamma\} = \mathbb{E}[I\{h(\mathbf{X}) \geq \gamma\}], \tag{2.20}$$

where I is the indicator function. Now l can be estimated by Monte Carlo integration,

and the crude estimate is

$$\hat{l} = \frac{1}{N}\sum_{i=1}^{N} I\{h(\mathbf{X}) \geq \gamma\}. \tag{2.21}$$

However, since $\{\mathbf{X}|h(\mathbf{x}) \geq \gamma\}$ are rare events, a very large number of samples is needed to get a sufficient number of non-zero contributions to the sum in (2.21). This is the main challenge of rare event simulation. Now we want to use importance sampling to overcome this challenge. Using equation (2.16), the optimal ISD is

$$g^*(\mathbf{x}) = \frac{I\{h(\mathbf{x}) \geq \gamma\} \cdot f(\mathbf{x})}{l}. \tag{2.22}$$

The cross entropy algorithm searches for a function g that is close to g^*, and the search is limited to functions that have the same functional form as f, but the distribution parameters are allowed to vary. That is: if f is parameterized as $f = f(\mathbf{x}; \mathbf{u})$ where $\mathbf{u} = (u_1, \ldots, u_n)$, then

$$g \in \{f(\mathbf{x}; \mathbf{v}), v_i \in \mathbb{R}\}. \tag{2.23}$$

Now one needs a measure of how "close" $f(\mathbf{x}; \mathbf{v})$ is to g^*, and the *Kullback-Liebler distance*, also known as *cross entropy* is used for this purpose. The Kullback-Liebler distance between two distributions f and g, is defined as

$$D(g, f) \equiv \mathbb{E}_g\left[\ln \frac{g(\mathbf{X})}{f(\mathbf{X})}\right] = \int g(\mathbf{x}) \cdot \ln \frac{g(\mathbf{x})}{f(\mathbf{x})} d^n x. \tag{2.24}$$

The cross entropy algorithm aims to find the paramerters that minimize the Kullback-Liebler distance between the importance sampling distribution and g^*,

$$\mathbf{v} = \arg\min_{\mathbf{v}} \; D\left[g^*(\mathbf{x}), f(\mathbf{x}; \mathbf{v})\right]. \tag{2.25}$$

By writing out the expression for the Kullback-Liebler distance, the optimization problem in equation (2.25) can be reduced to a simpler form.

$$\begin{aligned} D(g^*, f) &= \mathbb{E}_{g^*}\left[\ln(g^*/f)\right] \\ &= \int g^*(\mathbf{x}) \ln g^*(\mathbf{x}) d^n x - \int g^*(\mathbf{x}) \ln f(\mathbf{x}; \mathbf{v})\, d^n x \end{aligned} \tag{2.26}$$

The first integral in (2.26) is independent of $\mathbf{v}$, therefore (2.25) is equivalent to maximizing the second integral term

$$\begin{aligned} \int g^*(\mathbf{x}) \ln f(\mathbf{x}; \mathbf{v})\, d^n x &= \int l^{-1} I\{h(\mathbf{x}) \geq \gamma\} f(\mathbf{x}; \mathbf{u}) \ln f(\mathbf{x}; \mathbf{v})\, d^n x \\ &= l^{-1}\mathbb{E}_{\mathbf{u}}\left[I\{h(\mathbf{X}) \geq \gamma\} \ln f(\mathbf{X}; \mathbf{v})\right] \\ &= l^{-1}\mathbb{E}_{\mathbf{w}}\left[I\{h(\mathbf{X}) \geq \gamma\} W(\mathbf{X}; \mathbf{u}, \mathbf{w}) \ln f(\mathbf{X}; \mathbf{v})\right], \end{aligned} \tag{2.27}$$

the last equality was obtained by a change of measure, where $\mathbf{w}$ is arbitrary, $\mathbf{u}$ are the reference parameters and $W(\mathbf{X}; \mathbf{u}, \mathbf{w}) \equiv \frac{f(\mathbf{X};\mathbf{u})}{f(\mathbf{X};\mathbf{w})}$. Now equation (2.25) can be rewritten as

$$\mathbf{v} = \arg\max_{\mathbf{v}} \; \mathbb{E}_{\mathbf{w}}\left[I\{h(\mathbf{X}) \geq \gamma\} W(\mathbf{X}; \mathbf{u}, \mathbf{w}) \ln f(\mathbf{X}; \mathbf{v})\right]. \tag{2.28}$$

If the problem is convex in $\mathbf{v}$, it can be solved by setting the gradient with respect to $\mathbf{v}$ equal to zero, which reduces to

$$\begin{gathered} E_{\mathbf{w}}\left[I\{h(\mathbf{X}) \geq \gamma\} W(\mathbf{X}; \mathbf{u}, \mathbf{w})\, \partial_k \ln f(\mathbf{X}; \mathbf{v})\right] = 0, \\ k = 1, \ldots, n. \end{gathered} \tag{2.29}$$

The set of equations (2.29) yields the deterministic solution to the cross entropy problem, and is usually not computable. However, the deterministic solution can be estimated by replacing (2.29) by its *stochastic counterpart*

$$\frac{1}{N}\sum_{j=1}^{N} I\{h(\mathbf{X}_j) \geq \gamma\} W(\mathbf{X}_j; \mathbf{u}, \mathbf{w})\, \partial_k \ln f(\mathbf{X}_j; \mathbf{v}) = 0. \tag{2.30}$$

The solution to this set of equations will depend on the form of the PDF, and an especially important case with broad application is when f belongs to the nautural exponential family (NEF). In this case the set of equations are uncoupled, and the solution reads

$$v_k = \frac{\sum_{j=1}^{N} I\{h(\mathbf{X}_j) \geq \gamma\} W(\mathbf{X}_j; \mathbf{u}, \mathbf{w})\, X_{jk}}{\sum_{i=1}^{N} I\{h(\mathbf{X}_j) \geq \gamma\} W(\mathbf{X}_j; \mathbf{u}, \mathbf{w})}. \tag{2.31}$$

For simplicity f is assumed to belong to the NEF in the remainder of this section. If $\mathbf{X}_i$ are sampled from the reference distribution, the stochastic counterpart (2.31) will suffer from low precision since $P[I\{h(\mathbf{X}_j) \geq \gamma\}] << 1$ leaving very few non-zero terms in the sum. To overcome this challenge, the cross entropy algorithm uses an iterative approach where both γ and v_k are adjusted iteratively. The main algorithm consists of the 5 following steps [4]:

Step 1 Set $\mathbf{v}^{(0)} \leftarrow \mathbf{u}$, where $\mathbf{u}$ are the reference parameters. Define a value for γ.

Step 2 Draw N samples $\{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$ from $f(\mathbf{X}; \mathbf{v}^{(t-1)})$.

Step 3 Sort the values of $h(\mathbf{X_i})$ such that $h^{(1)} \leq \ldots \leq h^{(N)}$ and set γ_t to be the $(1-\rho)$ quantile of $\{h^{(i)}\}$ where ρ is not too small, for example $\rho = 0.01$. If $\gamma_t > \gamma$ set $\gamma_t \leftarrow \gamma$.

Step 4 For all parameters v_k, $k = 1, \ldots, n$ set

$$v_k^{(t)} \leftarrow \frac{\sum_{i=1}^{N} I(h(\mathbf{X}_i) \geq \gamma_t) \cdot W(\mathbf{X}_i; \mathbf{u}, \mathbf{v}^{(t-1)}) \cdot X_{ik}}{\sum_{i=1}^{N} I(h(\mathbf{X}_i) \geq \gamma_t) \cdot W(\mathbf{X_i}; \mathbf{u}, \mathbf{v}^{(t-1)})}. \tag{2.32}$$

Step 5 If $\gamma_t < \gamma$, repeat step 2-5.

When the convergence criterion is fulfilled, namely $\gamma_t = \gamma$, use the final parameter vector $\mathbf{v}^{(t)}$ for importance sampling as described in section 2.3.

Note that the toy example in 2.3.1 concerns estimation of a rare event probability. In fact, figure 3 was not only constructed to illustrate degeneracy, but also to visualize the cross entropy method. The mean μ_g of the gaussian distribution g in figure 3c was centered at the point that minimizes the cross entropy between g and g^*. The parameter value was found by numerical integration of equation (2.29). Thus μ_g is the solution to the deterministic CE problem when μ is the only free parameter. In fact, the mean of g and g^* were seen to coincide.

Finally, an important remark on the area of use for the CE algorithm must be made. In the above discussion, the CE algorithm was presented as a tool to estimate a rare event probability limited to the form $l = \mathbb{E}[I\{h(\mathbf{X}) \geq \gamma\}]$. The CE algorithm is however often used to estimate the expected value of h instead, i.e $\theta = \mathbb{E}[h(\mathbf{X})]$. This works well if the region of sample space where $h(\mathbf{X}) \geq \gamma$ constitutes the important samples for estimating θ.

2.5 The bootstrap

Since Monte Carlo estimates are stochastic quantities, it is crucial to obtain a measure of uncertainty. A straight forward way of doing so is to run multiple independent simulations and calculate e.g the standard deviation from the independent estimates, but this is obviously inefficient. *The bootstrap* [5] offers an automatic procedure of assessing both the uncertainty and distribution of the Monte Carlo estimator based on the samples from a single simulation. The general principle is to use a so called *bootstrap distribution* as a representation for the true distribution of the estimator.

Suppose that one uses a sample of N random vectors $\mathbb{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$ to estimate the expected value of the target function

$$\hat{\theta} = \hat{\theta}(\mathbb{X}) = \frac{1}{N}\sum_{j=1}^{N} h(\mathbf{X}_j). \tag{2.33}$$

A bootstrap sample $\mathbb{X}^\dagger$ is obtained by resampling observations from the sample N times *with replacement*.

$$\mathbb{X}^\dagger = \{\mathbf{X}_1^\dagger, \ldots, \mathbf{X}_N^\dagger\ \}. \tag{2.34}$$

Thus the bootstrap sample is a random permutation of the original sample where each observation may appear several times. The bootstrap distribution f_B is the distribution of the estimator $\hat{\theta}(\mathbb{X}^\dagger)$

$$\hat{\theta}(\mathbb{X}^\dagger) \sim f_B. \tag{2.35}$$

If N is sufficiently large, the bootstrap distribution is a good representation of the estimator's true distribution

$$\hat{\theta}(\mathbb{X}) \sim f. \tag{2.36}$$

A theoretical explanation of why this principle works is outside the scope of this thesis, the reader is directed to [5] for a comprehensive and educational text book on the bootstrap. In the following, the bootstrap principle will be illustrated through a specific example.

Example. (Bootstrap standard error.) Assume one wants to evaluate the standard error of the estimator $\hat{\theta}(\mathbb{X})$ in equation (2.33). This is done as follows:

1. Draw B bootstrap samples $\mathbb{X}_1^\dagger, \ldots, \mathbb{X}_B^\dagger$.
2. Compute the bootstrap standard error as

$$\sigma_B = \left(\frac{1}{B-1}\sum_{b=1}^{B}\left[\hat{\theta}(\mathbb{X}_b^\dagger) - \theta_B\right]^2\right)^{1/2}, \tag{2.37}$$

   where θ_B is the bootstrap mean

$$\theta_B = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}(\mathbb{X}_b^\dagger)\ . \tag{2.38}$$

Other statistics such as confidence intervals, quantiles and even bias can be evaluated in the same manner. And what makes the bootstrap an extremely useful tool is that no knowledge of the distribution of $\mathbf{X}$ is required.

2.6 Sequential Monte Carlo simulation

There are two main simulation approaches for reliability evaluation of power systems, sequential and non-sequential Monte Carlo simulation. The sequential MCS simulates the stochastic processes in the system chronologically, and each MC sample describes the history of the system state throughout the simulation period. Non-sequential MCS assumes that the system follows a continuous-time Markov chain, and system states are sampled from the stationary distribution. This section will give a general description of sequential MCS as well as a condition dependent stochastic model for transformers developed at SINTEF.

In the sequential MCS, the evolution of each component state $X_i(t)$ is simulated chronologically. The component states are assumed independent throughout this thesis, thus each component can be simulated individually. The component states are combined to a system state time series $\mathbf{X}(t) = [X_1(t), \ldots, X_n(t)]$, and the system time series can be used to find the system available capacity and interrupted power as a function of time, this is illustrated in figure 4. Reliability indices can be estimated by Monte Carlo integration when a sufficient number of system state time series have been simulated. The (crude) $EENS$ estimator is defined as follows

$$\widehat{EENS} = \frac{1}{N} \sum_{i=1}^{N} \int_0^T P_{\text{interr}}[t, \mathbf{X}_i(t)] dt. \tag{2.39}$$

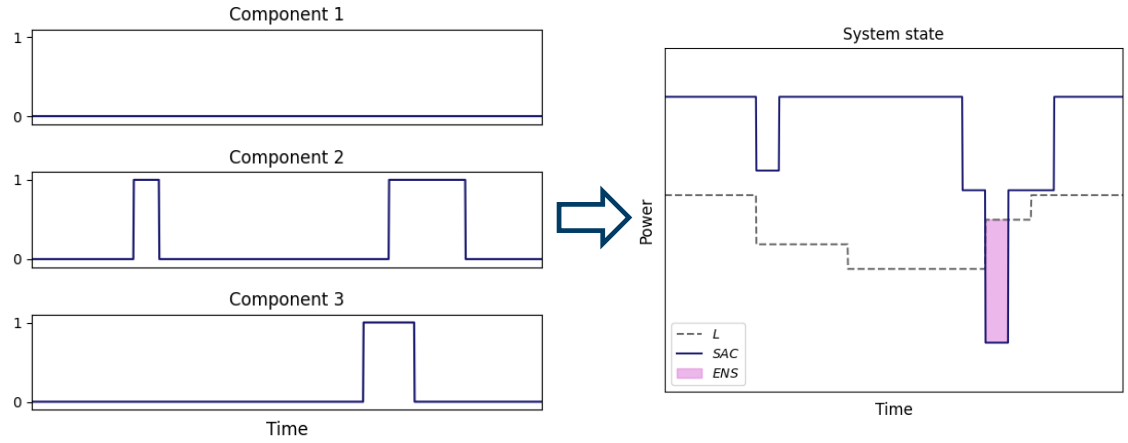


Figure 4: The sequential MCS combines component state time series to a time series of system available capacity. The area of the shaded regions where the load is higher than SAC is ENS.

