

**TECHNIQUES AND APPLICATIONS FOR THE EFFICIENT SIMULATION OF  
GAUSSIAN RARE EVENTS**

A Dissertation  
Presented to  
The Academic Faculty

by

Richard G. Birge

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy in the  
School of Industrial and Systems Engineering

Georgia Institute of Technology

December 2016

Copyright © Richard G. Birge 2016

**TECHNIQUES AND APPLICATIONS FOR THE EFFICIENT SIMULATION OF  
GAUSSIAN RARE EVENTS**

Approved by:

Dr. Antonius Dieker, Advisor  
Industrial Eng. and Op. Research  
*Columbia University*

Dr. Robert Foley  
Industrial and Systems Engineering  
*Georgia Institute of Technology*

Dr. David Goldsman  
Industrial and Systems Engineering  
*Georgia Institute of Technology*

Dr. Enlu Zhou  
Industrial and Systems Engineering  
*Georgia Institute of Technology*

Dr. Heinrich Matzinger  
School of Mathematics  
*Georgia Institute of Technology*

Date Approved: November 11, 2016

## **ACKNOWLEDGEMENTS**

I greatly appreciate everyone who has helped me at Georgia Tech. First and foremost, I would like to thank my advisor, Professor Antonius Dieker, for supporting me for four years and for guiding me along the way. Without him, I would not have grown from a graduate student to an independent researcher. I am also deeply grateful to the graduate community at Georgia Tech and in particular among them, my friends Guido Barrios, Tonghoon Suk, Justin Kirby, Amelia Musselman, and Yuan Li, who have inspired me with their insights and conversation, and been an invaluable part of my time here at Georgia Tech. I would also like to thank my committee members Dr. Robert Foley, Dr. David Goldsman, Dr. Enlu Zhou, and Dr. Heinrich Matzinger, who have also provided valuable input to me over my time in crafting my thesis, from my proposal to its version today.

## TABLE OF CONTENTS

|   |      |
|---|------|
| <b>Acknowledgments</b> . . . . .                                  | iii  |
| <b>List of Tables</b> . . . . .                                   | vii  |
| <b>List of Figures</b> . . . . .                                  | viii |
| <b>Summary</b> . . . . .  | x    |
| <b>Chapter 1: Introduction</b> . . . . .                          | 1    |
| 1.1 Gaussian processes in practice. . . . .                       | 1    |
| 1.2 Theoretical context. . . . .                                  | 3    |
| 1.3 Problem formulation. . . . .                                  | 4    |
| 1.4 Research contributions. . . . .                               | 5    |
| 1.5 Structure of thesis. . . . .                                  | 6    |
| <b>Chapter 2: Mathematical Framework and Motivation</b> . . . . . | 8    |
| 2.1 Preliminaries. . . . .  | 8    |
| 2.1.1 Traditional Monte Carlo. . . . .                            | 8    |
| 2.1.2 Importance Sampling. . . . .                                | 9    |
| 2.1.3 Notions of efficiency. . . . .                              | 11   |

|  |  |           |
|--|--|-----------|
| 2.2  | One-dimensional Gaussian tail probabilities. . . . .   | 14        |
| 2.3  | Two-dimensional Gaussian maximum. . . . .              | 21        |
| <b>Chapter 3: Constrained Gaussian Maximum . . . . .</b> |  | <b>28</b> |
| 3.1  | Overview. . . . .                                      | 28        |
| 3.1.1  | Primary assumptions and definitions. . . . .           | 29        |
| 3.1.2  | Simulation set-up. . . . .                             | 30        |
| 3.2  | Theoretical results. . . . .                           | 32        |
| 3.2.1  | Estimator is unbiased. . . . .                         | 32        |
| 3.3  | Numerical experiments. . . . .                         | 34        |
| 3.3.1  | Overview. . . . .                                      | 34        |
| 3.3.2  | IID Study. . . . .                                     | 34        |
| 3.3.3  | Perfect correlation study. . . . .                     | 38        |
| 3.4  | Appendix. . . . .                                      | 41        |
| <b>Chapter 4: General Gaussian Functionals . . . . .</b> |  | <b>44</b> |
| 4.1  | Overview. . . . .                                      | 44        |
| 4.1.1  | Primary definitions. . . . .                           | 44        |
| 4.1.2  | Simulation set-up. . . . .                             | 45        |
| 4.2  | Theoretical results. . . . .                           | 48        |
| 4.2.1  | Estimator is unbiased. . . . .                         | 48        |
| 4.2.2  | Estimator has desirable asymptotic properties. . . . . | 50        |
| 4.3  | Numerical experiments. . . . .                         | 56        |

|   |   |           |
|---|---|-----------|
| 4.3.1   | Overview. . . . .                           | 56        |
| 4.3.2   | Perfect correlation study. . . . .          | 57        |
| 4.3.3   | Cosine correlation study. . . . .           | 61        |
| 4.4   | Appendix. . . . .                           | 64        |
| <b>Chapter 5: Conditional Functionals . . . . .</b> |   | <b>66</b> |
| 5.1   | Overview. . . . .                           | 66        |
| 5.2   | Applications in finance. . . . .            | 67        |
| 5.2.1   | Primary definitions. . . . .                | 68        |
| 5.2.2   | Simulation set-up. . . . .                  | 70        |
| 5.3   | Theoretical results. . . . .                | 71        |
| 5.3.1   | Numerator of estimator is unbiased. . . . . | 71        |
| 5.4   | Numerical experiments. . . . .              | 74        |
| 5.4.1   | Overview. . . . .                           | 74        |
| 5.4.2   | Perfect correlation study. . . . .          | 75        |
| 5.4.3   | Bivariate study. . . . .                    | 77        |
| <b>Bibliography . . . . .</b>                       |   | <b>83</b> |

## LIST OF TABLES

|     |  |    |
|-----|--|----|
| 3.1 | IID Study: Monte Carlo Estimates vs. True Values . . . . .                           | 35 |
| 3.2 | Perfect Correlation Study: Monte Carlo Estimates vs. True Values . . . . .           | 38 |
| 4.1 | Perfect Correlation Study: Monte Carlo Estimates vs. True Values . . . . .           | 58 |
| 4.2 | Cosine Study: Monte Carlo Estimates vs. True Values . . . . .                        | 62 |
| 5.1 | Perfect Correlation Study: $\mathbb{E}[X_1; \max(X_1, \dots, X_{16}) > b]$ . . . . . | 75 |
| 5.2 | Bivariate Study: $\mathbb{E}[X_1; \max(X_1, X_2) > b]$ . . . . .                     | 77 |
| 5.3 | Covariance Table for Monte Carlo Estimates . . . . .                                 | 78 |

## LIST OF FIGURES

|     |   |    |
|-----|---|----|
| 3.1 | IID Study: Complementary CDF . . . . .  | 36 |
| 3.2 | IID Study: Relative Error vs. Thresholds . . . . .  | 37 |
| 3.3 | Perfect Correlation Study: Complementary CDF . . . . .  | 39 |
| 3.4 | Perfect Correlation Study: Relative Error vs. Thresholds . . . . .  | 40 |
| 4.1 | Illustrations of Assumption (A3), left, and Assumption (A4), right. The darkest colored set is $A_b \cap \{x : \ x - w^j x_j\  < K\}$ . . . . . | 52 |
| 4.2 | Perfect Correlation Study: Complementary CDF . . . . .  | 59 |
| 4.3 | Perfect Correlation Study: Relative Error vs. Thresholds . . . . .  | 60 |
| 4.4 | Cosine Study: Complementary CDF . . . . .   | 62 |
| 4.5 | Cosine Study: Relative Error vs. Thresholds . . . . .   | 63 |
| 5.1 | Perfect Correlation Study: Estimated Relative Error . . . . .   | 76 |
| 5.2 | Bivariate Study, 0.5 Correlation: Estimated Relative Error . . . . .  | 78 |
| 5.3 | Monte Carlo Estimates versus Covariance, Threshold $b = 6$ . . . . .  | 79 |

## SUMMARY

This dissertation focuses on rare-event simulation in the context of Gaussian random processes, which are in widespread use amongst industry practitioners. In particular, Gaussian processes have been used in financial disciplines such as risk management and portfolio valuation, because these processes are computationally tractable and come with an abundance of vendor tools. The present work provides both a theoretical and practical treatment of simulating Gaussian rare events, from algorithm construction and asymptotics, to empirical analysis via computational studies across multiple problem instances. The preliminary chapters of this dissertation provide the mathematical framework and theoretical justification for our methodology, including a case study in a simplified problem setting; this treatment is then extended to the later chapters, wherein we generalize previous theory to broader classes of functionals, and also consider specific applications of our method.