The main advantage of the sequential MCS is that it does not put any restrictions on the type of stochastic process that govern component states. The two state continuous-time Markov chain described in section 2.1 is a useful and simple stochastic model for component states. However, this model assumes a constant failure rate and is unable to

describe processes that lead to time dependencies such as the effect of component aging and maintenance schemes. The following subsection will describe a condition dependent stochastic model for the failure of transformers which has been developed at SINTEF. A large part of the work that will be presented in this paper uses this model in sequential MCS.

2.6.1 A condition dependent stochastic model for component failure

The condition dependent probabilistic model for transformer failure [8] has been developed by Jørn Foros and Maren Istad at SINTEF Energy Research, and will be reffered to as the Foros-Istad model from here on. The model is further adapted and integrated in reliability of supply analysis in [25, 26]. The Foros-Istad model separates failures in the active and non-active part of the transformer and both failure modes will put the trasformer out of operation. The stochastic model for the active part is the only condition dependent and explicitly time dependent failure mode.

Failures in the active part of the transformer, i.e core, windings and oil are labeled *wear-out failures*. A Norwegian database for scrapped transformers is used to curve fit a cumulative distribution for the time to failure. The cumulative distribution $F_w(s)$ is a function of *apparent age* s, and not chronological time t. Apparent age is a direct measure of technical condition, and is introduced to reflect the fact that the lifetime of a transformer should depend on condition and not directly on age. As an example, an old transformer can have a low apparent age after maintenance. Apparent age is in turn related to a health index (HI), which is a number between 0 and 1, where 1 is the best technical condition

$$\begin{aligned} HI &= HI(s), \\ s &= HI^{-1}(hi). \end{aligned} \tag{2.40}$$

Time to failure in the non-active part is assumed to be exponentially distributed with rate λ_{ml}, where the subscript "ml" is short for mid-life failure. After a failure of any kind, the sojourn time before the transformer is repaired or replaced is modeled as an exponential random variable with rate μ. In addition to failure and repair, the Foros-Istad model incorporates a simple maintenance scheme. When the transformer is in operation, preventive maintenance is conducted at random times which are exponentially distributed with rate λ_{pm}. The effect of preventive maintenance is to update the health index to $HI = 1$ giving an apparent age $s = HI^{-1}(1) = 0$.

The wear-out failure mode introduces time dependence in the failure rate. Although it is the time to failure PDF and not the failure rate which is directly involved in the simulation, it is instructive to examine how the failure rate depends on the failure modes. The two failure modes are independent, and thus the total failure rate λ_{tot} is the sum of the wear-out failure rate λ_w and the mid-life failure rate

$$\lambda_{tot} = \lambda_w + \lambda_{ml}\,, \tag{2.41}$$

this is known as a competing risk model. λ_w is a function of apparent age and is given by the definition of the failure rate function [18]

$$\lambda_w(s) = \frac{f_w(s)}{1 - F_w(s)}\,, \tag{2.42}$$

where $f \equiv \frac{\mathrm{d}F}{\mathrm{d}s}$. The wear-out failure rate as a function of chronological time is

$$\lambda_w(t) = \lambda_w(s) \cdot \frac{ds}{dt} = \frac{f_w(s)}{1 - F_w(s)} \cdot \frac{ds}{dt}\,. \tag{2.43}$$

The full stochastic process including failure, repair and maintenance is illustrated in figure 5.

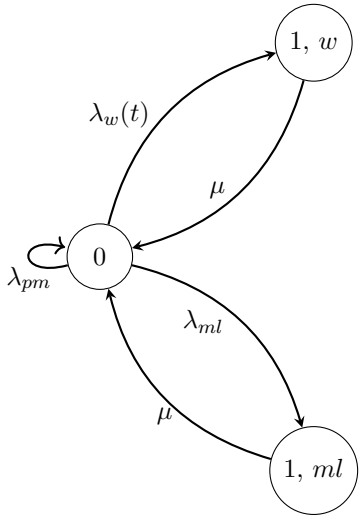


Figure 5: State diagram describing the probabilistic failure model. Figure courtesy of Håkon Toftaker.

Sequential simulation of this stochastic process results in a sequence of alternating times-to-failure τ_0 and repair times τ_1

$$\mathbb{T} = \{\tau_0^{(1)}, \tau_1^{(1)}, \tau_0^{(2)}, \tau_1^{(2)}, \ldots\}. \tag{2.44}$$

This sequence specifies the evolution of the component state $X(t)$. The τ_0 marks the points in time when the component makes a transition from up-state to down state $0 \to 1$, and the τ_1 marks transitions from down-state to up-state. Naturally, the simulation period must be finite, this leads to a partial censoring of the last transition time in (2.44). The censoring effect will have an important impact on the importance sampling discussed in section 6.3.

The list at the end of section 2.2 about the applications of Monte Carlo integration can underline why it is challenging to calculate reliability indices by analytical methods when the component states follow stochastic processes like the one described by the Foros-Istad model. In fact the problem exhibits all of the first three features that were listed. 1) The target function $h(\mathbf{X})$ is in this case ENS, and evaluating ENS requires an optimal power flow analysis of the whole power system. 2) The number of random variables involved is unknown since the number of transitions in (2.44) within the simulation period will vary. Thus the dimension of the integration domain can in principle be infinite. 3) The samples $\mathbf{X}(t)$ are generated from an ivolved stochastic process and are not sampled from a known PDF.

2.7 Non-sequential Monte Carlo simulation

The non-sequential Monte Carlo simulation of the power system assumes that the component states follows a continuous-time Markov Chain as described in section 2.1. The

stationary probabilies for a single component are

$$P_1 = P\{X = 1\} = \frac{\mu^{-1}}{\lambda^{-1} + \mu^{-1}}, \tag{2.45}$$

$$P_0 = P\{X = 0\} = \frac{\lambda^{-1}}{\lambda^{-1} + \mu^{-1}}. \tag{2.46}$$

P_1 is known as the *unavailability* and is conventially denoted u. Using this notation, the stationary distribution for each component can be written as

$$P(X = x) = u^x(1 - u)^{1-x}, \tag{2.47}$$

this is a Bernoulli distribution with parameter u. Since component states are independent, the joint distribution of the system is simply the product of the component distributions

$$f(\mathbf{x}) = \prod_{i=1}^{n} u_i^{x_i}(1 - u_i)^{1-x_i}. \tag{2.48}$$

In the non-sequential approach, system states $\mathbf{X}$ are sampled from f, and the crude estimate of $EENS$ is

$$\widehat{EENS} = \frac{1}{N}\sum_{i=1}^{N}\int_0^T P_{\text{interr}}(t, \mathbf{X})dt = \frac{1}{N}\sum_{i=1}^{N} T \cdot \overline{P}_{\text{interr}}(\mathbf{X}), \tag{2.49}$$

where $\overline{P}_{\text{interr}}(\mathbf{X})$ is the time-average of interrupted power over all operating states.

3 Constructing a reference case for model verification

Since this work concerns development of new Monte Carlo techniques, it is practical to use a simple test system for interpretability of results and low run time in the testing phase. A simple transmission network was constructed which is shown in the line diagram in figure 6. The test system was constructed during the authors summer internship and this section is a recapitulated from [3]. The network consists of two generating units, two transmission lines, two transformers and two delivery points. The new test system is based on the OPAL test network [14], the only difference being that two of the transmission lines have been replaced by transformers which convert the voltage of 132kV supplied by the generators to 66kV at the delivery points. The transformers are introduced for use of the Foros-Istad transformer failure model in sequential MCS as described in section 2.6.1. The reliability analysis considers only failure in lines and transformers, which will sometimes be referred to by the common term *branch* in this thesis.

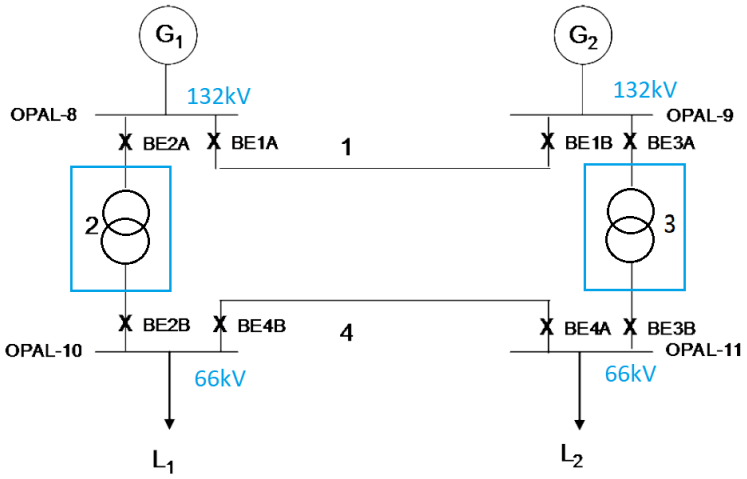


Figure 6: Line diagram of the test network.

The OPAL test network comes with a load curve which is also used for the new test network. The load curve consists of 2 operational states, a state of high load through the winter months December-February and a state of low demand the rest of the year. The load on delivery point L_1 is higher than at L_2 in both operating states.

In order to obtain reference results for the new test network, an optimal power flow analysis with all contingencies up to 2^{nd} order was performed, and reliability indices were calculated with the analytical OPAL model [14, 23]. The results show that there is redundancy in the system, and only contingencies of 2^{nd} order or higher give loss of load with the provided load curve. The $EENS$ contribution from each 2^{nd} order contingency is listed in table 2. Note that one of the 4 possible 2^{nd} order contingencies, namely simultaneous outage of branch (1, 4) does not lead to loss of load. The failure and repair rates used in this reference case are given in table 1.

Table 1: Probability parameters of the reference case.

Branch #	Failure Rate λ $[yr^{-1}]$	Repair time $\frac{1}{\mu}$ $[hrs]$
1	2	20
2	0.0036	367.6
3	0.0036	367.6
4	5	10

Table 2: Annual $EENS$ for 2^{nd} order contingencies. The simultaneous outage of branch i and j is denoted (i, j).

Contingency	$EENS$ [kW h]
(3, 4)	311.6
(2, 4)	528.8
(2, 3)	22.2

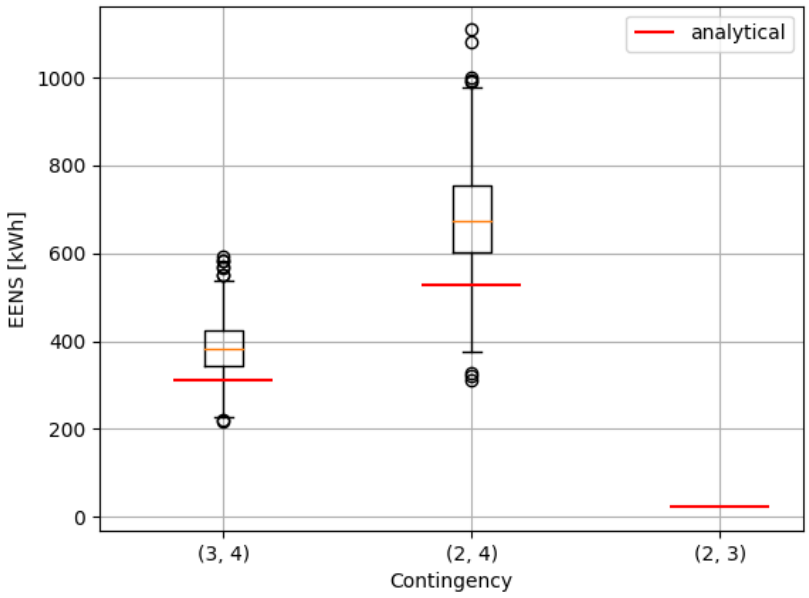
4 Long-term reliability analysis

This section will demonstrate sequential MCS using the Foros-Istad model and how the time dependent failure rate of transformers is expressed on system level in terms of time varying $EENS$. This effect makes the sequential model useful for *long term reliability analysis*. The work presented in this section was done during the authors summer internship, and several different case studies on long term reliability analysis are presented in [3]. Prior to using the Foros-Istad model, the sequential simulation model was verified by comparison with the reference case in section 3.

4.1 Verification of the sequential Monte Carlo model

Sequential simulation of each component state where time-to-failure τ_0 and repair times τ_1 are exponentially distributed with parameters λ and μ is a direct simulation of the continuous time Markov Chain described in section 2.1. Therefore the estimated $EENS$ defined in equation (2.39) should converge towards the analytical reference case. This specific sequential simulation model will be referred to as the *simple sequential model* from now on. The estimated annual $EENS$ for 2^{nd} order contingencies is plotted together with the analytical values in figure 7. The uncertainty in the estimate is visualized by box plots of the bootstrap distribution which was introduced in section 2.5.

Figure 7: $EENS$ per contingency state for the simple sequential model with N=10^5 samples.



The boxplots show that the uncertainty in $EENS$ is big, even for a relatively large sample size, and the relative standard deviation for the two first contingencies in the plot is 36% and 21 % respectively. Interestingly, the simulation captures no ENS from contingency (2, 3) which is the simultaneous outage of both transformers. This is reasonable since the failure rate of the transformers is far smaller than that of the lines, making the probablility of simultaneous outage very low. However the expected repair time of the transformers is

considerably longer than for lines which explains the fact that the analytical $EENS$ for (2, 3) is small but not negligible. The large variance in the $EENS$ estimates, and the fact that the (2, 3) contingency was not captured demonstrates the computational challenge of estimating means that are dominated by rare events with crude Monte Carlo simulation.

4.2 Long term EENS prediction with the sequential model

As described in section 2.6.1, the effect of aging and preventive maintenance in the Foros-Istad model makes the instantaneous failure rate of transformers time dependent. This section presents a case study with $EENS$ prediction over 35 years where the inital health index of both transformers is set to $HI_0 = 0.8$. This corresponds to an OK initial condition. The health index relation and time-to-failure distribution used in this thesis and in case study [26] are

$$\begin{aligned} s &= 10 \ln\left(\frac{1-hi}{hi}\right) + 53, \\ f(s) &= n(\mu = 60,\ \sigma = 18), \\ [s] &= \text{years}. \end{aligned} \tag{4.1}$$

A time horizon of 35 years was chosen because this was seen to capture the transient behaviour of the failure rate well. The estimated failure frequency $\omega(t)$ of a transformer with $HI_0 = 0.8$ is shown in figure 8. Note that the failure frequency is conceptually different to the failure rate $\lambda(t)$. λ is the instantaneous transition rate from up-state to down-state, while ω is the expected frequency of failure which is also influenced by the outage time. However, the numerical value of ω and λ is similar since the outage time is short compared to the time between consecutive failures. The failure frequency is seen to increase to a maximum around year 20 before it decreases steadily until the end of the period. This can be explained by competing factors which drive the failure rate up or down. The effect of component aging and deteriorating condition drives the failure rate up while preventive maintenance and replacement due to failure both have the effect of restoring the transformer to "new" condition, which drives the failure rate down. The effect of aging is dominating until year 20, but in the following years the transformer will have been replaced in a considerable portion of the MC samples and the failure rate decreases.

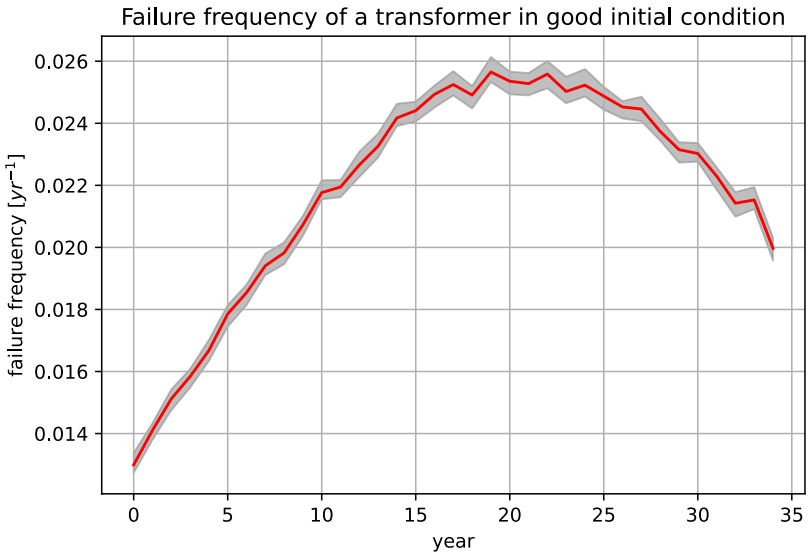


Figure 8: Estimated failure frequency of a transformer with $HI_0 = 0.8$. The grey envelope marks the bootstrap standard deviation.

The $EENS$ estimated from $N = 10^5$ samples is plotted in figure 9. The transient behaviour of the transformer failure rate is clearly expressed on system level as the $EENS$ curve has the same general shape as the failure frequency. The confidence intervals are made from bootstrap quantiles and again shows the challenge of estimating $EENS$ accurately even for a small test system. It should be stressed that the confidence intervals represent the distribution of the $EENS$ estimator, and not the distrubution of ENS. The distribution of ENS is extremely skewed as most of the samples give zero ENS while the few samples that contain contingencies leading to interruptions give very large values. The relative standard deviation of ENS is seen to lie in the range 20-60. The reader is directed to [3] for a more detailed discussion on the ENS distribution.

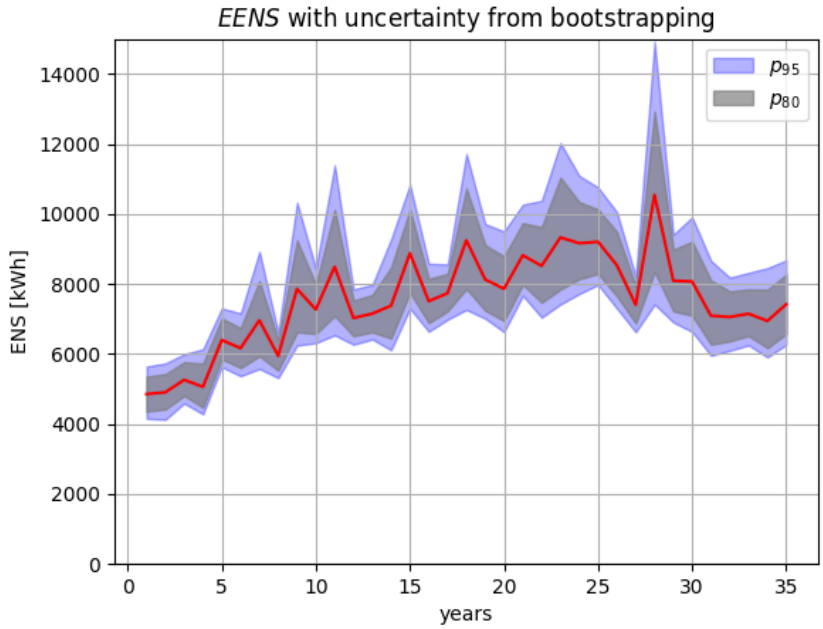


Figure 9: Long term $EENS$ prediction with transformers in good initial condition. The $P95$ and $P80$ envelopes mark confidence intervals made from bootstrap quantiles.

5 Non-sequential Monte Carlo Simulation and Importance Sampling

The non-sequential Monte Carlo model as described in section 2.7 was implemented as a test bed for importance sampling and the cross entropy method. With the non-sequential model, the joint distribution of the system (2.48) is straightforward to evaluate. This is an essential point since the joint distribution appears in the importance sampling estimator (2.14) as the numerator in the likelihood ratio. Another advantage of using the non-sequential model in this first approach is that estimating reliability of supply indices using importance sampling with non-sequential MCS is relatively well studied, see e.g [22, 7] for application of the cross entropy algorithm or see [15] for a similar but more general approach called the *variance minimization method* [20]. The two following subsections will present results from importance sampling with a user defined importance sampling density, and then application of the cross entropy algorithm which automates the search for a good importance sampling density. The results were obtained during the authors summer internship [3], but the discussion and interpretation in this thesis is based on a better theoretical foundation.

5.1 Using a predefined importance sampling density

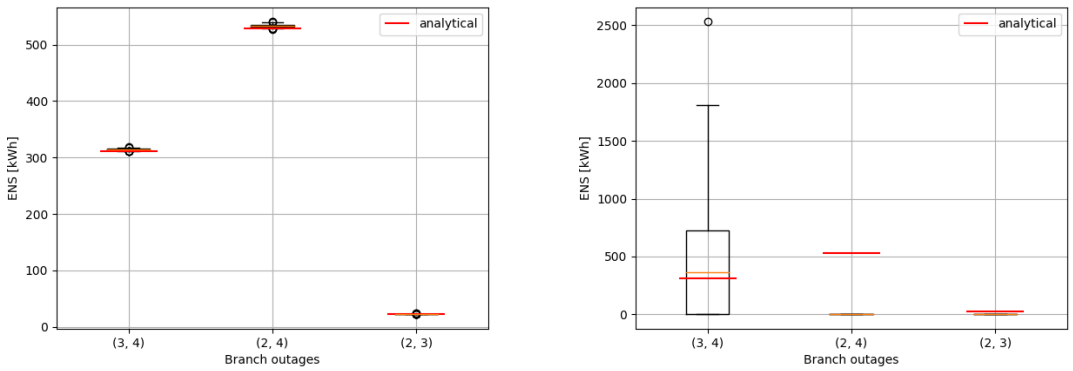
Using the definition of the importance sampling estimator (2.14), the joint distribution of the system (2.48) and the crude $EENS$ estimator (2.49), an importance sampling estimator of $EENS$ can be constructed

$$\widetilde{EENS} = \sum_{i=1}^{N} T \cdot \overline{P}_{\text{interr}}(\mathbf{X}_i) \cdot \frac{f(\mathbf{X}_i; \mathbf{u})}{g(\mathbf{X}_i)}. \tag{5.1}$$

When the importance sampling density is limited to have the same form as f but with different distribution parameters, the likelihood ratio reads

$$W(\mathbf{x}; \mathbf{u}, \mathbf{v}) = \frac{f(\mathbf{x}; \mathbf{u})}{f(\mathbf{x}; \mathbf{v})} = \prod_{i=1}^{n} \frac{u_i^{x_i}(1-u_i)^{1-x_i}}{v_i^{x_i}(1-v_i)^{1-x_i}}. \tag{5.2}$$

From table 1 and the definition of unavailability (2.45), the reference parameters of the test system are $\mathbf{u} = [4.94 \cdot 10^{-3}, 1.68 \cdot 10^{-4}, 1.68 \cdot 10^{-4}, 6.16 \cdot 10^{-3}]$. The importance sampling parameters were set to $\mathbf{v} = [0.5, 0.5, 0.5, 0.5]$ which corresponds to a system where the unavailability is much higher for all components. Since the non-sequential model assumes a steady state distribution, the results should converge towards the analytical reference case in section 3. $EENS$ for 2^{nd} order contingencies from importance sampling is plotted together with the corresponding crude estimate for comparison in figure 10. The importance sampling estimate shows a vast improvement in precision over the crude estimate. The crude simulation only captures ENS for one of the contingencies, but when importance sampling is used, contributions from all contingencies are estimated with negligible uncertainty. It should be pointed out that the non-sequential model with importance sampling consequently give slightly higher $EENS$ values of roughly 1% compared to the analytical model, and this systemic error has not been accounted for.



(a) (b)

Figure 10: Annual $EENS$ with (a): Importance sampling, (b): Crude Monte Carlo. $N = 10^6$ samples were used in both simulations.

In this case "guessing" on a suitable parameter vector $\mathbf{v}$ was very efficient and the method works well for a large range of parameter values. In cases where the dimension of the sample space is high, or the expression for the likelihood ratio is more complicated, finding suitable parameters becomes increasingly challenging. This is due to a problem called *degeneracy* of the likelihood ratio which was briefly discussed in section 2.3.1 leading to high variance in the importance sampling estimate.

5.2 Using the cross entropy algorithm

A slightly modified version of the Cross Entropy algorithm described in section 2.4 was applied to the non-sequential model. Recall that the goal of the cross entropy algorithm is to find the parameter vector $\mathbf{v}$ that minimizes the distance in the cross entropy sense between $f(\mathbf{x}; \mathbf{v})$ and the theoretically optimal importance sampling density. In the non-sequential model, each value of $\overline{P}_{\text{interr}}$ corresponds uniquely to a value of ENS by equation (2.49)

$$ENS(\mathbf{X}) = T \cdot \overline{P}_{\text{interr}}(\mathbf{X}), \tag{5.3}$$

therefore $\overline{P}_{\text{interr}}$ was used as the target function instead of ENS. Since the joint distribution of the system given in (2.48) belongs to the natural exponential family, the stochastic counterpart in equation (2.32) can be used directly and reads

$$v_k^{(t)} \leftarrow \frac{\sum_{j=1}^{N} I\{\overline{P}_{\text{interr}}(\mathbf{X}_j) \geq \gamma_t\} \cdot W(\mathbf{X}_j; \mathbf{u}, \mathbf{v}^{(t-1)}) \cdot X_{jk}}{\sum_{j=1}^{N} I\{\overline{P}_{\text{interr}}(\mathbf{X}_j) \geq \gamma_t\} \cdot W(\mathbf{X}_j; \mathbf{u}, \mathbf{v}^{(t-1)})} \tag{5.4}$$

Considering contingencies up to 2^{nd} order, there are only four possible values for $\overline{P}_{\text{interr}}$, three values for the contingencies that lead to interruptions in table 2, and zero for all other contingencies. Their numerical values are

$$\overline{P}_{\text{interr}} : \{0, 41250\text{kW}, 70000\text{kW}, 111250\text{kW}\}. \tag{5.5}$$

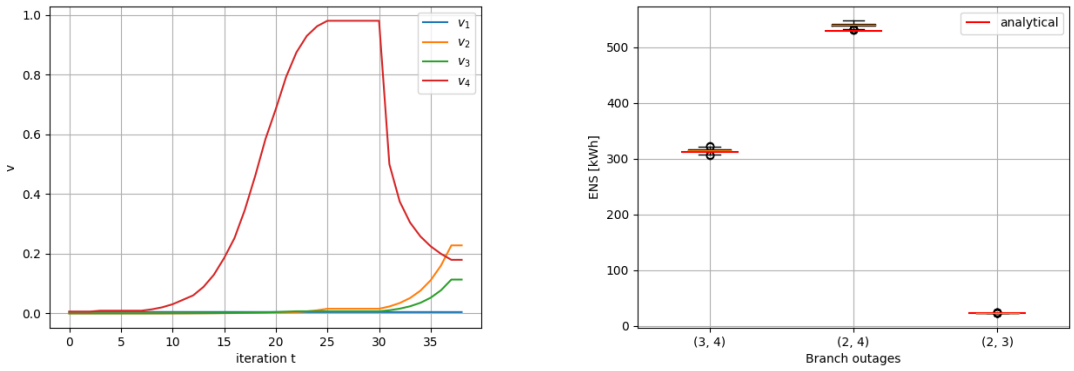
In the first approach, the parameters were updated directly using (5.4). And the stopping criterion was set to $\gamma = 1000$kW. Note that any choice of gamma between $0 < \gamma < 41250$kW would have the same effect. With this approach, the importance sampling distribution was seen to degenerate to an atomic distribution with the entire mass centered

at one point in the sample space. This is a known behaviour of the cross entropy algorithm when applied to finite support discrete distributions and it can even be desirable in cases where the cross entropy method is used in optimization problems. Degeneration of the importance sampling distribution is further discussed in [4, 20]. The ISD was seen to degenerate to either of the three contingencies that lead to interruptions, this happens when the unavailability of two of the components is set to 1, and the two remaining are set to 0. As an example, the parameters $\mathbf{v} = [0, 1, 0, 1]$ leaves the full mass of the ISD on contingency (2, 4). The effect of this degeneration is that the importance sampling estimator finds the analytical $EENS$ value of the corresponding contingency with $N = 1$ sample.

Although interesting, this is clearly unwanted behaviour. To prevent the parameter vector from changing too rapidly or ending up at boundaries in parameter space, the updating formula for the parameter vector (5.4) was modified to

$$v_k^{(t)} \leftarrow v_k^{(t-1)} + \alpha \cdot (v_k^{*(t)} - v_k^{(t-1)}) \cdot v_k^{(t-1)} \tag{5.6}$$

Where v_k^* is calculated from (5.4) and $\alpha \in (0, 1]$ is a tuning parameter that controls the step length. In addition, the stopping criterion was changed to $\gamma = 100000$kW which lies between the highest and second highest value in (5.5). The trajectory of the parameter vector in the modified cross entropy method and the corresponding $EENS$ estimates are shown in figure 11. The tuning parameter was set to $\alpha = 0.5$, and 10^5 samples were used in each iteration. $N = 10^6$ samples were used in the final estimate. Figure 11 (a) shows that the unavailability of branch 4 is initially increased towards 1, but the convergence criterion is not met until v_4 is turned down to a moderate level such that contributions from contingencies where branch 4 is up can also be captured. The unavailability of branch 2 and 3 are also increased, but the unavailability of branch 1 remains very low. This is promising since there are no 2^{nd} order contingencies where branch 1 is out that lead to loss of load. The importance sampling estimates using the final values in the parameter-plot again show a huge improvement over the crude estimate, and the uncertainty is negligible.