First, in Chapters 1 and 2, we introduce the research topic area, mathematical context and goals. We give further detail in Chapter 1 on our problem motivation, which is the design of an efficient, generalizable Monte Carlo algorithm for simulating exceedance probabilities of rare-event-type functionals, such as  $\max_i X_i$ , for discrete Gaussian random vectors  $X$ . We follow this introductory chapter with a discussion in Chapter 2 of the mathematical foundations for our proposed simulation method, including a background to importance sampling theory and mathematical notions of algorithmic complexity and efficiency, particularly in asymptotic settings. We analyze two small problem instances where standard variance-reduction tools have computational setbacks; we outline the subtleties that make these problems non-trivial, and show their relevance for the alternative methodology explored in this dissertation. To provide additional understanding for how our method operates and the underlying mathematical intuition, in Chapter 3 we give an expository study focusing on a special problem instance of the constrained maximum functional. In this chapter, we detail the construction of an efficient Monte Carlo estimator wherein the un-

derlying Gaussian vector has standard marginals, for which the minimum is also restricted to be non-negative. Through the estimation problem we study in Chapter 3, we provide the key principles and performance insights for our simulation algorithm, which we highlight through several computational studies.

The latter parts of this dissertation, Chapters 4 and 5, extend the earlier theoretical framework to cover a wide class of functions, taken with respect to Gaussian marginals having general variance structures, and also including conditional functionals relevant to financial risk management. In Chapter 4, we give the construction of a general, efficient Monte Carlo estimator to compute exceedance probabilities for rare-event-type functionals of discrete Gaussian vectors, taken with nonstandard marginals, for what we show to be a broad class of functions. The mathematical treatment in Chapter 4 serves to bridge the theory from the special case of the constrained maximum in Chapter 3 to a general abstract problem setting. We provide proofs in Chapter 4 that our general estimator is unbiased and also has desirable asymptotic properties. We substantiate these claims with detailed computational studies. Lastly, in Chapter 5, we present an additional extension to types of functionals relevant to applied settings, for example to assess financial risk. Focusing on topics important to simulation practitioners, particularly the computation of certain risk measures such as expected shortfall or conditional value-at-risk, we describe the usefulness of our method for wide problem applications. We close Chapter 5 with numerical examples that highlight the potential practical uses and flexibility of our simulation method.

# CHAPTER 1

## INTRODUCTION

### 1.1 Gaussian processes in practice.

Our research has focused on rare-event simulation in the context of Gaussian random processes, which are in widespread use amongst broad classes of industry practitioners. In particular, Gaussian processes have enormous influence in financial disciplines such as risk management and portfolio valuation, because these processes are computationally tractable and come with an abundance of vendor tools. By computationally tractability, we refer to the attractive mathematical properties of Gaussian distributions, which allows for fast computation in high-dimensional settings. This property of computational tractability leads to the other property, the abundance of vendor tools available to practitioners. Because Gaussian processes not only have the benefit of extensive academic literature, but also have sophisticated software tools available, marketed through vendors such as RiskMetrics or MSCI Barra, practitioners have the ability to start with these existing tools, and then add their own customizations.

A major application area for Gaussian processes is in financial risk settings, particularly in the computation of risk metrics. For computing risk metrics, a value is assigned to quantify the risk position of a security or portfolio relative to other financial instruments in consideration. A frequent risk measure in finance is known as value-at-risk, or VaR, which can also be considered a measure of tail risk. VaR produces a percentile for a given probability distribution, which has been used to model the value of a financial instrument; a low percentile, for example, would signify a low value which could be expected on a bad trading day.

Because Gaussian processes allow for quick computation of VaR both for single positions and also in portfolio settings, they are often used. The two papers [24] and [19] look at efficient impor-

tance sampling techniques for the simulation of portfolio value-at-risk. The importance sampling tools explored by Glasserman in these works were shown to be asymptotically efficient but were usually model-specific, dependent on the model used for the risk factors in the portfolio, for example a Delta-Gamma approximation. Differences in the results obtained by various risk measures, depending on the underlying portfolio model, are also explored in [46] and [33].

More advanced risk metrics may also be considered beyond the traditional measures such as value-at-risk. For these metrics, additional criteria are considered, such as the size of tail losses or moment properties. Recent works that focus on the simulation of portfolio value-at-risk consider cases in which model pitfalls may be avoided through a robust optimization problem. This problem is known as robust risk measurement and is explored in [26], as well as in more general papers in mathematical finance, with notable examples being [35] and [34]. Alongside those risk measures just mentioned, such as value-at-risk and conditional value-at-risk, examples of alternative criteria include metrics such as worst-case risk evaluation and maximum-drawdown.

Beyond specific applications in traditional risk measures, which produce a value to quantify the risk of a position or portfolio, another related area is the concept of default or credit risk. In this field, the motivational question is the probability of a default or loss event. In the default scenario, the notion of default may be considered similar to the notion of survival in reliability estimation.

As with the computation of portfolio risk measures for value-at-risk, Gaussian processes frequently arise in the study of the rare default probabilities, just mentioned, and in this setting known as credit risk. Here the problem focus is on modeling the chance of survival of the single security or portfolio, until some default event, based on dependence to an underlying system of risk factors. Typically, a Gaussian framework is used to model the risk factors, with the security or portfolio value fluctuating in according to the volatility experienced through the underlying risk factors. In traditional measurement of credit risk, a default event may be defined by a given lower-bound value, but more complicated definitions of default may also be considered.

## 1.2 Theoretical context.

Previous work in rare event simulation for Gaussian processes has focused on the construction of Monte Carlo estimators with low variance. In most instances, the asymptotic variance of the Monte Carlo estimator is the primary criterion of judging efficiency. Notions of asymptotic efficiency were first outlined by Glynn and Whitt in [28].

Within the context of simulating rare events for Gaussian processes, no study to date has achieved the construction of an estimator which is strongly efficient, that is, having bounded relative error, in all problem settings, although progress has been made under certain restrictive conditions. Significant advancements in the field were made by Adler et al. [1], which was a study of efficient estimators of suprema of Gaussian random fields. This work by Adler et al. yielded simulation algorithms which the authors showed to be polynomially efficient in the general case, but only strongly efficient under certain special conditions. Notably, the technique proposed by the authors adopted an importance-sampling density which could also be applicable to conditional estimators and Gaussian functions. For example, the paper [42] explores expanding the change-of-measure technique first adopted by Adler, Blanchet and Liu in [1] for the theoretical problem of simulating exponential Gaussian integrals.

In addition to this work by Adler et al., another importance impetus for the present research is a change-of-measure argument first advanced by Dieker and Yakir [12] in their study of simulation methods for Pickands' constant. In this paper, the authors developed a change-of-measure technique involving a quotient of exponentials which forms the general theory for the importance sampling method followed in this work. We give a more detailed perspective on the linkage between this paper and this thesis in the appendix of Chapter 3.

### 1.3 Problem formulation.

Given an arbitrary,  $n$ -dimensional Gaussian random vector  $X$ , our research motivation has been to devise an efficient simulation algorithm for computing the probability that a rare-event-type functional, such as  $\max_i X_i$ , exceeds a certain level-crossing parameter  $b$ . In the first part of this work, we focus in the special case of a constrained max functional, with standard Gaussian marginals, wherein we introduce and explain the pedagogical motivation behind our methodology in a simple setting. In the latter parts, as a first extension, we generalize the problem to more broad problem classes, such as general max functionals, or related rare-event functionals, taken over nonstandard Gaussian marginals. From there, we consider further extensions, in particular, conditional functionals such as conditional tail probabilities, and explore the usage of these functionals across widespread industry applications, focusing on financial risk management.

For our purposes here, we are focused on only Gaussian processes, and specifically in a discrete setting. For the general problem formulation, we define  $X$  to be an  $n$ -dimensional, Gaussian random vector, whose marginal components may have arbitrary drift and variance. We denote the  $i$ -th element of the vector  $X$  by  $X_i$  for  $i = 1, \dots, n$ .

This probability, as described, may be written mathematically, given a general vector-valued function  $g$ , and an exceedance threshold  $b$ , as:

$$\mathbb{P} [g(X) > b].$$

We may also consider the related problem, which is of particular interest for applications, given for another general functional  $f$  by

$$\mathbb{E} [f(X); g(X) > b].$$

Here, we assume the parameter  $b$  to be large, so that traditional Monte Carlo simulation would

not be practical. The two properties which we intend to study for our estimator are first, its mean, which to be unbiased must have the same expectation as the original probability, and second, its variance, an important criterion for algorithmic efficiency.

The simulation methodology followed in this research is Monte Carlo using importance sampling, whereby a change-of-measure is adopted to improve the efficiency of the Monte Carlo estimator. We note here that a sample realization of an (unbiased) estimator  $\Theta_b$  from a given simulation replication  $k$  may be denoted  $\Theta_b^{(k)}$ . After all simulation replications are finished, and assuming  $m$  replications in total, our final estimate for  $\mathbb{E}\Theta_b$  would be calculated according to the sample average, with representation

$$\bar{\Theta}_b = \frac{1}{m} \sum_{k=1}^m \Theta_b^{(k)}.$$

The efficiency goal, and motivation for importance sampling, in our context, is to construct a Monte Carlo estimator  $\Theta_b$  so that  $\Theta_b/\mathbb{E}\Theta_b$  is asymptotically independent of the exceedance parameter  $b$ . Hence, for this setting, the coefficient of variation would remain uniformly bounded as the exceedance parameter, also known as a threshold parameter or level-crossing barrier, increases.