(a) (b)

Figure 11: (a): Trajectory of the parameter vector in the cross entropy algorithm. (b): Importance sampling estimate of $EENS$ using the parameter vector from the cross entropy algorithm.

A natural extension would be to investigate how the CE-algorithm works with the non-sequential model on a larger and more realistic system. However, the non-sequential model

was mainly implemented for the purpose of familiarizing with importance sampling and the CE-method, and the main goal of this research is to apply variance reduction techniques to sequential MCS. It should be mentioned that there are established methods to tackle high dimensionality with the CE-method. The so called *screening* method identifies *bottleneck parameters* which are the parameters that are most important for increasing the precision. Adjusting only the bottleneck parameters limits the problems connected to high dimensionality in the likelihood ratio. This idea will be discussed in a different context in section 7.3. The screening method is described in [20] and the method is applied to reliability analysis in [12, 16].

6 Sequential Monte Carlo Simulation and Importance Sampling

This chapter will discuss and present results from importance sampling with the simple sequential model. Recall from section 4.1 that simple sequential means direct simulation of the continuous time Markov chain with transition rates μ and λ. Two different likelihood ratios were tested, one based on the steady state distribution, and one based on the observed transition times. The first gave similar improvements in precision as was seen with the non-sequential model, the drawback is that its use is limited Markov processes. The latter can in principle be applied to any semi-Markov process e.g sequential simulation of the Foros-Istad model. Although it was verified that the importance sampling estimator is unbiased with this choice of likelihood ratio, severe problems were encountered with regards to degeneracy.

6.1 Two different likelihood ratios

A derivation of the two different likelihood ratios used with the simple sequential model will follow. The first is simply the same likelihood ratio used with the non-sequential model, namely

$$W_x(\mathbf{x};\mathbf{u},\mathbf{v}) = \frac{f(\mathbf{x};\mathbf{u})}{f(\mathbf{x};\mathbf{v})} = \prod_{i=1}^{n} \frac{u_i^{x_i}(1-u_i)^{1-x_i}}{v_i^{x_i}(1-v_i)^{1-x_i}}. \tag{6.1}$$

In order for this likelihood ratio to be valid, the Markov chain must have been in operation long enough to reach its stationary distribution. A way of avoiding to simulate the "burn in" time of the chain is to draw the initial system state directly from the stationary distribution (2.48). This likelihood ratio will be referred to as the *steady state likelihood ratio* in the remainder of this thesis, and it is denoted W_x implying that it is evaluated directly with the state vector of the system with no explicit time dependency.

The second likelihood ratio that was used assumes that the system can be described by a semi-Markov process. Formally, this is a process that can be in any one of M states labeled $1,\ldots,M$ and each time it enters state i it stays there for a random amount of time with distribution f_i and mean μ_i before it jumps to any of the other states with transition probability P_{ij} [18]. Note that the continuous time Markov chain is a special case of a semi-Markov process where the times between transitions are exponentially distributed. Adopting the notation from [9], the trajectory $\mathbf{J}$ of the state vector $\mathbf{X}(t)$ is uniquely defined by a sequence of tuples $(t_1,s_1),\ldots,(t_K,s_K)$ where t_i are the transition times and s_i are the corresponding states. The likelihood of the trajectory for a fixed number of transitions K is then

$$l(\mathbf{J}) = \prod_{i=1}^{K} f_{s_i}(t_i - t_{i-1}) \cdot P_{s_{i-1},s_i}. \tag{6.2}$$

The states of components in the system are independent, and therefore one can assign an independent trajectory to each component and write the likelihood for the system as a product of these. A component can be in either of two states, and if the component is in state 1 it will transition to state 0 with probability 1 and vice-versa, thus $P_{1,0} = P_{0,1} = 1$. The likelihood in (6.2) is then reduced to

$$l(\mathbf{J}) = \prod_{i=1}^{K} f_{s_i}(t_i - t_{i-1}), \qquad s_i \in \{0,1\}, \tag{6.3}$$

where the sequence of states s_i alternates between 0 and 1. and f_0 and f_1 are the distributions of the times-to-failure τ_0 and repair times τ_1 correspondingly. However, the simulation is not run for a fixed number of transitions, but rather a fixed time period. Therefore, a variable number of transitions will be observed and the last transition time will be partially censored in the sense that one knows only that the transition time exceeds the observation window, but not by how much. [17] explains how the likelihood can be expressed with different types of censoring, and the censored version of (6.3) is

$$l(\mathbf{J}) = \left(\prod_{i=1}^{K-1} f_{s_i}(t_i - t_{i-1}) \right) \cdot (1 - F_{s_K}(T - t_{K-1})), \tag{6.4}$$

where T is the end time of the simulation and G_{s_K} is the cumulative distribution of g_{s_K}. It should be stressed that the number K is known only after the time series has been observed. Now the likelihood ratio for a single component can be written as

$$w_t(\mathbf{J}) = \left(\prod_{i=1}^{K-1} \frac{f_{s_i}(t_i - t_{i-1})}{g_{s_i}(t_i - t_{i-1})} \right) \cdot \frac{1 - F_{s_K}(T - t_{K-1})}{1 - G_{s_K}(T - t_{K-1})}, \tag{6.5}$$

where f_{s_i} are the reference distributions, g_{s_i} are the ISDs and F_{s_K} and G_{s_K} are the corresponding cumulative distributions. The likelihood ratio for the system trajectory is then

$$W_t(\mathbf{J}_1, \ldots, \mathbf{J}_n) = w_{t,1}(\mathbf{J}_1) \cdot \ldots \cdot w_{t,n}(\mathbf{J}_n), \tag{6.6}$$

and it will simply be referred to as the *trajectory likelihood ratio* from now on.

6.2 Using the steady-state likelihood

Similar to the non-sequential case from section 5.1, importance sampling was done with a user defined parameter vector. In contrast to the non-sequential representation which is fully specified by the unavailability, the sequential model is parameterized by both μ and λ. There is an infinite set of values for λ and μ that give the same value for u, therefore either μ or λ should be fixed when the parameter vector is distorted for importance sampling. The repair rate was fixed to the reference values, and the failure rates were changed such that the resulting unavailability was $\mathbf{v} = [0.1, 0.1, 0.1, 0.1]$. $EENS$ estimates from a simulation with $N = 10^4$ samples are shown in figure 12. The results should be compared to the crude estimate in figure 7, and the improvement in accuracy is comparable to the non-sequential case. The sequential model was seen to consistently overestimate $EENS$ when compared to the analytical model and the cause was identified to be an inconsistency between the models in the number of hours that represent a full year. No attempt was made to fix the inconsistency since it has little or no practical consequence.

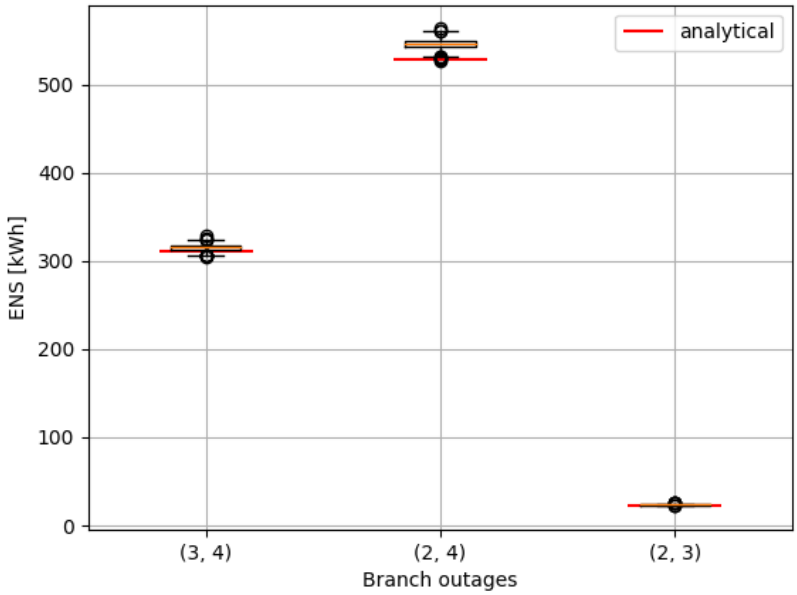


Figure 12: Annual $EENS$ using the steady state likelihood ratio.

As mentioned, use of the steady state likelihood ratio is limited to the simple sequential model. Sequential simulation generally requires more computational effort than non-sequential models, and one might be tempted to draw the conclusion that sequential simulation of a continuous time Markov Chain is obsolete since the same results can be obtained from non-sequential or analytical methods with considerably less CPU time. This is indeed true for the simple test case considered here. However, a sequential simulation can give valuable information which is not readily obtained from non-sequential methods if there are time-dependencies in other parts of the system, examples could be stochastic variation in load or generating capacity. The latter is highly relevant for modeling renewable energy sources. One could also apply this importance sampling method to systems that contain both Markovian and non-Markovian components. This is done by applying importance sampling with the steady state likelihood ratio to the Markovian part of the system while the distributions of non-Markovian components are left unchanged. Since the stationary distribution of the simple sequential model is the same as in the non-sequential case, the cross entropy algorithm described in section 5.2 can readily be applied to the sequential model with minor adjustments. The literature on sequential cross entropy Monte Carlo simulation is rather scarce compared to that on non-sequential methods. Reference [6] gives a good comparison between non sequential, quasi-sequential and sequential methods and applies the cross entropy method to a system of Markovian components combined with a stochastic wind power model. [27] applies a similar sequential cross entropy method with emphasis on including samples from the optimization stage in the final estimates.

6.3 Using the trajectory likelihood ratio with censoring

The main motivation for using the trajectory likelihood ratio is that it can be applied to a general semi-Markov process such as the Foros-Istad model. No references that apply this type of likelihood ratio to system reliability analysis have been found by the author, and therefore a validation of the trajectory likelihood is called for. A test case was constructed

where the unavailability of a single Markovian component was estimated by importance sampling. This problem was chosen since the exact value of the unavailability is known. Recall that the definition of unavailability originates from the stationary distribution of the continuous time Markov Chain, and the simplest way of obtaining a stationary chain is to draw the initial state from the stationary distribution. This introduces an additional factor in the likelihood, namely the likelihood of the initial state under the stationary distribution. The trajectory likelihood then reads

$$l(\mathbf{J}) = P_{s_0} \cdot \left(\prod_{i=1}^{K-1} f_{s_i}(t_i - t_{i-1}) \right) \cdot (1 - F_{s_K}(T - t_{K-1})), \qquad s_i \in \{0, 1\} \tag{6.7}$$

where

$$P_s = u^s(1-u)^{1-s}, \qquad f_0(t) = \text{Exp}(1/\lambda), \qquad f_1(t) = \text{Exp}(1/\mu).$$

The importance sampling estimator for component unavailability is

$$\tilde{u} = \frac{1}{N} \sum_{i=1}^{N} \frac{T_{out}(\mathbf{J}_i)}{T} \cdot W_t(\mathbf{J}_i), \tag{6.8}$$

Where $T_{out}(\mathbf{J})$ is the total time spent in state 1 for a given sample. For the test case, a failure rate of $\lambda = 0.0036$ and repair rate $\mu = 21.94$ was used, which corresponds to the parameters for one of the transformers in table 1. These parameters give an unavailability of $u = 1.64 \cdot 10^{-4}$. For the importance sampling, the repair rate was fixed to the reference value and different scaling factors for λ were tested. The importance sampling estimates and their bootstrap distributions for the different scaling factors are shown in figure 13, $N = 10^5$ samples are used for each estimate. Note that the leftmost box plot with scaling factor 1 is the crude estimate. The precision of the estimate is clearly improved as the failure rate is increased, and a scaling factor of roughly 100-500 seems to give very good precision. The increased variance for the highest failure rates are a consequence of the degeneracy of the likelihood ratio, and the box plot for the highest scaling factor of 4096 shows that the estimate is starting to collapse.

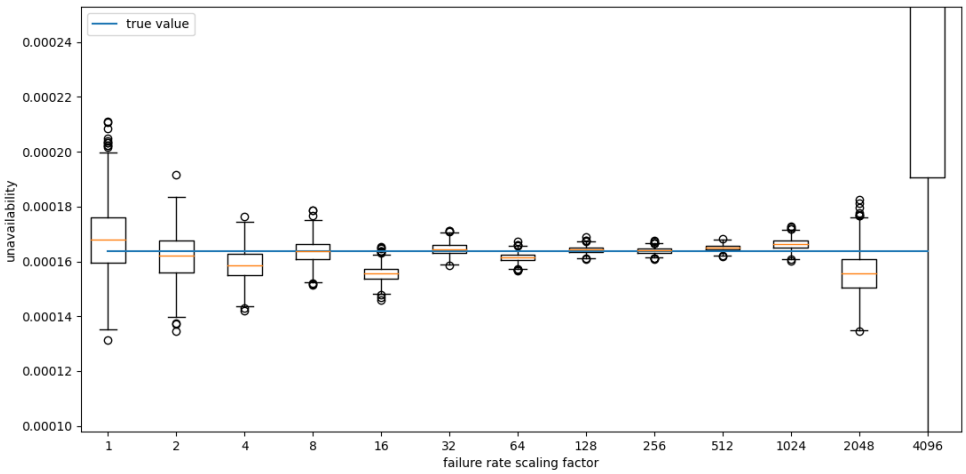


Figure 13: Estimated unavailability for different scaling factors of the failure rate using W_t.

The cause of the degeneration of the importance sampling estimator is in this case twofold. The most obvious cause is that each factor in the likelihood ratio involving λ will typically take on more extreme values when the importance sampling parameter is far from the reference value. The less obvious source of degeneracy comes from increasing dimension. Recall that a variable number of transitions K can occur for each trajectory that is sampled, and each transition is associated with a random variable, namely the time since the previous transition. Phrased differently, both the target function and the likelihood ratio is a function of a random vector with variable dimension. The dimension will tend to increase when the failure rate is higher since transitions from $0 \to 1$ are more frequent. A peculiar consequence of this issue is that the importance sampling estimate will be more prone to degeneracy for longer simulation periods T.