## 1.4 Research contributions.

### Flexibility and impact of methodology for risk applications.

We believe our research has considerable practical significance because of the flexibility our method allows across multiple problem settings, with extensions to finance, insurance, and risk disciplines. This flexibility arises from the ability to use our work to construct estimators for exceedance probabilities related to general Gaussian functionals, which can be interpreted according to various financial risk metrics, as well as portfolio analytic tools. In particular, our technique allows for the efficient simulation of Gaussian tail probabilities which are used in risk measures popular in finance, such as value-at-risk, known as VaR, as well as conditional value-at-risk, CVaR, also known as expected shortfall. Additionally, our work also, we believe, allows for the fast computation of

portfolio credit analytics, that rely on assessing probability of default according to default scenarios calibrated based on Gaussian default intensities. Seen from this perspective, our estimator allows for the calculation of default probabilities, where the level-crossing parameter from our original problem may be interpreted as a default barrier.

### Efficient, generalizable Monte Carlo tools.

Our research allows, first, for the development of a new change-of-measure technique and importance sampling methodology, and second, and most importantly, it is noteworthy for the general flexibility across general Gaussian functionals. Furthermore, our work not only allows for the construction of estimators with attractive efficiency properties, for the study of exceedance probabilities for Gaussian maxima, but also, as we shall show, allows for the construction of estimators related to general Gaussian functionals.

Because our methodology is not limited to a specific functional, it has the important benefit of being related to broad problem classes. Moreover, this flexibility of our method has the benefit of allowing a single underlying simulation methodology which utilizes a common importance sampling approach. As mentioned, this ease of generalization has not only important theoretical considerations, but practical benefits as well, by allowing practitioners to have a unifying simulation framework in risk or portfolio calculations they may need to perform.

## **1.5 Structure of thesis.**

In Chapter 2, Mathematical Preliminaries, we provide the theoretical foundations for the present work, including a background to importance sampling theory and mathematical notions of efficiency. In Chapter 3, we present an expository study to our more general methodology, by focusing on a special problem case of the constrained maximum functional, assuming standard Gaussian marginals. In Chapter 4, we extend this framework to cover quite general classes of functionals, as well as non-standard Gaussian marginals including general variance structures. In the final chap-

ter, we present an additional extension to conditional functionals and explore areas of relevance for important application topics, such as value-at-risk and loss-given-default, across insurance and financial risk management industries.

## CHAPTER 2

### MATHEMATICAL FRAMEWORK AND MOTIVATION

#### Overview.

In this chapter, we provide a mathematical overview of the method of Monte Carlo simulation. In particular, we discuss the variance-reduction technique of importance sampling, and additionally, we show how standard variance-reduction tools for the general problem setting discussed in this thesis are not easily applied. To support this assertion, as a case study, we analyze two small problem instances, and for each instance, compare and contrast the relative (in)efficiencies of the standard tools. The limitation posed by these standard tools, as we show, is what motivates our search for an alternative method with superior algorithmic performance.

#### 2.1 Preliminaries.

##### 2.1.1 Traditional Monte Carlo.

To begin, we assume an event  $A$  defined on a probability space  $(\Omega, \mathcal{F}, P)$  and unknown quantity of interest  $z \equiv \mathbb{P}[A]$ . Standard Monte Carlo estimates  $z$  by repeatedly sampling from the original probability distribution  $\mathbb{P}$ . A typical estimator would be:

$$Z = \mathbf{1}(A),$$

and hence,  $\mathbb{E}_P[Z] = \mathbb{P}[A]$ . When  $A$  is a rare event (e.g.  $\mathbb{P}[A] < 10^{-3}$ ), computational cost is high for achieving reasonable levels of accuracy.

In addition to estimating probabilities, standard Monte Carlo may also be used to compute functions of random variables. For example, for  $Y$  a  $\mathcal{F}$ -measurable random variable, we may wish

to estimate  $\mathbb{E} h(Y)$ . In this case, the Monte Carlo estimator  $T$  could be derived as

$$T(\omega) = \sum_{\omega \in \Omega} h(Y(\omega)) \mathbf{1}(\omega).$$

The high cost of traditional Monte Carlo in these cases prompts the usage of alternative techniques, among them, importance sampling, which we discuss next.

### 2.1.2 Importance Sampling.

Importance sampling is a variance-reduction technique, whereby random samples are drawn from an alternative probability measure which differs from the natural (or original) measure. We assume here an unbiased Monte Carlo estimator  $Z$  under the original measure, to be denoted  $P$ , such that  $z = \mathbb{E}_P Z$ . Informally, the goal is to sample under a new measure  $Q$ , rather than the original measure  $P$ , so that more mass is placed in that region of the state space which contributes the most to the quantity being studied.

Furthermore, importance sampling is a technique which applies not just to single random variables but also to (continuous) functionals of those variables. Because random variable functionals are the main interest in this thesis, we will emphasize the benefits of importance sampling from that perspective.

The insight for importance sampling utilizes what is known as a likelihood ratio to modify the sampling procedure from measure  $P$  to  $Q$ . We will assume here continuous probability densities; specifically, the random variable  $Z$  admits a probability density under  $P$  ( $p_Z$ ), and also under  $Q$  ( $q_Z$ ), though the framework holds in more general settings as well. We assume one measure-space  $(\Omega, \mathcal{F})$  and that  $Z$  is  $\mathcal{F}$ -measurable. Given  $h(\cdot)$  a continuous functional over the support of  $Z$ , we have the integral representation:

$$\int_{\Omega} h(\omega) p_Z(\omega) d\omega = \int_{\Omega} h(\omega) \left( \frac{p_Z(\omega)}{q_Z(\omega)} \right) q_Z(\omega) d\omega.$$

The above relation may also be written using the expectation operator. Letting  $\mathbb{E}_P$  represent an expectation taken under measure  $P$  and similarly noted for measure  $Q$ , we have

$$\mathbb{E}_P [h(Z)] = \mathbb{E}_Q \left[ h(Z) \frac{p(Z)}{q(Z)} \right].$$

In the above instance, the Radon-Nikodym derivative  $L$  is given by the quotient of the probability densities, that is  $L(Z) := p(Z)/q(Z)$ . This quantity is also known as the likelihood ratio.

Using the likelihood ratio, we can relate the expectation under the original measure to the expectation from the importance sampling measure, shown as follows:

$$\mathbb{E}_P [h(Z)] = \mathbb{E}_Q [h(Z)L(Z)]. \tag{2.1}$$

More formally, for any given random variable  $Z$  which admits a Radon-Nikodym derivative  $L$ , we have that:

$$P(d\omega) = L(\omega) Q(d\omega).$$

It follows under the above relation that  $z = \mathbb{E}_P Z = \mathbb{E}_Q ZL$ . Thus, the importance sampling method for Monte Carlo is to simulate independent and identically-distributed (IID) replicates of  $ZL$  drawn from measure  $Q$ , rather than replicates of  $Z$  drawn from  $P$ .

One important category or family of importance sampling densities, which we will explore shortly, is known as *exponential twisting*. Because the technique of exponential twisting is widely used by practitioners, and also factors heavily into our discussion of alternative sampling procedures, we provide a formal definition here.

Given a random variable  $Z$  and measure-space  $(\Omega, \mathcal{F})$  as defined previously, with  $P$  the original probability measure, the exponential twisting density is constructed as follows. Using a real parameter, let us call it  $\theta$ , which is also known as the “twist”, the new exponential twisting density,

which we may denote  $Q_\theta$ , has the form,

$$Q_\theta(d\omega) := \frac{e^{\theta\omega}}{\mathbb{E}_P[e^{\theta Z}]} P(d\omega). \quad (2.2)$$

It can be readily verified that the new density  $Q_\theta$  will integrate to 1 over the sample space  $\Omega$ , as required, given the division by the moment-generating function in the denominator. The likelihood ratio  $L_\theta(\omega)$  is given by the reciprocal to this quantity, which is

$$L_\theta(\omega) := e^{-\theta\omega} \mathbb{E}_P[e^{\theta Z}].$$

It should be noted, that although the mechanics of exponential twisting appear straightforward, the choice of the parameter  $\theta$  plays quite an important role. The twisting parameter  $\theta$  is, in fact, critical to the evaluation of the quality of the estimator, due to its influence on the sampling variance.

As has been stated, a chief goal of any importance sampling procedure is variance reduction, because of the computational savings. Variance-reduction is achieved provided that

$$\text{Var}_Q ZL < \text{Var}_P Z. \quad (2.3)$$

Any probability measure  $Q$  with a likelihood ratio  $L$  with respect to  $P$  which lowers the sampling variance is a candidate for importance sampling.

However, while multiple choices may be available for the importance sampling density, not all measures may be optimal from a computational perspective. Indeed, the degree of variance reduction is closely related to the concept of algorithmic efficiency, which we discuss next.

### 2.1.3 Notions of efficiency.

A significant criterion for evaluating the quality of a Monte Carlo estimator is the notion of relative error. By relative error, we properly refer to the statistical term, root mean squared error, or coef-

ficient of variation. (Relative error is the more typical term in simulation literature when referring to efficiency criterion of algorithms, as discussed in such studies as [43] and [45].)

Definition: Relative error.

Let  $Z$  be an unbiased estimator for the quantity  $z(b)$ , given a threshold (e.g., exceedance or level-crossing) parameter  $b$ , with  $\mathbb{E}Z = z(b)$ . Then we define the relative error, to be denoted  $e_Z(b)$ , and it holds that

$$e_Z(b) := \frac{\sqrt{\text{Var}Z(b)}}{z(b)}. \quad (2.4)$$

Two of the most relevant criteria for measuring the efficiency of a Monte Carlo algorithm are *strong efficiency* and *logarithmic efficiency*, with logarithmic efficiency being the weaker of the two notions. These two notions are defined below, and can also be related to the concept of relative error, given above in (2.4) and additionally shown in [43].

Definition: Logarithmic (log) efficiency.

An estimator  $Z$  is said to be *logarithmically efficient* (also known as “weakly efficient” or “asymptotically efficient”) in estimating  $z(b)$  if  $\mathbb{E}Z = z(b)$  and it holds that:

$$\liminf_{b \rightarrow \infty} \frac{|\log \text{Var}Z(b)|}{|\log z(b)^2|} \geq 1,$$

or equivalently

$$\limsup_{b \rightarrow \infty} z(b)^\epsilon e_Z(b) < \infty \quad \text{for all } \epsilon > 0.$$

Definition: Strong efficiency.

An estimator  $Z$  is said to be *strongly efficient* (or have *bounded relative error*) in estimating  $z(b)$  if  $\mathbb{E}Z = z(b)$  and it holds that

$$\limsup_{b \rightarrow \infty} \frac{\mathbb{V}\text{ar}Z(b)}{z(b)^2} < \infty,$$

or equivalently

$$\limsup_{b \rightarrow \infty} e_Z(b) < \infty.$$

Thus, in the notion of relative error, strong efficiency can be understood as the property of having uniformly bounded relative error.

Because it is relevant to the present study, we also introduce here the special (stronger) case within the class of strong efficiency, which occurs for an estimator with relative error that asymptotically approaches zero. This notion is called *vanishing relative error* and is defined as given below.

Definition: Vanishing relative error.

Let  $Z$  be an unbiased estimator for the quantity  $z(b)$ , given a threshold (e.g., exceedance parameter)  $b$ , with  $\mathbb{E}Z = z(b)$ . Then we define vanishing relative error by,

$$\lim_{b \rightarrow \infty} e_Z(b) = 0. \tag{2.5}$$

## 2.2 One-dimensional Gaussian tail probabilities.

In this section we work out the techniques for the estimation of single-dimension Gaussian tail-probabilities under three different methods, which we show to have varying efficiency properties. We will later contrast these efficiency results with what is available in two-dimensions, to show that the one-dimension results cannot be generally extended to higher dimensions or more general settings.

Here, we assume  $X$  follows a standard normal distribution. We let  $\varphi(\cdot)$  denote the standard normal density. We wish to estimate the upper-tail probability given a threshold value  $b$ , where

$$z := \mathbb{P}(X > b) = \int_b^\infty \varphi(x) dx = \int_b^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx,$$

for  $b$  a large, positive real number (e.g.  $b \gg 3$ ).

In the paragraphs below, we describe the three alternative procedures we will examine for estimating Gaussian tail probabilities:

1. Numerical evaluation of tail integral.
2. Importance sampling via exponential twisting.
3. Importance sampling via the exponential density.

Moreover, we compare the quality of the estimator  $Z$  obtained under each method by analyzing the relative error asymptotics, given by  $e_Z(b)$  for  $b \rightarrow \infty$ . The asymptotic behavior of the relative error for each of these estimators is well-documented; the proofs below are provided to aid the reader. (The details of each proof are our own, although the results are otherwise established in existing literature.)

Numerical integral approximation.

**Proposition 2.2.1.** *The numerical approximation method for the Gaussian upper-tail integral, with estimator to be denoted  $Z_\Psi$ , given*

$$Z_\Psi := \Psi(b) = \int_b^\infty \varphi(x) dx,$$

*and with  $\varphi(\cdot)$  denoting the standard normal density, then  $Z_\Psi$  has the relative error property:*

$$e_{Z_\Psi}(b) = 0 \quad \text{for all } b > 0,$$

*where the integral calculation is done numerically.*

*Proof.* Here we assume an exact numeric approximation for the tail integral is available, and because the calculation is considered exact, there is zero relative error. □

Importance sampling: exponential twisting.

**Proposition 2.2.2.** *The importance sampling estimator via exponential twisting, to be denoted  $Z_{IT}$ , with twist parameter  $b$ , as defined by*

$$Z_{IT} := e^{-bX + b^2/2} \mathbf{1}(X > b),$$

*has relative error which grows like  $\sqrt{b}$ :*

$$e_{Z_{IT}}(b) \sim \left(\frac{\pi}{2}\right)^{\frac{1}{4}} \sqrt{b}, \quad b \rightarrow \infty.$$

*Proof.* Here, the intuition is to shift the probability density under the original measure  $P$ , which we may denote by the function  $p(\cdot)$ , to a new density, some function  $q_b(\cdot)$ , under a new probability

measure,  $Q_b$ , centered at the exceedance threshold  $b$ :

$$q_b(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-b)^2/2}.$$

The threshold  $b$ , in fact, also corresponds to a choice of exponential twisting parameter.

Using the set-up for the exponential twist defined in (2.2), then for a twist  $\theta = b$ , the likelihood ratio,  $L$ , is given by

$$\frac{dP(x)}{dQ_b(x)} = e^{-bx+b^2/2}.$$

Taking  $Z = \mathbf{1}(X > b)$  to be the standard Monte Carlo estimator under  $P$ , the new estimator under  $Q_b$  will be

$$Z_{IT} := ZL = e^{-bX+b^2/2} \mathbf{1}(X > b).$$

Recalling that the estimator  $Z_{IT} = ZL$  is unbiased, with  $\mathbb{E}_{Q_b}[Z_{IT}] = \mathbb{P}[X > b] =: z$ , we may compute the squared relative error for  $Z_{IT}$  by the ratio,

$$\begin{aligned} \frac{\text{Var}_{Q_b}[ZL]}{[\mathbb{P}(X > b)]^2} &= \frac{(\mathbb{E}_{Q_b}[(ZL)^2] - (\mathbb{E}_{Q_b}[ZL])^2)}{[\mathbb{P}(X > b)]^2} \\ &= \frac{\mathbb{E}_{Q_b}[(ZL)^2]}{z^2} - 1 \\ &= [e_Z(b)]^2. \end{aligned}$$

To aid with the asymptotic analysis, we make use of the identity known as Mills' ratio. This result states that given a positive threshold  $b$ , with  $\varphi(\cdot)$  the standard normal density, then as  $b$  grows, the upper-tail integral has the behavior,

$$\int_b^\infty \varphi(x) \sim \frac{\varphi(b)}{b}, \quad b \rightarrow \infty. \quad (2.6)$$

We may derive by the usual calculus that  $\mathbb{E}_{Q_b} [(ZL)^2]$  has the following tail-integral form,

$$\mathbb{E}_{Q_b} [(ZL)^2] = e^{b^2} \int_{2b}^{\infty} \varphi(x) dx.$$

Thus, using the Mills' ratio asymptotic result given in (2.6), we find that

$$\mathbb{E}_{Q_b} [(ZL)^2] \sim e^{b^2} \frac{\varphi(2b)}{2b}, \quad b \rightarrow \infty.$$

And so, again by using Mills' ratio, we have the relation:

$$\frac{\mathbb{E}_{Q_b} [(Z_{IT}L)^2]}{[\mathbb{E}_{Q_b} (Z_{IT})]^2} \sim \frac{e^{b^2} (\varphi(2b)/2b)}{(\varphi(b)/b)^2}, \quad b \rightarrow \infty.$$

As can be readily verified from the above steps, the relative error  $e_{Z_{IT}}(b)$  behaves asymptotically for  $b \rightarrow \infty$  as

$$\sqrt{\frac{e^{b^2} (\varphi(2b)/2b)}{(\varphi(b)/b)^2}} = \left(\frac{\pi}{2}\right)^{1/4} \sqrt{b},$$

which grows polynomially with  $b$ . Although this estimator does meet the criterion for logarithmic efficiency, it will not be strongly efficient, since the relative error grows without bound as the threshold  $b$  approaches infinity.  $\square$

Importance sampling: exponential density.