The trajectory likelihood ratio was also tested for $EENS$ estimation, and result were compared to the reference case. With multiple components there is an additional increase in dimension since the system likelihood ratio is a product of each component likelihood ratio, thus one should expect that the importance sampling estimator is even more vulnerable to degeneracy. Finding suitable parameters for importance sampling required some trial and error. It was found that leaving the failure rates of the lines i.e branch 1 and 4 unchanged and only increasing the failure rate of the transformers gave good results. When the failure rate of the lines was changed by a moderate factor, the estimate was seen to break down completely in the sense that all observed values of the likelihood ratio were very close to zero giving practically zero $EENS$. Figure 14 shows the importance sampling estimate of $EENS$ with $N = 10^5$ samples when the failure rate of both transformers is scaled by a factor of 50, and clearly the precision is considerably better than the crude estimate in figure 7. Although this is an interesting result in itself, importance sampling with W_t is not feasible in practice. Even for a test system of only four components, finding suitable parameters by manual search was challenging and for a larger and realistic system this task would be intractable. The cross entropy algorithm can not easily be applied to automatically find optimal parameters. The main reason is that one knows only the expression for the likelihood, and not the PDF of the system trajectory, and the PDF must be known in order to solve the cross entropy optimization problem (2.29).

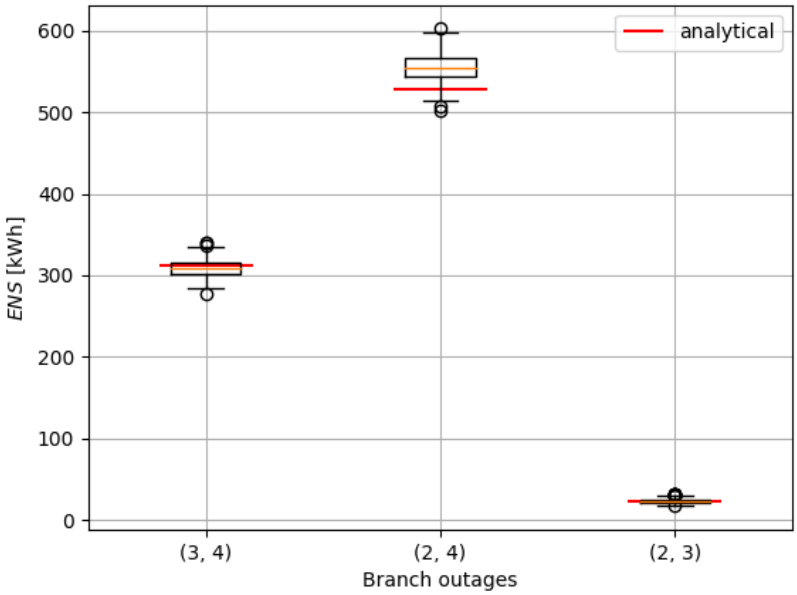


Figure 14: Importance sampling estimate of $EENS$ using W_t.

7 Resampling

The research in section 6 did not result in a viable importance sampling technique for non-markovian components. This chapter will present *importance resampling* as a way to circumvent the use of intricate likelihood ratios, and section 7.2.3 proposes a novel method named *subspace partitioning* to find a close-to-optimal importance resampling distribution. Finally, the subspace partition method is used to reduce the variance of long term $EENS$ predictions with the Foros-Istad model in section 7.3.

7.1 Resampling from the empirical distribution

In the resampling approach, a set of component samples, i.e component trajectories, are considered as an *empirical distribution*. Trajectories are independently resampled from the empirical distribution of each component and are then recombined to form new system trajectories. First, a definition of the empirical distribution is in place.

Definition. (The empirical distribution [5].) Having observed a set of N random samples $\{x_1, \ldots, x_N\}$ from a distribution f, the empirical distribution is the discrete distribution $\hat{f}$ that assigns the probability $1/N$ to each observation x_i. $\hat{f}$ can be expressed as

$$\hat{f}(x) = \prod_{i=1}^{N} N^{-I\{x=x_i\}}. \tag{7.1}$$

When N trajectories $\{\mathbf{j}_1, \ldots, \mathbf{j}_N\}$ are sampled for each component, the joint empirical distribution is

$$\hat{f}(\mathbf{j}_1, \ldots, \mathbf{j_n}) = \prod_{k=1}^{n} \prod_{i=1}^{N} N^{-I\{\mathbf{j}_k=\mathbf{j}_{ki}\}}. \tag{7.2}$$

Since each component trajectory is resampled independently, there is a vast number of possible combinations as the sample space contains N^n points. Therefore it is reasonable to resample $M > N$ system trajectories, and resampling offers a computationally inexpensive increase in sample size which again leads to better precision. However, increasing the sample size must be done with caution, and the improvement in precision is somewhat limited. Consider the example where one has sampled $N = 100$ component trajectories for the 4-branch test system. The sample space consists of $100^4 = 10^8$ elements, and it is tempting to think that it is appropriate to choose M in the order of millions. However, the case may be that the precision of reliability indices is terrible no matter how large M is. The reason is that successful resampling is dependent on the empirical distribution $\hat{f}$ to be a good representation of f. When $N = 100$, $\hat{f}$ may be a very poor representation of the true distribution. The failure rate of the transformers is $\lambda = 0.0036$, and if the simulation period is $T = 1$ year, only about 0.36% of the trajectories will visit the failure state on average. Thus $N = 100$ trajectories for the transformers will typically contain no, or very few outages and obviously $\hat{f}$ has very poor resolution in the interesting region of the sample space. The takeaway is that N must be large enough to capture a considerable number of outages, and it is pointless to increase the sample size by a very large factor.

To found the above discussion on a better theoretical basis, it is useful to introduce the concept of a *resample mean*, which is the expected value under the empirical distribution defined by a sum over all N^n points in sample space.

$$\theta^\dagger = \mathbb{E}_{\hat{f}}[h(\mathbf{X}^\dagger)] = \sum_{x_1} \ldots \sum_{x_n} \hat{f}(\mathbf{x}) \cdot h(\mathbf{x}). \tag{7.3}$$

Where the dagger denotes that $\mathbf{X}^\dagger$ is resampled from $\hat{f}$. Note that in the univariate case $n = 1$, the resample mean is a superficial construction since it is equivalent to the sample average or the crude MC estimate. By resampling, one estimates $\theta^\dagger$ by

$$\hat{\theta}^\dagger = \frac{1}{M} \sum_{i=1}^{M} h(\mathbf{X}^\dagger). \tag{7.4}$$

Thus increasing M gives a better estimate of $\theta^\dagger$

$$\lim_{M \to \infty} \frac{1}{M} \sum_{i=1}^{M} h(\mathbf{X}^\dagger) = \theta^\dagger. \tag{7.5}$$

While increasing N makes $\theta^\dagger$ a better approximation of θ

$$\lim_{N \to \infty} (\theta - \theta^\dagger) = 0, \tag{7.6}$$

where $\theta = \mathbb{E}_f[h(\mathbf{X})]$ as usual.

7.2 Importance resampling

The combined methods of resampling and importance sampling is called importance resampling. There are some interesting papers concerning importance resampling to estimate bootstrap distributions e.g [13, 11]. The resampling approach described in the previous section is not a variant of the bootstrap principle, therefore these articles are not directly relevant to the problem. Nonetheless, these references have come to use in defining some of the mathematical framework which is somewhat similar.

Importance resampling is done by resampling from an ISD $\hat{g}$ that is different from the empirical distribution $\hat{f}$, and the goal is to estimate $\theta^\dagger$ with better precision. $\hat{g}$ is constructed by assigning a probability weight v_{ik} to each observation X_{ij}, where $i = 1, \ldots, N$ and $k = 1, \ldots, n$. The importance resampling estimator is

$$\tilde{\theta^\dagger} = \frac{1}{M} \sum_{i=1}^{M} h(\mathbf{X}^\dagger) \cdot \frac{\hat{f}(\mathbf{X}^\dagger)}{\hat{g}(\mathbf{X}^\dagger)}. \tag{7.7}$$

As for conventional importance sampling, there exists an optimal resampling distribution, namely

$$\hat{g}^*(\mathbf{x}) = \frac{h(\mathbf{x}) \cdot \hat{f}(\mathbf{x})}{\theta^\dagger}. \tag{7.8}$$

Note that the empirical distribution $\hat{f}$ is a finite support discrete distribution, and the solution to the cross entropy problem for a general finite support discrete distribution will be derived in the following subsection.

7.2.1 The cross entropy solution for a finite support discrete distribution

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector where each component is an independent discrete random variable with finite support

$$X_k \in \{a_{1k}, \ldots, a_{mk}\}.$$

The joint distribution of $\mathbf{X}$ can be written as

$$f(\mathbf{x};\mathbf{u}_1,\ldots,\mathbf{u}_n)=\prod_{k=1}^{n} f_k(x_k,\mathbf{u}_k), \tag{7.9}$$

where

$$f_k(x_k,\mathbf{u}_k)=\prod_{i=1}^{m} {u_{ik}}^{I\{x_k=a_{ik}\}}. \tag{7.10}$$

Now one wants to find the parameters $\mathbb{V}\equiv(\mathbf{v}_1,\ldots,\mathbf{v}_m)$ that minimizes $D[g(\mathbf{x})^*,f(\mathbf{x};\mathbb{V})]$. Using equation (2.28), the stochastic counterpart of this optimization problem can be expressed as

$$\mathbb{V}=\arg\max_{v_{ik}}\frac{1}{N}\sum_{j=1}^{N} h(\mathbf{X}_j)W(\mathbf{X}_j,\mathbb{U},\mathbb{W})\ln f(\mathbf{X}_j,\mathbb{V}), \tag{7.11}$$

where

$$\begin{aligned}\ln f(\mathbf{X}_j,\mathbb{V}) &= \ln\prod_{k=1}^{n}\prod_{i=1}^{m} {v_{ik}}^{I\{X_{jk}=a_{ik}\}} \\ &= \sum_{k=1}^{n}\sum_{i=1}^{m} I\{X_{jk}=a_{ik}\}\ln v_{ik}.\end{aligned} \tag{7.12}$$

The maximization problem (7.11) is subject to the n constraints that the marginal distributions f_k must be normalized

$$\sum_{i=1}^{m} v_{ik}-1=0. \tag{7.13}$$

When the constraints have the form $g=0$, the maximization problem (7.11) can be solved by the method of Lagrange multipliers. Define the Lagrange function

$$\begin{gathered}\mathcal{L}(\mathbf{v}_1,\ldots,\mathbf{v}_n,\lambda_1,\ldots,\lambda_n)\equiv \\ \sum_{j=1}^{N} h(\mathbf{X}_j)W(\mathbf{X}_j,\mathbb{U},\mathbb{W})\sum_{k=1}^{n}\sum_{i=1}^{m} I\{X_{jk}=a_{ik}\}\ln v_{ik}-\sum_{k=1}^{n}\lambda_k\left(\sum_{i=1}^{m} v_{ik}-1\right).\end{gathered} \tag{7.14}$$

Now the cross entropy solution is given by the critical point of the Lagrange function

$$\partial_{v_{ik}}\mathcal{L}=0, \tag{7.15}$$

$$\partial_{\lambda_k}\mathcal{L}=0. \tag{7.16}$$

To solve (7.15), we first calculate

$$\begin{aligned}\partial_{v_{ik}}\ln f=\partial_{v_{ik}}\sum_{l=1}^{n}\sum_{q=1}^{m} I\{X_{jl}=a_{ql}\}\ln v_{ql} &= \sum_{l=1}^{n}\sum_{q=1}^{m}\delta_{kl}\delta_{iq}I\{X_{jl}=a_{ql}\}\frac{1}{v_{ql}} \\ &= I\{X_{jk}=a_{ik}\}\frac{1}{v_{ik}},\end{aligned} \tag{7.17}$$

and

$$\partial_{v_{ik}}\sum_{l=1}^{n}\lambda_l\left(\sum_{q=1}^{m} v_{ql}-1\right)=\sum_{l=1}^{n}\lambda_l\sum_{q=1}^{m}\delta_{kl}\delta_{iq}=\lambda_k. \tag{7.18}$$

Then (7.15) is reduced to

$$\begin{aligned} &\sum_{j=1}^{N} h(\mathbf{X}_j) W(\mathbf{X}_j, \mathbb{U}, \mathbb{W}) I\{X_{jk} = a_{ik}\} \frac{1}{v_{ik}} - \lambda_k = 0 \\ \implies & v_{ik} = \frac{1}{\lambda_k} \sum_{j=1}^{N} h(\mathbf{X}_j) W(\mathbf{X}_j, \mathbb{U}, \mathbb{W}) I\{X_{jk} = a_{ik}\}. \end{aligned} \tag{7.19}$$

Now use (7.16) to determine λ_k

$$\begin{aligned} \sum_{i=1}^{m} v_{ik} = \frac{1}{\lambda_k} \sum_{j=1}^{N} h(\mathbf{X}_j) W(\mathbf{X}_j, \mathbb{U}, \mathbb{W}) \underbrace{\sum_{i=1}^{m} I\{X_{jk} = a_{ik}\}}_{=1} = 1 \\ \implies \lambda_k = \lambda = \sum_{j=1}^{N} h(\mathbf{X}_j) W(\mathbf{X}_j, \mathbb{U}, \mathbb{W}). \end{aligned} \tag{7.20}$$

Finally, the cross entropy solution for a multivariate finite support discrete distribution with independent variables is

$$v_{ik} = \frac{\sum_{j=1}^{N} h(\mathbf{X}_j) W(\mathbf{X}_j, \mathbb{U}, \mathbb{W}) I\{X_{jk} = a_{ik}\}}{\sum_{j=1}^{N} h(\mathbf{X}_j) W(\mathbf{X}_j, \mathbb{U}, \mathbb{W})}. \tag{7.21}$$

Strictly, the target function must not change sign to guarantee non-negativity of the weights v_{ik}. This derivation is inspired from a similar proof found in [11] and generalizes it to multiple dimensions.