**Proposition 2.2.3.** *The importance sampling estimator via the exponential density, denoted  $Z_{IE}$ , which given a standard exponential variable  $Y$ , has the construction*

$$Z_{IE}(b) := \frac{e^{-b^2/2}}{b\sqrt{2\pi}} e^{-Y^2/(2b^2)},$$

then  $Z_{IE}$  has vanishing relative error,

$$e_{Z_{IE}}(b) \rightarrow 0, \quad b \rightarrow \infty.$$

*Proof.* The estimator, denoted  $Z_{IE}(b)$ , is given in the general case where  $Y$  has the standard exponential distribution, by

$$Z_{IE}(b) := \frac{e^{-b^2/2}}{b\sqrt{2\pi}} e^{-Y^2/(2b^2)}.$$

To see how this estimator is derived, then, starting from the problem  $\mathbb{P}[X > b]$ , where  $X$  is standard normal, we have

$$\begin{aligned} \int_b^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx &= \int_0^\infty \frac{e^{-(x+b)^2/2}}{\sqrt{2\pi}} dx \\ &= \frac{e^{-b^2/2}}{\sqrt{2\pi}} \int_0^\infty e^{-x^2/2} e^{-xb} dx \\ &= \frac{e^{-b^2/2}}{b\sqrt{2\pi}} \int_0^\infty e^{-y^2/(2b^2)} e^{-y} dy. \end{aligned}$$

Alternatively, we can also assume a general parameter  $\lambda$  which we leave open, along with a standard exponential variable  $\Lambda$ . Done in this manner, the derivation is as follows.

$$\begin{aligned} \int_b^\infty \phi(x) dx &= \int_0^\infty \phi(x+b) dx \\ &= \frac{1}{\lambda} \int_0^\infty \lambda e^{-\lambda x} e^{\lambda x} \phi(x+b) dx \\ &= \frac{1}{\lambda} \mathbb{E} \left[ e^{\Lambda} \phi \left( \frac{\Lambda}{\lambda} + b \right) \right]. \end{aligned}$$

To see that  $Z_{IE}$  has vanishing relative error, we look at the ratio of  $\mathbb{E}(Z_{IE})^2$  to the square of the first moment. Because by the unbiased property  $\mathbb{E}Z_{IE} = \mathbb{P}[X > b]$ , this latter quantity is  $\mathbb{P}[X > b]^2$ .

Using the above calculations, therefore, we have that:

$$\begin{aligned}
& \mathbb{E}(Z_{IE})^2 / [\mathbb{E}(Z_{IE})]^2 \\
&= \frac{\frac{1}{\lambda^2} \mathbb{E} [e^{2\lambda} \phi(\Lambda/\lambda + b)^2]}{\left[ \frac{1}{\lambda} \mathbb{E} [e^{\Lambda} \phi(\Lambda/\lambda + b)] \right]^2} \\
&= \frac{\frac{1}{\lambda^2} \frac{e^{-b^2}}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp(2y(1 - b/\lambda) - y) \exp(-y^2/\lambda^2) dy}{\frac{1}{\lambda^2} e^{-b^2} \left( \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp(y(1 - b/\lambda) - y) \exp(-y^2/(2\lambda^2)) dy \right)^2} \\
&= \frac{\exp((\lambda^2(1 - b/\lambda - 1/2))^2/\lambda^2) \int_0^\infty \left( \frac{1}{\sqrt{2\pi}} \right)^2 \exp(-(y - \lambda^2(1 - b/\lambda - 1/2))^2/\lambda^2) dy}{\left( \exp((\lambda^2(1 - b/\lambda - 1))^2/(2\lambda^2)) \right)^2 \left( \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp(-(y - \lambda^2(1 - b/\lambda - 1))^2/(2\lambda^2)) dy \right)^2} \\
&= \frac{\exp(\lambda^2(b/\lambda)(b/\lambda - 1) + \lambda^2/4) \int_{-\lambda^2(1-b/\lambda)+\lambda^2/2}^\infty \exp(-z^2/\lambda^2) dz}{\exp(\lambda^2(b^2/\lambda^2)) \left( \int_{-\lambda^2(1-b/\lambda)+\lambda^2}^\infty \exp(-z^2/(2\lambda^2)) dz \right)^2} \\
&= e^{\lambda^2/4 - b\lambda} \frac{\int_{b\lambda - \lambda^2/2}^\infty \exp(-z^2/\lambda^2) dz}{\left( \int_{b\lambda}^\infty \exp(-z^2/(2\lambda^2)) dz \right)^2}.
\end{aligned}$$

From here, we use the asymptotic approximation for  $\xi \rightarrow \infty$  that

$$\int_\xi^\infty e^{-y^2/2} dy \sim \frac{1}{\xi} e^{-\xi^2/2}.$$

This asymptotic approximation gives us for  $b \rightarrow \infty$  that

$$\begin{aligned}
& \mathbb{E}(Z_{IE})^2 / [\mathbb{E}(Z_{IE})]^2 \\
&\sim \exp(\lambda^2/4 - b\lambda) \frac{\frac{1}{2(b\lambda - \lambda^2/2)/\lambda^2} \exp(-(b\lambda - \lambda^2/2)^2/\lambda^2)}{\left( \frac{1}{2b\lambda/(2\lambda^2)} \exp(-b^2\lambda^2/(2\lambda^2)) \right)^2} \\
&= \exp(\lambda^2/4 - b\lambda - \lambda^2/4 + b\lambda) \frac{b^2/\lambda^2}{2(b/\lambda - 1/2)} \\
&= \frac{(b/\lambda)^2}{2b/\lambda - 1}.
\end{aligned}$$

For the above asymptotics to hold, we require  $b\lambda \rightarrow \infty$  in the lower limit for the integral in the

denominator, and in the numerator, we also require  $b\lambda - \lambda^2/2 \rightarrow \infty$ . Together these conditions imply  $\lambda < 2b$  and also that  $\lambda \gg 1/b$ .

The remaining terms give us a minimization which we may write as

$$\min_{\lambda: 0 < \lambda < 2b} \frac{b^2/\lambda^2}{2b/\lambda - 1}. \quad (2.7)$$

We may check that the first derivative with respect to  $\lambda$  for the above yields

$$\frac{d}{d\lambda} \left( \frac{b^2/\lambda^2}{2b/\lambda - 1} \right) = \frac{2b^2(\lambda - b)}{\lambda^2(\lambda - 2b)^2}.$$

Over the range  $\lambda \in (0, 2b)$  the minimum is attained at  $\lambda = b$  for which the above derivative is zero, and for which the moment ratio equals 1.

Separately, we may also verify the condition  $\lambda < 2b$ , since for any  $\lambda > 2b$ , the scaling factor would be of the form  $\exp\{\lambda^2/4 - b\lambda + b^2\}$ , which would tend to infinity for  $\lambda > 2b$  with  $b \rightarrow \infty$ . □

### 2.3 Two-dimensional Gaussian maximum.

In this section, similar to the prior section, we work out alternative techniques for the estimation of Gaussian tail-probabilities using standard tools, but now in a two-dimensional setting, which turns out to pose mathematical hurdles.

As we show here, the standard tools from the prior section break down and deteriorate, when measured in terms of efficiency. These limitations, particular as they relate to the ability to handle quite general problem settings, motivate the need for an alternative, more flexible, and algorithmically superior procedure, which is the subject of this thesis.

For the two-dimensional example setting, we take the problem set-up to be that  $X$  is a bivariate Gaussian vector, i.e.  $X \sim \mathcal{N}(0, \Sigma_2)$ , where we have the functional  $g$  given by,

$$g(x) = \max(x_1, x_2),$$

and a covariance matrix  $\Sigma_2$ , which is defined for  $\rho > 0$  as:

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

We further note that the above instance is a special case of the problem  $\mathbb{P}[\max_{i=1, \dots, n} X_i > b]$ , which is one of our chief research interests.

For the cases to follow, in two instances, which are the naive (standard) Monte Carlo estimator, and the exponential-twisting estimator, the relative error asymptotics have previously been established in existing literature. (As in the prior section, proofs are provided for the understanding of the reader, with the proof details being our own.)

However, in one instance, for what we call the “integrated” estimator, the proof of the relative error asymptotic behavior has not otherwise been established to our knowledge, and is a new result.

### Standard estimator.

The “standard” Monte Carlo estimator may be derived as:

$$Z_b^{\text{MC}} := \mathbf{1} \{ \max(X_1, X_2) > b \}.$$

However, it is readily verified that this estimator is *not* computationally efficient (in any mathematical sense), since

$$\mathbb{E}(Z_b^{\text{MC}})^2 \gg (\mathbb{E}Z_b^{\text{MC}})^2, \quad b \rightarrow \infty.$$

### Tail integral or integrated estimator.

Under this approach, first brought to our attention by J. Blanchet, the idea is to integrate in one dimension, and then sample in the others. For this technique, it is helpful to define:

$$X^{(1)} = X - w^1 X_1, \quad w^1 := (1, \rho),$$

$$\xi_b^1(x) = \min(b - x_1, (b - x_2)/\rho),$$

given a general input vector  $x := (x_1, x_2)$ . The vector  $X^{(1)}$  has the first coordinate equal to zero.

To see this, note that  $w_1^1 = 1$  hence  $X_1^{(1)} = X_1 - X_1 = 0$ .

The new estimator as constructed we call the “integrated” estimator, to be abbreviated “IgS”, and we define it as

$$Z_b^{\text{IgS}} := \Psi(\xi_b^1(X^{(1)})), \tag{2.8}$$

where  $\Psi(b) = 1 - \Phi(b)$  denotes the standard Gaussian upper-tail probability for a point  $b$ , given as one less the cdf  $\Phi(\cdot)$  evaluated at  $b$ . This estimator has the construction, which also shows it to be

unbiased, via the derivation,

$$\begin{aligned}
\mathbb{P}[\max(X_1, X_2) > b] &= \mathbb{E} [\mathbb{P}[\max(X_1, X_2) > b \mid X^{(1)}]] \\
&= \mathbb{E} [\mathbf{1}(X_1 > \xi_b^1(X^{(1)}))] \\
&= \mathbb{E} [\Psi(\xi_b^1(X^{(1)}))] .
\end{aligned}$$

To aid with the efficiency analysis, we present the asymptotic result as a proposition first without proof. The full proof of the proposition follows after the explanatory text and required lemmas found next below.

**Proposition 2.3.1.** *The integrated estimator  $Z_b^{IgS}$ , with form*

$$Z_b^{IgS} := \Psi(\xi_b^1(X^{(1)}),$$

*given  $\Psi(\cdot)$  a standard Gaussian upper-integral, has relative error which has exponential form for  $c > 0$ , behaving according to the form*

$$\log e(b) \sim cb^2, \quad b \rightarrow \infty,$$

*for  $\rho \in [0, 1]$ .*

As a preliminary insight into the proposition, we note that

$$\xi_b^1(X^{(1)}) = \min(b, (b - X_2^{(1)})/\rho).$$

Therefore, by evaluating the variable  $X_2^{(1)}$  under the disjoint events  $\{X_2^{(1)} \leq b(1-\rho)\}$  and  $\{X_2^{(1)} >$

$b(1 - \rho)\}$ , we have that  $\xi_b^1(X^{(1)})$  has the following form:

$$\Psi(\xi_b^1(X^{(1)})) = \begin{cases} \Psi(b) & X_2^{(1)} \leq b(1 - \rho) \\ \Psi((b - X_2^{(1)})/\rho) & X_2^{(1)} > b(1 - \rho). \end{cases}$$

Since  $(b - X_2^{(1)})/\rho$  has a normal distribution with mean  $b/\rho$  and variance  $(1 - \rho^2)/\rho^2$ , it also follows that

$$\begin{aligned} & \mathbb{E} \left[ \Psi(\xi_b^1(X^{(1)}))^\alpha \right] \\ &= \Psi(b)^\alpha P(X_2^{(1)} \leq b(1 - \rho)) + \mathbb{E} \left[ \Psi\left(\frac{b - X_2^{(1)}}{\rho}\right)^\alpha ; X_2^{(1)} > b(1 - \rho) \right] \\ &= \Psi(b)^\alpha P(Y_b \geq b) + \mathbb{E}[\Psi(Y_b)^\alpha ; Y_b < b], \end{aligned}$$

where  $Y_b \sim \mathcal{N}(b/\rho, (1 - \rho^2)/\rho^2)$ .

To show the inefficiency of the ‘‘integrated’’ estimator, we bring in support the following lemma.

**Lemma 2.3.1.** *Suppose  $Y_b \sim \mathcal{N}(b/\rho, (1 - \rho^2)/\rho^2)$  for some  $\rho \in (0, 1)$ . For  $\alpha \geq 2$ , we have*

$$\lim_{b \rightarrow \infty} \frac{1}{b^2} \log \mathbb{E} \left[ \exp(-\alpha Y_b^2/2) ; Y_b < b \right] > -\alpha/2.$$

*Proof.* Noting that  $Y_b$  has the same distribution as  $b/\rho + \sqrt{1 - \rho^2}N/\rho$  for a standard normal random variable  $N$ , we find that

$$\mathbb{E} \left[ \exp(-\alpha Y_b^2/2) ; Y_b < b \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-b\sqrt{\frac{1-\rho^2}{1+\rho}}} \exp \left( -\frac{\alpha}{2} \left( \frac{b}{\rho} + \frac{\sqrt{1-\rho^2}}{\rho} z \right)^2 - z^2/2 \right) dz.$$

We next compute the maximizer of the integrand by setting the derivative of the exponent equal to 0. We find that  $z^* = b\nu$ , where

$$\nu = -\frac{\alpha\sqrt{1-\rho^2}}{\rho^2 + \alpha(1-\rho^2)},$$

so that  $z^*$  lies in the integration interval if  $\alpha > \rho/(1 + \rho)$ . The Laplace method yields in that case that

$$\begin{aligned} \lim_{b \rightarrow \infty} \frac{1}{b^2} \log \mathbb{E} [\exp(-\alpha Y_b^2/2); Y_b < b] &= -\frac{\alpha}{2} \left( \frac{1}{\rho} + \frac{\sqrt{1-\rho^2}}{\rho} \nu \right)^2 - \nu^2/2 \\ &= -\frac{\alpha[1 + \alpha(1-\rho^2)]}{2(\rho^2 + \alpha(1-\rho^2))^2}, \end{aligned}$$

and it is readily verified that this is larger than  $-\alpha/2$  if  $\alpha \geq 2$ . □

*Proof of Proposition 2.3.1.* As derived earlier, we utilize the representation,

$$\begin{aligned} &\mathbb{E} \left[ \Psi \left( \xi_b^1(X^{(1)}) \right)^\alpha \right] \\ &= \Psi(b)^\alpha P(X_2^{(1)} \leq b(1-\rho)) + \mathbb{E} \left[ \Psi \left( (b - X_2^{(1)})/\rho \right)^\alpha; X_2^{(1)} > b(1-\rho) \right] \\ &= \Psi(b)^\alpha P(Y_b \geq b) + \mathbb{E} [\Psi(Y_b)^\alpha; Y_b < b], \end{aligned}$$

where  $Y_b \sim \mathcal{N}(b/\rho, (1-\rho^2)/\rho^2)$ .

Now, with lemma (2.3.1) also at hand, and using the fact that  $P(Y_b \geq b) \rightarrow 1$ , we may conclude that, as  $b \rightarrow \infty$ ,

$$\lim_{b \rightarrow \infty} \frac{1}{b^2} \log \mathbb{E} \left[ \Psi \left( \xi_b^1(X^{(1)}) \right)^2 \right] > -1.$$

On the other hand, we also may establish,

$$\frac{1}{b^2} \log \mathbb{E} \left[ \Psi \left( \xi_b^1(X^{(1)}) \right) \right] = \frac{1}{b^2} \log P(\max(X_1, X_2) > b),$$

which converges to  $-1/2$  as  $b \rightarrow \infty$ . □

### Exponential twisting.

To improve upon the above estimator, one may consider the *importance sampling* estimator, to be abbreviated “IS”, which we define here as:

$$Z_b^{\text{IS}} := e^{-bX_1 - b^2/2} \mathbf{1}\{\max(X_1, X_2 - b(1 - \rho)) > 0\}.$$

We may further establish that the above estimator is unbiased using the below identity,

$$\begin{aligned} \mathbb{P}[\max(X_1, X_2) > b] &= \mathbb{E} \left[ e^{-bX_1 - b^2/2}; \max(X_1 + b, X_2 + \rho b) > b \right] \\ &= \mathbb{E} \left[ e^{-bX_1 - b^2/2}; \max(X_1, X_2 - (1 - \rho)b) > 0 \right]. \end{aligned}$$

**Proposition 2.3.2.** *The importance sampling estimator  $Z_b^{\text{IS}}$  via exponential twisting, with form*

$$Z_b^{\text{IS}} := e^{-bX_1 - b^2/2} \mathbf{1}\{\max(X_1, X_2 - b(1 - \rho)) > 0\},$$

*has relative error for some  $c > 0$ ,*

$$\log e(b) \sim cb^2, \quad b \rightarrow \infty.$$

As stated above, the above estimator is not a good choice for the reason that the second moment is unbounded in  $b$ . This result was also shown in [25]. In other words, we may establish for the above IS estimator that

$$\mathbb{E}(\Theta_b^{\text{IS}})^2 \gg (\mathbb{E}\Theta_b^{\text{IS}})^2, \quad b \rightarrow \infty.$$

*Proof.* By the Laplace method (see also Dieker and Mandjes, e.g. [11]), we have

$$\begin{aligned} & \lim_{b \rightarrow \infty} \frac{1}{b^2} \log \mathbb{E} \left[ e^{-\alpha b X_1 - \alpha b^2/2}; \max(X_1, X_2 - (1 - \rho)b) > 0 \right] \\ &= - \inf_{\{x \in \mathbb{R}^2: \max(x_1, x_2 - (1 - \rho)) \geq 0\}} \left[ \alpha x_1 + \alpha/2 + \frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1 - \rho^2)} \right], \end{aligned}$$

and it is readily seen that  $x = (-(1 - \rho)(2 + \rho), 1 - \rho)$  is the minimizer for  $\alpha = 2$  while  $x = (-(1 - \rho), 1 - \rho)$  and  $x = (0, 0)$  are the minimizers for  $\alpha = 1$ . Since the minimizer for the second moment is not a minimizer for the first moment, the estimator is not logarithmically efficient. □

## CHAPTER 3

### CONSTRAINED GAUSSIAN MAXIMUM

#### 3.1 Overview.

In this chapter, our focus is the construction of an efficient estimator for simulating the exceedance probability with respect to a discrete, standard Gaussian vector, but in the presence of simplifying constraints. This chapter is meant an expository study for the more general techniques introduced later in this thesis. The estimation problem focused on here is chosen so that the most important, basic ideas underlying our method may be distilled from the subtleties necessary for more general estimation problems.

Specifically, in this chapter, we study the estimation of

$$\mathbb{P} \left[ \max_{i=1, \dots, n} X_i > b, \min_{i=1, \dots, n} X_i > 0 \right], \quad (3.1)$$

where each of the  $X_i$  are standard normal and  $\text{Cov}(X_j, X_k) > 0$  for all  $j, k$ . As before, we assume the parameter  $b$  to be large, so that traditional Monte Carlo simulation would not be practical. This complication motivates the use of a variance-reduction technique such as importance sampling, as we utilize in our method.