7.2.2 Estimating optimal probability weights for importance resampling ENS

This subsection will use the results from section 7.2.1 to construct an importance resampling estimator for $EENS$. When N trajectories are sampled for each component, the joint empirical distribution of the system is

$$\hat{f}(\mathbf{j}_1, \ldots, \mathbf{j}_n) = \hat{f}_1(\mathbf{j}_1) \cdot \ldots \cdot \hat{f}_n(\mathbf{j}_n) = \prod_{k=1}^{n} \prod_{i=1}^{N} N^{-I\{\mathbf{j}_k = \mathbf{j}_{ki}\}}. \tag{7.22}$$

This is a multivariate finite support discrete distribution, thus the resampling probability weights that solve the CE problem can be estimated by equation (7.21). The stochastic counterpart reads

$$\begin{gathered} v_{ik} = \frac{\sum_{j=1}^{M} ENS(\mathbf{J}_{j1}^{\dagger}, \ldots, \mathbf{J}_{jn}^{\dagger}) I\{\mathbf{J}_{jk}^{\dagger} = \mathbf{j}_{ik}\}}{\sum_{j=1}^{M} ENS(\mathbf{J}_{j1}^{\dagger}, \ldots, \mathbf{J}_{jn}^{\dagger})}, \\ i = 1, \ldots, N, \\ k = 1, \ldots, n. \end{gathered} \tag{7.23}$$

And the corresponding ISD is

$$\hat{g}(\mathbf{j}_1, \ldots, \mathbf{j}_n; \mathbb{V}) = \prod_{i=1}^{n} \prod_{k=1}^{N} v_{ik}{}^{I\{\mathbf{j}_i = \mathbf{j}_{ik}\}}. \tag{7.24}$$

The full simulation process is illustrated in figure 15, note that there are two resampling steps, one for optimization and one for estimation. First M_1 samples are drawn from the empirical distribution, one may call this uniform resampling as all observations have the same probability. The M_1 samples are used to estimate the optimal probability weights by equation (7.23). In the second resampling stage M_2 samples are drawn according to the ISD, these samples are used to estimate reliability indices.

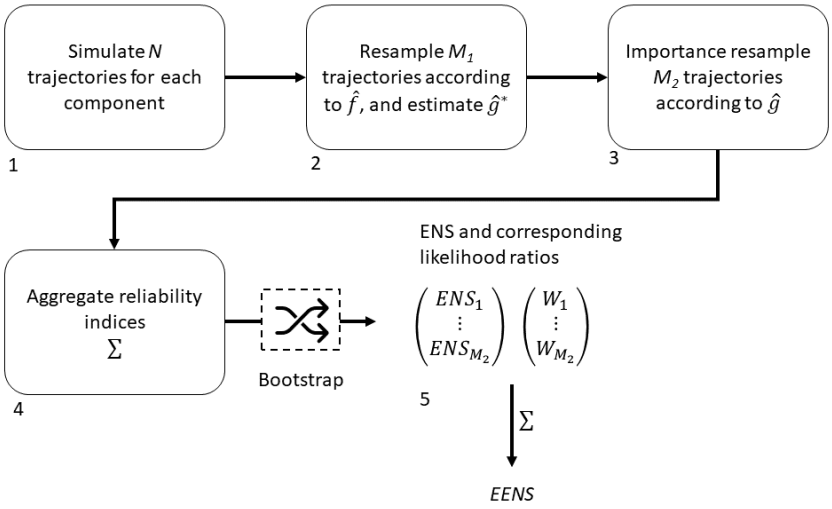


Figure 15: Flow chart of the importance resampling method

Results from 3 different simulations, all with $N = 10^4$ and $M_1 = M_2 = 10^6$ are shown in figure 16 and figure 17. When bootstrapping is applied between step 4 and 5 in the process as illustrated in figure 15, the bootstrap distribution represents the distribution of the resampling estimator for a given empirical distribution. I.e. the bootstrap distribution corresponds to the conditional distribution of $\hat{\theta}^\dagger | \hat{f}$, where $\hat{\theta}^\dagger$ is defined in equation 7.4. Ideally, bootsrapping should be applied between step 1 and 2 instead as this would yield the unconditioned distriution of $\hat{\theta}^\dagger$. However, steps 2-4 would then have to be repeated as many times as the number of bootstrap samples which would be very time consuming. In summary, the boxplots can be viewed as confidence intervals of the resample mean for a given empirical distribution. The plots show that the variance is consequently lower for importance resampling compared to uniform resampling. The estimates are relatively far off from the analytical values which indicates that the resample mean is a poor approximation of the true mean for a sample size of $N = 10^4$. Figure 17 shows that importance resampling is able to estimate $EENS$ for contingency (2, 3), which is a rare event, while the uniform resampling approach fails.

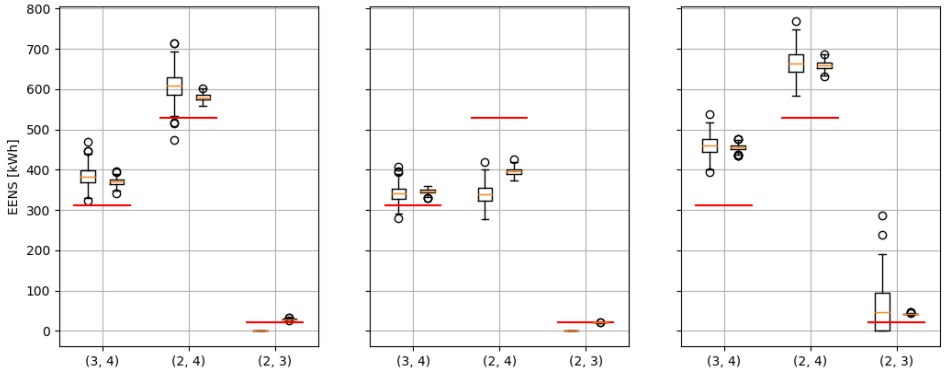


Figure 16: Comparison of $EENS$ estimates from uniform resampling and importance resampling. Each plot corresponds to an independent simulation, and the pairs of boxlots show uniform resampling (left) and importance resampling (right). The analytical values are shown as red lines.

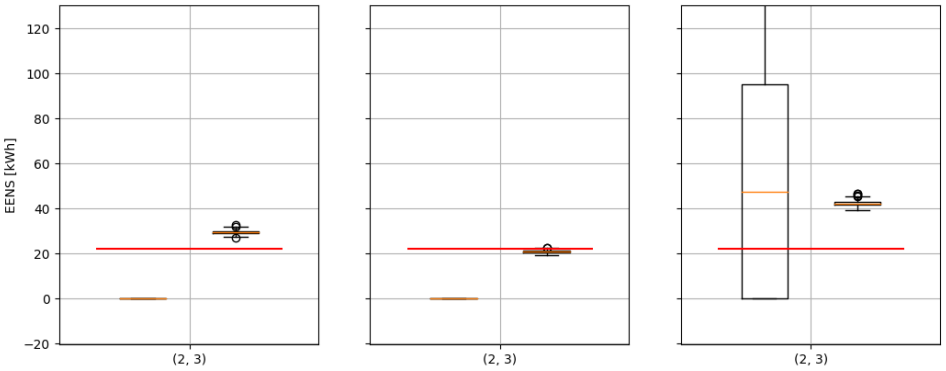


Figure 17: The same plot as in figure 16 focused at contingency (2, 3).

Looking at figure 16 and 17, it is clear that a sample size of $N = 10^4$ is insufficient, i.e the empirical distribution does not contain enough information to provide an accurate resample mean. The caveat with this approach to importance resampling is that the number of probability weights that needs to be estimated for each component is N. Therefore, if the sample size N is increased, more parameters needs to be estimated. This leads to a substantial increase in the computational load of the optimization step since the number of resamples during optimization should be much larger than the number of parameters to estimate, $M_1 >> N$. Because of the challenging optimization stage, this approach is mainly of theoretical interest. The following section will present a new method named *subspace partitioning*, which gives a dramatic reduction in the number of free parameters.

7.2.3 The subspace partition method

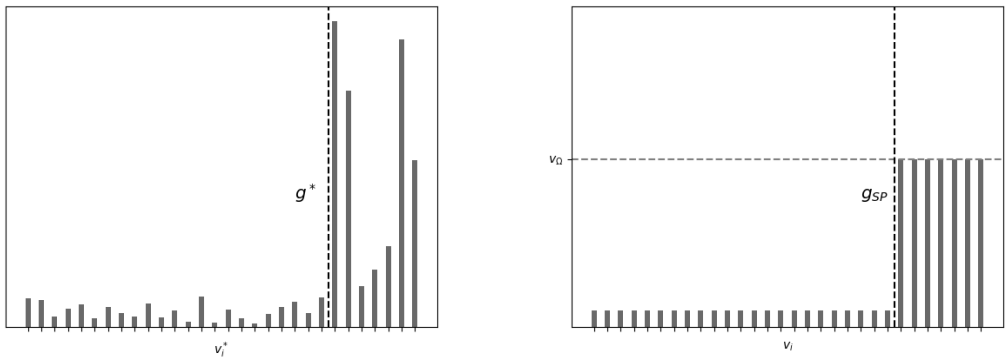
The subspace partitioning method (SPM) is a cross entropy solution for finite support discrete distributions where the ISD is constrained in a way that reduces the number of

free parameters in the optimization stage. Before the cross entropy solution is derived, the idea behind the method will be illustrated.

Example. (The subspace partitioning method visualized.) Consider the generic case where one wants to use importance sampling to estimate the expected value of a target function $\theta = \mathbb{E}[h(X)]$, where X is a discrete random variable with finite support. The optimal ISD is

$$g^*(x) = \frac{h(x)\prod_{i=1}^{m} u_i^{I\{x=a_i\}}}{\theta}.$$

A hypothetical sketch of g^* is shown in figure 18a. Note that the average likelihood in the subspace to the right of the dashed line is much higher than in the subspace to the left. This indicates that the right subspace is the important region to sample from. The idea of the subspace partition method is to constrain the ISD to have uniform likelihood in each subspace equal to the average likelihood under g^*. This reduces the number of free parameters of the ISD to the number of subspaces chosen for the problem. The subspace partition ISD (g_{SP}) is sketched in figure 18b. When sampling form g_{SP}, the same portion of samples will lie in the important region as if one had sampled directly from g^*.



(a) (b)

Figure 18: (a): The optimal ISD. (b): The subspace partitioning ISD.

When the subspace partition method is applied to importance resampling of ENS, the sample space is partitioned into two subspaces. One subspace contains all the component trajectories that visit a failure state, and the complementing subspace contains the remaining trajectories. There is one parameter for each of the two subspaces, but only one of them needs to be estimated since the other one is determined by normalization.

7.2.4 The cross entropy solution for the subspace partition method

Now a formal derivation of the cross entropy solution given the subspace parametrization will follow. Consider the same situation as in section 7.2.1 with a random vector $\mathbf{X} = (X_1, \ldots, X_n)$ where each variable can take the m values

$$X_k \in \Omega_k = \{a_{1k}, \ldots, a_{mk}\}.$$

As earlier, the PDF of $\mathbf{X}$ is

$$f(\mathbf{x}; \mathbb{U}) = \prod_{k=1}^{n}\prod_{i=1}^{m} {u_{ik}}^{I\{X_{jk}=a_{ik}\}}. \tag{7.25}$$

Partition the sample space Ω_k of each variable into G disjoint subspaces

$$\Omega_k = \Omega_{1k} \cup \ldots \cup \Omega_{Gk}, \tag{7.26}$$

$$\Omega_{ik} \cap \Omega_{jk} = \varnothing \quad i \neq j. \tag{7.27}$$

With the subspace paramterization, the ISD can be written as

$$g(\mathbf{x}; \mathbb{V}) = \prod_{k=1}^{n} \prod_{i=1}^{G} v_{ik}{}^{I\{X_{jk} \in \Omega_{ik}\}}. \tag{7.28}$$

where v_{ik} are the subspace parameters. Following a very similar procedure as in section 7.2.1, the cross entropy problem will be solved by Lagrange multipliers. The cross entropy problem reads

$$\mathbb{V} = \arg\max_{v_{ik}} \frac{1}{N} \sum_{j=1}^{N} h(\mathbf{X}_j) W(\mathbf{X}_j; \mathbb{U}, \mathbb{W}) \ln g(\mathbf{X}_j; \mathbb{V}). \tag{7.29}$$

The normalization constraints can be written as

$$\sum_{i=1}^{G} m_i \cdot v_{ik} - 1 = 0, \tag{7.30}$$

where m_i are the number of elements in each subspace, $m = m_1 + \ldots + m_G$. From equations (7.29) and (7.30), the Lagrange function can be defined as

$$\mathcal{L}(\mathbf{v}_1, \ldots, \mathbf{v}_n, \lambda_1, \ldots, \lambda_k) \equiv$$
$$\sum_{j=1}^{N} h(\mathbf{X}_j) W(\mathbf{X}_j; \mathbb{U}, \mathbb{W}) \sum_{k=1}^{n} \sum_{i=1}^{G} I\{X_{jk} \in \Omega_{ik}\} \ln v_{ik} - \sum_{k=1}^{n} \lambda_k \left(\sum_{i=1}^{G} m_i \cdot v_{ik} - 1 \right). \tag{7.31}$$

And the CE solution is given by

$$\partial_{v_{ik}} \mathcal{L} = 0, \tag{7.32}$$

$$\partial_{\lambda_k} \mathcal{L} = 0. \tag{7.33}$$

(7.32) reduces to

$$\begin{aligned} &\sum_{j=1}^{N} h(\mathbf{X}_j) W(\mathbf{X}_j; \mathbb{U}, \mathbb{W}) I\{X_{jk} \in \Omega_{ik}\} \frac{1}{v_{ik}} - \lambda_k m_i = 0 \\ &\implies v_{ik} = \frac{1}{\lambda_k m_i} \sum_{j=1}^{N} h(\mathbf{X}_j) W(\mathbf{X}_j; \mathbb{U}, \mathbb{W}) I\{X_{jk} \in \Omega_{ik}\}. \end{aligned} \tag{7.34}$$

And using equation 7.33 one finds that

$$\lambda_k = \lambda = \sum_{j=1}^{N} h(\mathbf{X}_j) W(\mathbf{X}_j; \mathbb{U}, \mathbb{W}). \tag{7.35}$$

Thus the stochastic counterpart of the CE solution for the subspace partition approach is

$$v_{ik} = \frac{1}{m_i} \cdot \frac{\sum_{j=1}^{N} h(\mathbf{X}_j) W(\mathbf{X}_j; \mathbb{U}, \mathbb{W}) I\{X_{jk} \in \Omega_{ik}\}}{\sum_{j=1}^{N} h(\mathbf{X}_j) W(\mathbf{X}_j; \mathbb{U}, \mathbb{W})}. \tag{7.36}$$

7.2.5 The subspace partition method for importance resampling ENS

One may use the stochastic counterpart for the subspace parameters in (7.36) directly, but there are strong arguments for using the iterative CE algorithm as described in section 2.4. Using iteration, the estimate for the optimal parameters typically becomes more accurate, and fewer samples are needed. In addition, the iterative algorithm incorporates an automatic stop criterion. Both approaches were tested with similar results, but the iterative CE algorithm gave more consistent values for the parameters. Therefore, this section will focus on the iterative approach. The Cross Entropy algorithm for importance resampling ENS is as follows.

Step 0. (Preparation.) Simulate the N trajectories for each component in the system which defines the empirical distribution $\hat{f}$. For each component k, define the subspace Ω_{fk} which contains all trajectories that visit a failure state, and let N_{fk} be the number of elements in Ω_{fk}. The complementing subspace Ω_{fk}^c contains all remaining trajectories.

Step 1. Set the initial value of the subspace parameters to the reference parameters $v_{fk}^{(0)} \leftarrow u_{fk} = \frac{1}{N}$, and set $\rho \leftarrow 0.1$.

Step 2. Resample M_1 trajectories according to the importance sampling distribution $\hat{g}(\mathbf{j}_1, \ldots, \mathbf{j}_n; \mathbb{V}^{(t)})$, and calculate the corresponding ENS values. Sort the ENS values such that $ENS^{(1)} \leq \ldots \leq ENS^{(M_1)}$ and set γ_t to be the $1 - \rho = 90\%$ quantile of $\{ENS^{(j)}\}$.

Step 3. For each component $k = 1, \ldots, n$ update the subspace parameter

$$v_{fk}^{(t+1)} \leftarrow \frac{1}{N_{fk}} \cdot \frac{\sum_{j=1}^{M_1} I\{ENS(\mathbf{X}_j^\dagger) > 0\} W(\mathbf{X}_j^\dagger; \mathbb{U}, \mathbb{V}^{(t)}) I\{X_{jk}^\dagger \in \Omega_{fk}\}}{\sum_{j=1}^{M_1} I\{ENS(\mathbf{X}_j^\dagger) > 0\} W(\mathbf{X}_j^\dagger; \mathbb{U}, \mathbb{V}^{(t)})}, \tag{7.37}$$

where $\mathbf{X}^\dagger \equiv (\mathbf{J}_1^\dagger, \ldots, \mathbf{J}_n^\dagger)$ for brevity. The parameter for the non-failure subspace is given by normalization.

Step 4. If $\gamma_t > 0$, i.e if at least 10% of the M_1 samples give non-zero ENS, use $\mathbb{V}^{(t+1)}$ as the final parameter values. If not, repeat step 2-4.

Step 5. Importance resample M_2 trajectories according to $\hat{g}(\mathbf{x}; \mathbb{V}^{(t+1)})$ and estimate $EENS$.

The results from the subspace partition method using $N = 10^6$, $M_1 = 10^4$ and $M_2 = 10^6$ samples for 3 consecutive, independent simulations are are shown in figure 21. The results from uniform resampling with $M = 10^6$ are plotted for comparison. The subspace partition method clearly outperforms the accuracy of uniform resampling for all contingencies, and the method excels at estimating $EENS$ for contingency (2, 3) which is the rarest event. In all 3 simulations, the algorithm converged after 2 iterations. In the second step of the iteration, about 18% of the samples have $ENS > 0$, while the corresponding number at step 1 which represents uniform resampling is around 0.1%. The ENS distribution in the two iteration steps is visualized in figure 19. Not only is the portion of non-zero samples much larger in the 2^{nd} iteration, but the range of observed ENS values is also considerably wider. The trajectories of the subspace parameters are shown in figure 20. The values of the subspace parameters are clearly much more consistent in the 2^{nd} iteration which justifies the iterative approach. Note that the parameters for branch 1 and 4 remain very close to their original values, while the parameters for the transformers which are branch 2 and 3 are scaled up to around 140 times the reference value. The number of samples

used in the optimization stage is $2 \cdot M_1 = 2 \cdot 10^4$ which is only 2% of the total sample size N, and the CPU time required by the optimization stage is negligible compared to both the initial MC simulation and the final estimation stage.

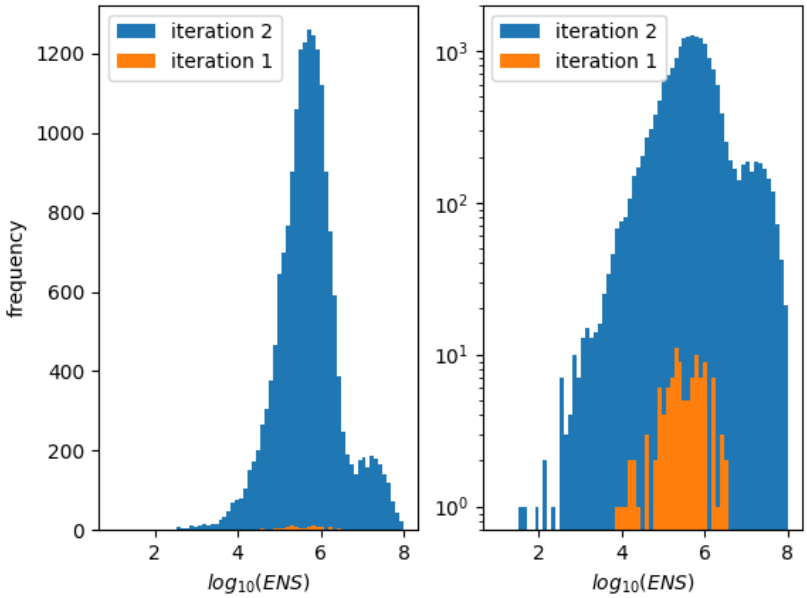


Figure 19: Histograms of samples with non-zero ENS from the 1^{st} and 2^{nd} iteration of the CE algorithm. The x-axis is logarithmic, and both plots show the same data. In the right plot, the y-axis is logarithmic to better show the distribution from the 1^{st} iteration.

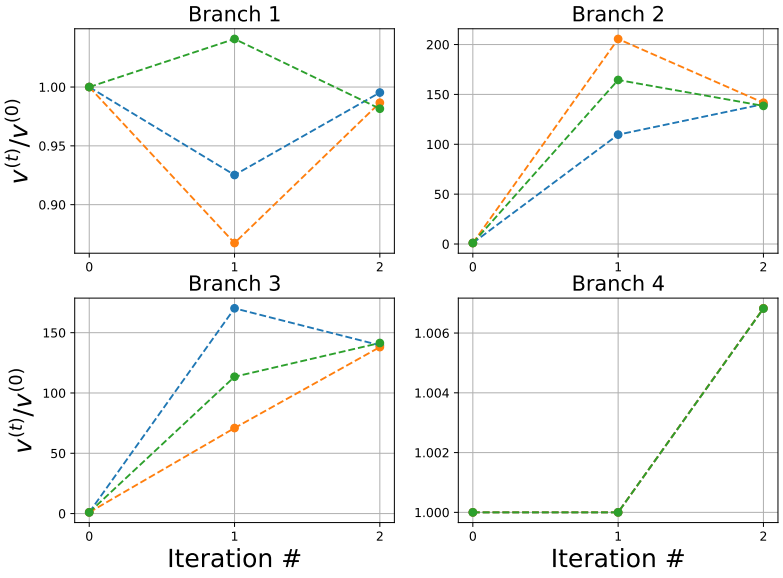
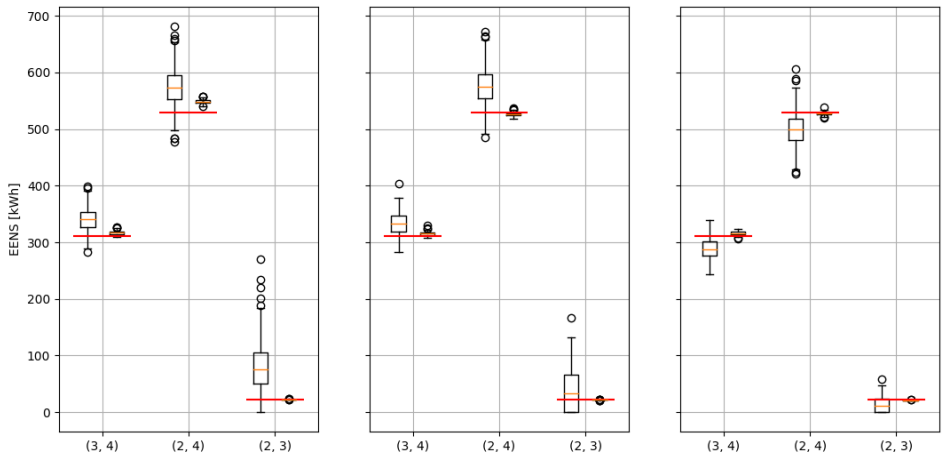
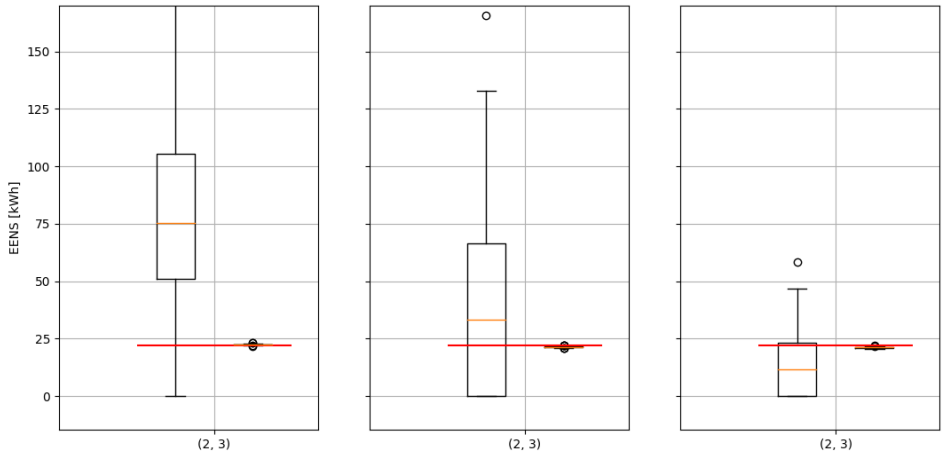


Figure 20: Evolution of all 4 subspace parameters for 3 simulations. The values are relative to the reference parameters (iteration 0). Only one line is visible for branch 4 since all 3 trajectories are virtually indentical.



(a)



(b)

Figure 21: $EENS$ per contingency for 3 consecutive, independent simulations using $N = 10^6$ and $M_2 = 10^6$ with uniform resampling (left box plots) and the SPM (right box plots). Analytical values are shown in red. (a): Showing all contingencies. (b): Focused at the (2, 3) contingency.

The SPM can only be as good as as the resample mean, and an important and partly unanswered question is how good the resample mean represents the true mean. The resample mean is a function of the empirical distribution, and thus it must be regarded as a random variable. It follows from the central limit theorem that the resample mean is normally distributed. Thus, the main unanswered question is how the variance of the resample mean scales with the number of samples N, the number of components n, and whether there exist a general asymtotic convergence rate as a function of these. Answering this question is far from trivial since the resample mean is not computable in practice for any relevant value of N. In summary, the SPM offers substantialy better accuracy than CMC or uniform resampling with the caveat that one looses the possibility of assesing

the true uncertainty in a simple manner. Table 3 gives a quantitative comparison of the precision between uniform resampling and and the subspace partition method, but the reader should keep in mind that these values represent the uncertainty relative to the resample mean, and not the true mean.

Table 3: Bootstrap standard deviation in kW h for each of the three independent simulations. The three main rows represent uniform resampling, the subspace partition method and the relative reduction in uncertainty. Each subrow represent an independent simulation.

Contingency	(3, 4)	(2, 4)	(2, 3)
Uniform [kW h]	19.6	30.5	42.4
	19.0	31.4	32.2
	16.9	27.5	12.4
SPM [kW h]	2.91	3.00	0.210
	3.06	2.87	0.217
	3.11	2.81	0.216
Improvement [%]	85.2	90.2	99.5
	83.9	90.9	99.3
	82.6	89.8	98.7

7.3 Application to the Foros-Istad model

To verify that the subspace partition method is applicable to a very general sequential simulation, the SPM was used with the Foros-Istad transformer model. In the previously used reference case, the lines have unrealistically high failure rates, $\lambda_1 = 2$ for branch 1 and $\lambda_4 = 5$ for branch 4. To test the method on a more reliable system, the failure rates for the lines were set to $\lambda_1 = 0.0045$ and $\lambda_4 = 0.027$. These are representative parameter values for power lines found in SINTEF's 4-area test network [24]. The initial health index for both the transformers were set to $HI_0 = 0.95$ which corresponds to good technical condition. As previously, 3 independent simulations were run in order to get some indication of the consistency in the results. $EENS$ for 2^{nd} order contingencies for all 3 simulations are presented in figure 22. One million samples were used both in the initial simulation and in the final estimation step, and the crude estimator is shown for comparison. An analytical approximation was computed as reference value using the same procedure as in [25].

The portion of samples containing a loss of load event is in the range $(1-3)\cdot 10^{-5}$ for the crude Monte Carlo simulations, thus the system is very reliable and interruptions are truly rare events. The results suggest that the subspace partition method retains its efficiency on more reliable grid configurations. As a measure of performance, the bootstrap root mean square error (RMSE) is reported in table 4 and compared to the crude estimator. The analytical approximation is used as the true mean to compute the RMSE.

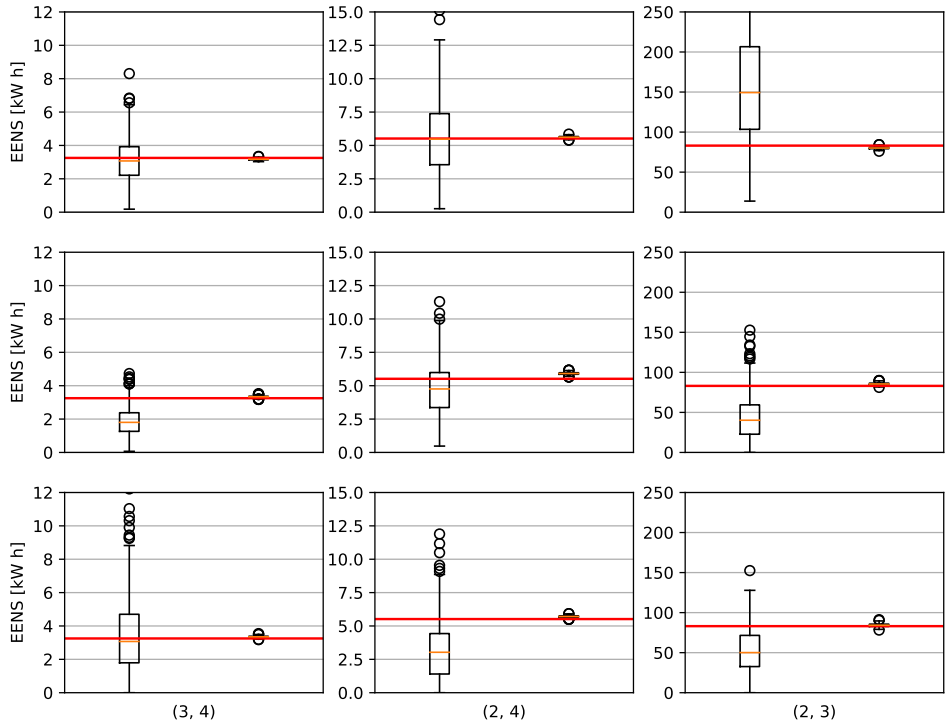


Figure 22: $EENS$ with the crude estimator (left box plots), the subspace partition method (right box plots) and an analytical approximation (red lines). Each row corresponds to an independent simulation, and each column corresponds to a given contingency.