The two properties which we intend to study for our estimator are first, its mean, which to be unbiased must have the same expectation as the original probability, and secondly, its variance. The variance property is critically related to the notion of algorithmic efficiency.

Thus, simply stated, our goal is the construction of an unbiased Monte Carlo estimator  $\Theta_b$ , whereby we can achieve the dual qualities:

1. *Unbiased property*:

$$\mathbb{E}\Theta_b = \mathbb{P} \left( \max_{i=1,\dots,n} X_i > b, \min_{i=1,\dots,n} X_i > 0 \right);$$

2. *Strong efficiency*: Having bounded relative error for  $b \rightarrow \infty$ , given by the inequality, for  $C$  a positive constant,

$$\lim_{b \rightarrow \infty} e_{\Theta}(b) \leq C.$$

In this chapter, we provide only empirical support for bounded relative error, stated above as point 2, and defer a theoretical discussion of the asymptotic properties of our estimator to Chapter 4. Numerical evidence for the efficiency performance of our algorithm is provided via several simulation studies which close this chapter.

### 3.1.1 Primary assumptions and definitions.

Mathematically, as stated, we define  $X$  to be an  $n$ -dimensional, multivariate Gaussian random vector, where each of the marginals  $X_i$  is standard normal for  $i = 1, \dots, n$ , and the covariance between any two marginal components is strictly positive, that is  $\text{Cov}(X_j, X_k) > 0$  for all  $j, k = 1, \dots, n$ . In the absence of a subscript, we let  $X$  stand for a vector representation, i.e.  $X := (X_1, \dots, X_n)$ .

In addition to the Gaussian vector  $X$ , we suppose that we have a uniform random variable,  $J$ , where  $J$  is uniformly distributed on  $\{1, \dots, n\}$ , as well as an exponential random variable,  $\Lambda$ , where  $\Lambda$ , has the standard exponential distribution, mutually independent of  $J$  and also independent of  $X$ .

With these observations, we then define:

$$X^{(j)} := X - X_j w^j, \quad j = 1, \dots, n \quad (3.2)$$

where  $w^j$  is a  $n$ -dimensional vector with

$$w_k^j = \text{Cov}(X_j, X_k), \quad j, k = 1, \dots, n.$$

As we will show, our estimator heavily exploits the fact that  $X^{(j)}$  is independent of  $X_j$ .

Furthermore, for  $b > 0$ , we set

$$A_b = \left\{ x \in \mathbb{R}^n : \max_{i=1, \dots, n} x_i > b, \min_{i=1, \dots, n} x_i > 0 \right\}, \quad (3.3)$$

so that the probability of interest from (3.1) is equivalently written  $\mathbb{P}[X \in A_b]$ .

### 3.1.2 Simulation set-up.

**Basic formulation.** The Monte Carlo estimator  $\Theta_b$  we have constructed utilizes the variables given above, as well as some additional ingredients specified here.

We note here that a sample realization of  $\Theta_b$  from a given simulation replication  $k$  may be denoted  $\Theta_b^{(k)}$ . After all simulation replications are finished, and assuming  $m$  replications in total, our final estimate for  $\mathbb{E}\Theta_b$  would be calculated according to the sample average, given by

$$\bar{\Theta}_b = \frac{1}{m} \sum_{k=1}^m \Theta_b^{(k)}.$$

It follows trivially that  $\bar{\Theta}_b$  is an unbiased estimator of  $\Theta_b$ . For notational simplification, in all subsequent references to our estimator, we refer to the underlying random variable  $\Theta_b$ .

We need a few definitions before we can formulate our algorithm. We first define a truncation

function  $t$  by setting  $t(\xi) = \max(|\xi|, 1)$  for  $\xi \in \mathbb{R}$ . We also define, for  $j = 1, \dots, n$ ,  $x \in \mathbb{R}^n$ , and  $y \in \mathbb{R}$ ,

$$h_b^j(y; x) = \frac{\phi(y)}{\frac{1}{n} \sum_{k=1}^n \exp(bx_k - b(1 - w_k^j)y)} \quad (3.4)$$

and

$$\xi_b^j(x) = \max \left( \max_{k=1, \dots, n} \frac{-x_k}{w_k^j}, \min_{k=1, \dots, n} \frac{b - x_k}{w_k^j} \right). \quad (3.5)$$

**Algorithm.** We now describe the steps for our algorithm to generate a single simulation replication.

**Algorithm 3.1.1.** *The algorithm takes the inputs defined below, with the main procedure and output following.*

*Inputs:* A Gaussian random vector  $X$  of dimension  $n$ , with standard marginals; a standard exponential random variable  $\Lambda$ , which has unit mean; and a uniform random variable  $J$  on  $\{1, \dots, n\}$ .

*Steps:* We generate sample realizations of  $X$  and  $\Lambda$  as well as an index  $j$  from the uniform variable  $J$ . From these inputs, we construct the sample vector for  $X^{(j)}$  as well as the related quantities  $\xi_b^j(X^{(j)})$  and  $h_b^j(y; X^{(j)})$ .

*Output:* We then output our estimator  $\Theta_b$ :

$$\Theta_b = \frac{1}{t(\xi_b^j(X^{(j)}))} e^{\Lambda} h_b^j \left( \frac{\Lambda}{t(\xi_b^j(X^{(j)}))} + \xi_b^j(X^{(j)}); X^{(j)} \right), \quad (3.6)$$

## 3.2 Theoretical results.

### 3.2.1 Estimator is unbiased.

In this section, we show that our estimator is unbiased for estimating  $\mathbb{P}[X \in A_b]$ . We start with writing the event of interest in terms of  $X_j$  and  $X^{(j)}$ , which are independent. This lemma is the reason for imposing the condition of positive covariances; we do not need this assumption in the later chapters of this thesis.

**Lemma 3.2.1.** *For  $b > 0$  and any  $j = 1, \dots, n$ , we have*

$$\{X \in A_b\} = \{X_j > \xi_b^j(X^{(j)})\}.$$

*Proof.* By definition of  $\xi_b^j$ , the statement  $\xi_b^j(x) < y$  is equivalent to the following two statements:

(1)  $-x_k < w_k^j y$  for  $k = 1, \dots, n$ , and (2) there exists some  $\ell$  such that  $b - x_\ell < w_\ell^j y$ .

Statement (1) is equivalent to  $x_k + w_k^j y > 0$  for every  $k$  and statement (2) is equivalent to the existence of some  $\ell$  such that  $x_\ell + w_\ell^j y > b$ . For  $x = X^{(j)}$  and  $y = X_j$ , (1) becomes  $\min_k X_k > 0$  and (2) becomes  $\max_\ell X_\ell > b$ . □

The following lemma is a key ingredient in our proof that our estimator is unbiased.

**Lemma 3.2.2.** *Suppose the random variable  $J$  is uniformly distributed on  $\{1, \dots, n\}$ , and independent of  $X$ . For  $b > 0$ , we then have*

$$\mathbb{P}[X \in A_b] = \mathbb{E} \left[ \int_{\xi_b^J(X^{(J)})}^{\infty} h_b^J(y; X^{(J)}) dy \right].$$

*Proof.* For  $b > 0$ , we have

$$\begin{aligned}
\mathbb{P}[X \in A_b] &= \sum_{j=1}^n \mathbb{E} \left[ \frac{\mathbf{1}[X \in A_b]}{\sum_{k=1}^n e^{bX_k - bX_j}} \right] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \frac{\mathbf{1}[X_j > \xi_b^j(X^{(j)})]}{\sum_{k=1}^n e^{bX_k^{(j)} - b(1-w_k^j)X_j}} \right] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \int_{-\infty}^{\infty} \frac{\phi(y)}{\sum_{k=1}^n e^{bX_k^{(j)} - b(1-w_k^j)y}} \mathbf{1}[y > \xi_b^j(X^{(j)})] dy \right] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \int_{\xi_b^j(X^{(j)})}^{\infty} \frac{\phi(y)}{\sum_{k=1}^n e^{bX_k^{(j)} - b(1-w_k^j)y}} dy \right] \\
&= \frac{1}{n} \sum_{j=1}^n \mathbb{E} \left[ \int_{\xi_b^j(X^{(j)})}^{\infty} h_b^j(y; X^{(j)}) dy \right],
\end{aligned}$$

and the claim follows.  $\square$

Recall (from calculation in Chapter 2) that for any  $\mu > 0$  and  $\xi \in \mathbb{R}$  and any function  $h$ , we have

$$\int_{\xi}^{\infty} h(y) dy = \frac{1}{\mu} \mathbb{E} [e^{\Lambda} h(\Lambda/\mu + \xi)],$$

where  $\Lambda$  is a standard exponential. The best choice for  $\mu$  is  $\mu = \xi$  as  $\xi \rightarrow \infty$ , but this choice is not possible for  $\xi = 0$ . That is why we use  $\mu = t(\xi)$ , and

$$\frac{1}{t(\xi)} e^{\Lambda} h(\Lambda/t(\xi) + \xi)$$

is therefore an unbiased estimator for  $\int_{\xi}^{\infty} h(y) dy$ .