Table 4: Bootstrap root mean square error for three independent, consecutive simulations. The three main rows report RMSE from crude Monte Carlo, the subspace partition method and the relative reduction in RMSE. The average reduction in RMSE over all contingencies is 92.8%.

Contingency	(3, 4)	(2, 4)	(2, 3)
CMC [kW h]	1.246	2.70	103
	1.64	2.09	50.5
	2.13	3.32	38.4
SPM [kW h]	0.0954	0.138	3.29
	0.102	0.403	2.84
	0.112	0.215	2.24
Improvement [%]	92.3	94.9	96.9
	93.7	80.7	94.4
	94.7	93.5	94.2

7.4 Remarks on extended application of the subspace partition method

To further test and develop the subspace partition method, it should be applied to a larger and more realistic grid, such as the 4-area test network developed at SINTEF [24]. The dimension of the problem increases with a larger system, and importance sampling is generally more challenging in high dimension. However, the inventors of the cross entropy algorithm have developed a technique called the two-stage screening method [20] to manage the degeneracy of the likelihood ratio in high dimension. This was briefly discussed in section 5.2. The screening method identifies the distribution parameters that have the largest relative change as bottleneck parameters, and resets the non-bottleneck parameters to the reference value. As an example, figure 20 suggests that the transformers are bottleneck components in the reference case since the parameters for the lines remain close to their original value. For a larger system, the screening method could have a substantial effect. One can regard the bottleneck components as the most important for the precision of the estimate. Therefore, the author suggests that the number of samples should be increased for the bottleneck components in order to obtain better resolution of their empirical distributions, but this has yet to be tested.

An attempt was also made to apply the SPM to long-term prediction of $EENS$ similar to what was done in section 4. It was found that the component trajectories should be split into periods of one year, and resampled individually for each period to achieve variance reduction. The reason behind this is that when the simulation period is longer, a larger portion of the trajectories will visit a failure state at some point, but the component spends very little time in the failure state compared to the total simulation time. Therefore, failures are unlikely to overlap in time for two individual trajectories which both contain one failure event. Splitting the time series post simulation is inefficient with the current implementation of the Monte Carlo model. Therefore, this was only attempted with a relatively small sample size which means that importance resampling is not very effective. Modest variance reduction was observed, but these results are omitted from the thesis.

8 Concluding remarks and suggestions for further research

For the final discussion, it is natural to restate the research question defined in section 1.3:

Q: Can importance sampling reduce the variance of reliability index estimates for a sequential Monte Carlo simulation with time dependent component failure rates?

The trajectory likelihood method proposed in section 6 revealed that changing the parameters which govern the underlying stochastic processes in the system quickly leads to complicated expressions for the likelihood ratio. This can in turn make the importance sampling estimator vulnerable to degeneracy, and the cross entropy optimization problem becomes intractable. To avoid this problem, the remainder of the work was focused on different variations of importance resampling. The main advantage of importance resampling is that, in principle, no knowledge of the system's PDF is required which makes it applicable to a very general model. The subspace partition method is the most important result from the research. It unites the ideas of resampling, importance sampling, and the cross entropy algorithm, and it has achieved substantial improvement in precision on a reliable system configuration. The answer to the research question is clear: The SPM can reduce the variance of reliability index estimates on a sequential MCS with time dependent component failure rates. This can in turn provide system planners with previously unavailable information on how time dependent processes such as maintenance and aging influence the reliability of supply.

Importance resampling can only perform as good as the resample mean which was introduced in section 7.1. The main weakness of the SPM, and importance resampling in general is that the accuracy and validity of the resample mean is largely unknown. From a limited number of empirical observations one can however say that it seems to coincide quite well with the true mean as long as the sample number is sufficiently big. A closely related problem is that the uncertainty of MC estimates can not easily be evaluated by bootstrapping. An unbiased estimate of the uncertainty can be obtained by brute force running multiple independent simulations, but this is unpractical for obvious reasons. Further research on the subspace partition method, or other resampling techniques should focus on establishing better theoretical knowledge of the resample mean. Additionally, the subspace partition method should be tested on a larger, realistic power system.

References

[1] Roy Billinton and Ronald N. Allan. *Reliability Evaluation of Power Systems, Second Edition*. Springer Science + Business Media LLC, 1996.

[2] Roy Billinton and Steve Andzanu. "Composite generation and transmission system adequacy assessment with time varying loads using a contingency enumeration approach". In: *IEEE WESCANEX 97 Communications, Power and Computing. Conference Proceedings* (1997), pp. 41–46.

[3] Ivar Bjerkebæk. "Documentation of Research During Summer Internship EN-05: Simulation of Reliability in the Power System". Project Memo, SINTEF Energy Research. 2022.

[4] Pieter-Tjerk De Boer and Dirk P. Kroese. "A Tutorial on the Cross-Entropy Method". In: *Annals of Operations Research* 134 (2005), pp. 19–67.

[5] Bradley Efron and Robert J Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, 1993.

[6] A. G. Fernández and M. Leite da Silva. "Reliability assesment of Time-Dependent Systems via Sequential Cross-Entropy Monte Carlo Simulation". In: *IEEE Transactions on Power Systems* 26 (2011), pp. 2381–2389.

[7] A. G. Fernández et al. "Composite Systems Reliability Evaluation Based on Monte Carlo Simulation and Cross-Entropy Methods". In: *IEEE Transactions on Power Systems* 28 (2013), pp. 4598–4606.

[8] Jørn Foros and Maren Istad. "Health Index, Risk and Remaining Lifetime Estimation of Power Transformers". In: *IEEE Transactions on Power Delivery* 35 (2019), pp. 2612–2620.

[9] Adam W. Grace, Dirk P. Kroese, and Werner Sandmann. "Automated State Dependent Importance Sampling for Markov Jump Processes via Sampling From the Zero-Variance Distribution". In: *Journal of Applied Probability* 51 (2014), pp. 741–755.

[10] Tito Homem-de-Mello and Reuven Y. Rubinstein. "Estimation of Rare Event Probabilities Using Cross-Entropy". In: *Proceedings of the 2002 Winter Simulation Conference* (2002), pp. 310–319.

[11] Jiaqiao Hu and Zheng Su. "Adaptive Resampling Algorithms for Estimating Bootstrap Distributions". In: *Journal of Statistical Planning and Inference* 138 (2008), pp. 1763–1777.

[12] K. P. Hui et al. "The Cross-Entropy Method for Network Reliability Estimation". In: *Annals of Operations Research* 134 (2005), pp. 101–118.

[13] M. Vernon Johns. "Importance Sampling for Bootstrap Confidence Intervals". In: *Journal of the American Statistical Association* 83 (1988), pp. 709–714.

[14] Gerd Kjølle and Oddbjørn Gjerde. "The OPAL methodology for reliability analysis of power systems". Report no. TR A7175, SINTEF Energy Research. 2012.

[15] Dmitrii Lieber, Arkadii Nemirovskii, and Reuven Y. Rubinstein. "A Fast Monte Carlo Method for Evaluatig Reliability Indices". In: *IEEE Transactions on Reliability* 48 (1999), pp. 256–261.

[16] Dmitrii Lieber, Reuven Y. Rubinstein, and David Elmakis. "Quick Estimation of Rare Events in Stochastic Networks". In: *IEEE Transactions on Reliability* 46 (1997), pp. 254–256.

[17] William Q. Meeker, Luis A. Escobar, and Francis G. Pascual. *Statistical Methods for Reliability Data, 2^{nd} Edition*. Wiley Series in Probability and Statistics, 2022. Chap. 2, Models Censoring and Likelihood for Failure Time Data.

[18] Sheldon M. Ross. *Introduction to Probability Models, Twelfth Edition*. Elsevier Academic Press, 2019.

[19] Reuven Y. Rubinstein. "Optimization of computer simulation models with rare events". In: *European Journal of Operational Research* 99 (1997), pp. 89–112.

[20] Reuven Y. Rubinstein and Dirk P. Kroese. *Simulation and the Monte Carlo Method, 2^{nd} Edition*. Wiley Series in Probability and Statistics, 2007.

[21] G. I. Schuëller, H. J. Pradlwarter, and P. S. Koutsourelakis. "A critical appraisal of reliability estimation procedures for high dimension". In: *Probabilistic Engineering Mechanics* 19 (2004), pp. 463–474.

[22] M. Leite da Silva, A. G. Fernández, and Singh. "Generating Capacity Evaluation Based on Monte Carlo Simulation and Cross-Entropy Methods". In: *IEEE Transanctions on Power Systems* 25 (2010), pp. 129–237.

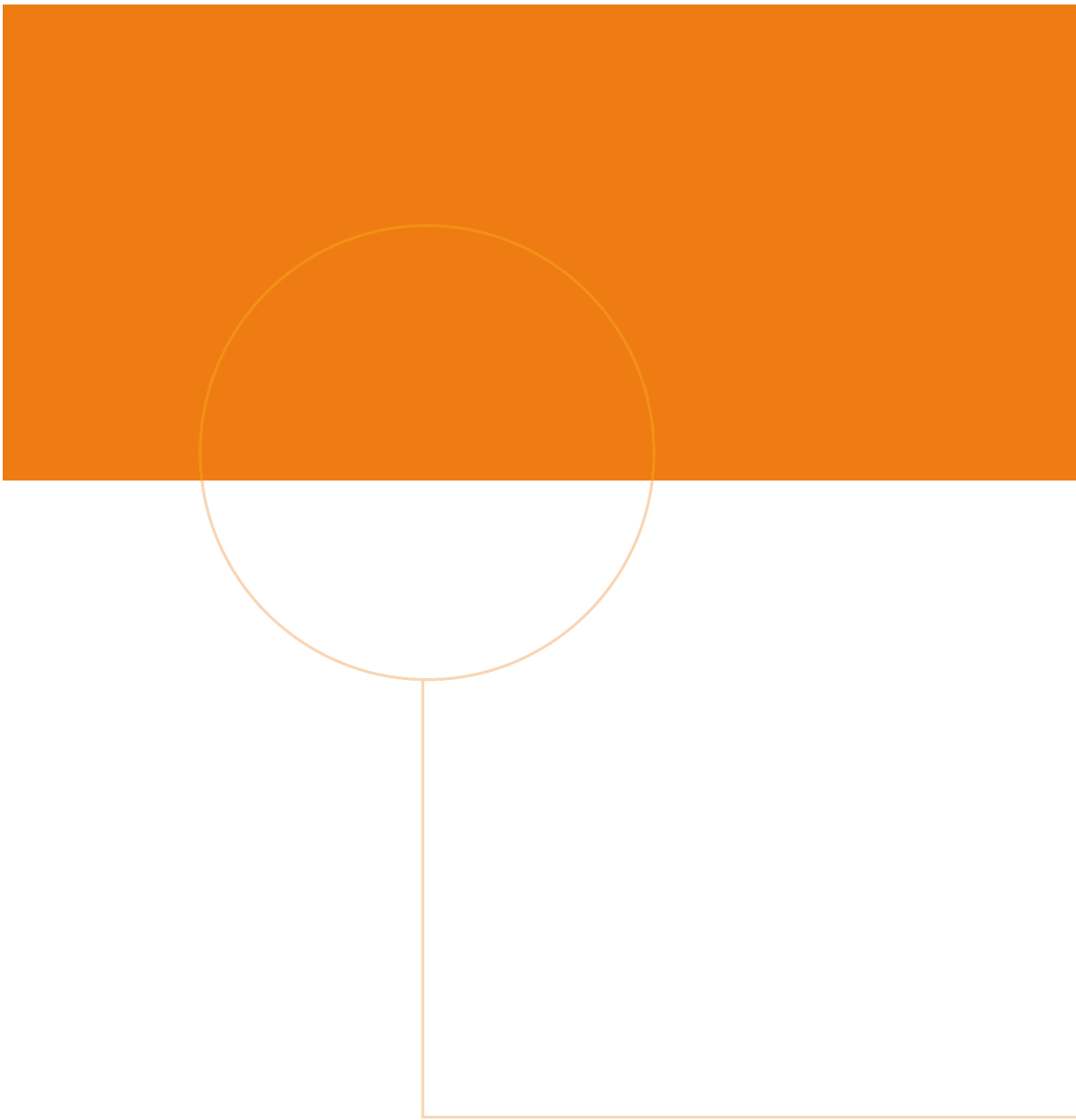
[23] Iver Bakken Sperstad. "Documentation of the OPAL prototype and SAMREL framework implementations". Project Memo no. AN 21.12.09, SINTEF Energy Research. 2021.

[24] Iver Bakken Sperstad et al. "Data set for power system reliability analysis using a four-area test network". In: *Data in Brief* 33 (2020). Article 106495.

[25] Håkon Toftaker, Jørn Foros, and Iver Bakken Sperstad. "Accounting for component condition and preventive retirement in power system reliability analyses". Pre-print version, DOI: 10.36227/techrxiv.19362131.v1. 2022.

[26] Håkon Toftaker and Iver Bakken Sperstad. "Integrating component condition in long-term power system reliability analysis". In: *Proceedings of the 32nd European Safety and Reliability Conference* (2022), pp. 1691–1698.

[27] Yuan Zhao et al. "Composite Power System Relibility Evaluation Based on Enhanced Sequential Cross-Entropy Monte Carlo Simulation". In: *IEEE Transactions on Power Systems* 34 (2019), pp. 3891–3901.



NTNU

Norwegian University of
Science and Technology