Combining this with Lemma 3.2.2, we obtain the following proposition.

**Proposition 3.2.1.** *For  $b > 0$ , we have*

$$\mathbb{E}\Theta_b = \mathbb{P} \left[ \max_{i=1, \dots, n} X_i > b, \min_{i=1, \dots, n} X_i > 0 \right].$$

### 3.3 Numerical experiments.

#### 3.3.1 Overview.

In this section, we present two numerical simulation studies, focused on two specific instances of the problem form we introduced at the beginning of this chapter, that is  $\mathbb{P}[X \in A_b]$ , equivalently,

$$\mathbb{P} \left[ \max_{i=1,\dots,n} X_i > b, \min_{i=1,\dots,n} X_i > 0 \right].$$

The two instances are chosen by changing the covariance structure of the Gaussian vector  $X$ , which also have the benefit of allowing closed-form solutions to which we can compare our method for accuracy. These instances are:

1. Independent and identically-distributed (“IID”) marginals  $X_i$ . Here the pairwise correlation is set approximately equal to 0.
2. Perfectly correlated marginals  $X_i$ . Here the pairwise correlation is set approximately equal to 1.

#### 3.3.2 IID Study.

For this case study, because the marginals are IID, the problem calculation admits a straightforward analytical solution:

$$\begin{aligned} & \mathbb{P} \left[ \max_{i=1,\dots,n} X_i > b, \min_{i=1,\dots,n} X_i > 0 \right] \\ &= \mathbb{P}[\min_{i=1,\dots,n} X_i > 0] - \mathbb{P} \left[ \max_{i=1,\dots,n} X_i \leq b, \min_{i=1,\dots,n} X_i > 0 \right] \\ &= [\mathbb{P}[X_1]]^n - [\mathbb{P}[0 < X_1 < b]]^n \\ &= (1/2)^n - [\Phi(b) - 1/2]^n, \end{aligned}$$

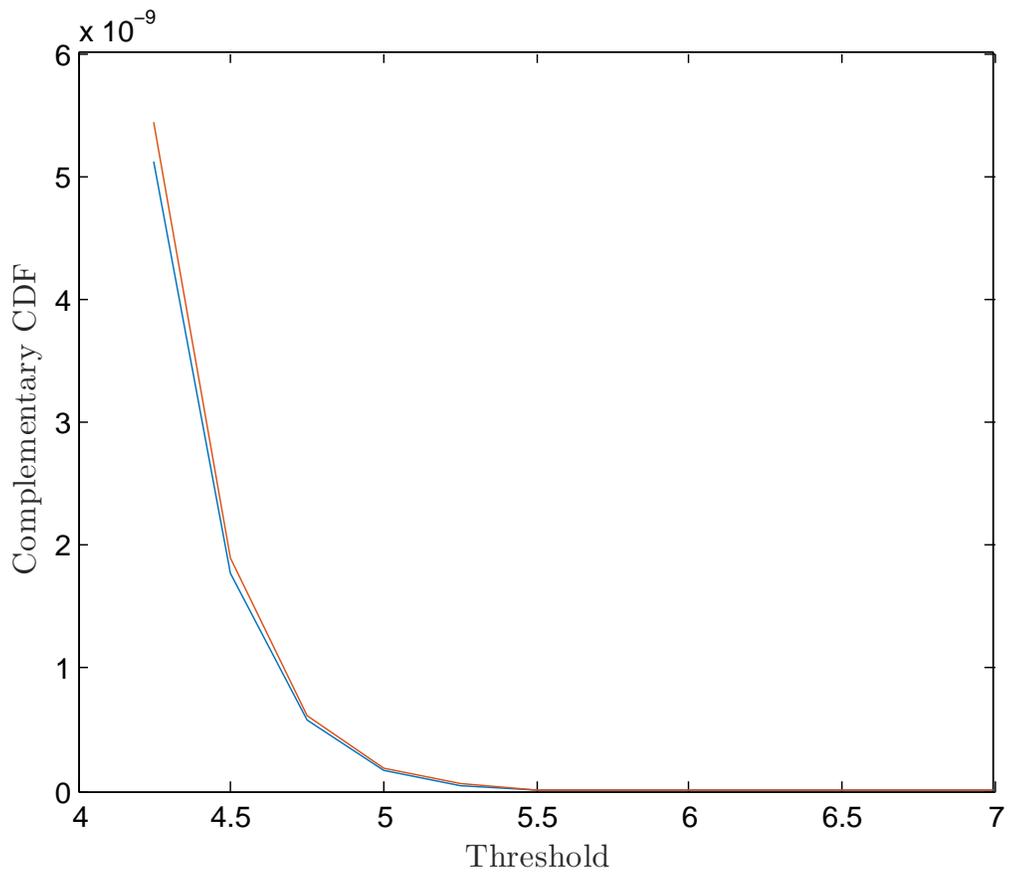
where  $\Phi(\cdot)$  represents the normal cdf. Thus, we may compare the values generated by our method against the true values from the analytical calculation to check for accuracy.

First, we present a table given in 3.1 which summarizes the simulated versus true values, and also includes the sample variance and relative error calculations of our Monte Carlo estimates. We use  $m = 10^6$  simulation replications and set  $n = 16$ . We also give graphical summaries in 3.1 and 3.2.

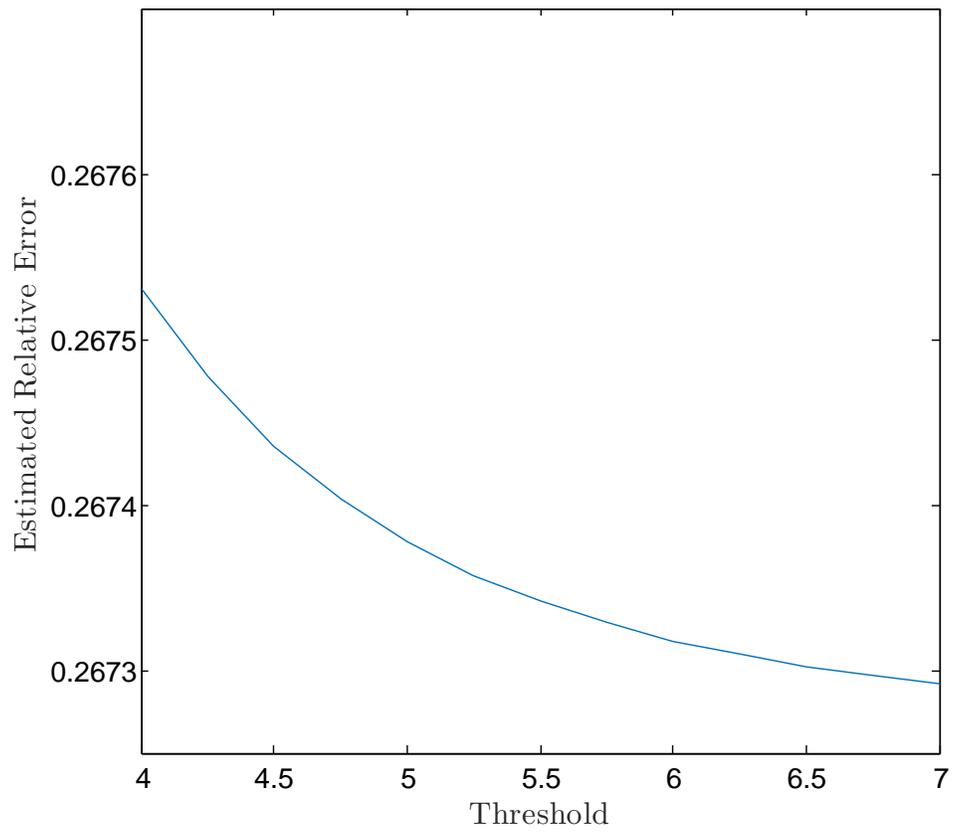
**Table 3.1:** IID Study: Monte Carlo Estimates vs. True Values

| Threshold | Sample Mean | True Values | Std. Error | Relative Error |
|-----------|-------------|-------------|------------|----------------|
| 4         | 7.729E-09   | 8.211E-09   | 2.068E-09  | 0.26753        |
| 4.5       | 8.266E-10   | 8.813E-10   | 2.211E-10  | 0.26744        |
| 5         | 6.949E-11   | 7.436E-11   | 1.858E-11  | 0.26738        |
| 5.5       | 4.591E-12   | 4.926E-12   | 1.227E-12  | 0.26734        |
| 6         | 2.380E-13   | 2.559E-13   | 6.361E-14  | 0.26732        |
| 6.5       | 9.669E-15   | 1.042E-14   | 2.585E-15  | 0.26730        |
| 7         | 3.077E-16   | 3.320E-16   | 8.224E-17  | 0.26729        |

Second, we present a graphical summary of the complementary CDF and relative error performance. The two graphs together show not only do the MC estimates quite closely approximate the true values, but also that the relative error of the estimation procedure stays bounded, even as the problem becomes computationally more difficult for higher thresholds.



**Figure 3.1:** IID Study: Complementary CDF



**Figure 3.2:** IID Study: Relative Error vs. Thresholds

### 3.3.3 Perfect correlation study.

For this case study, because the marginals are perfectly correlated, the problem has the simple solution,

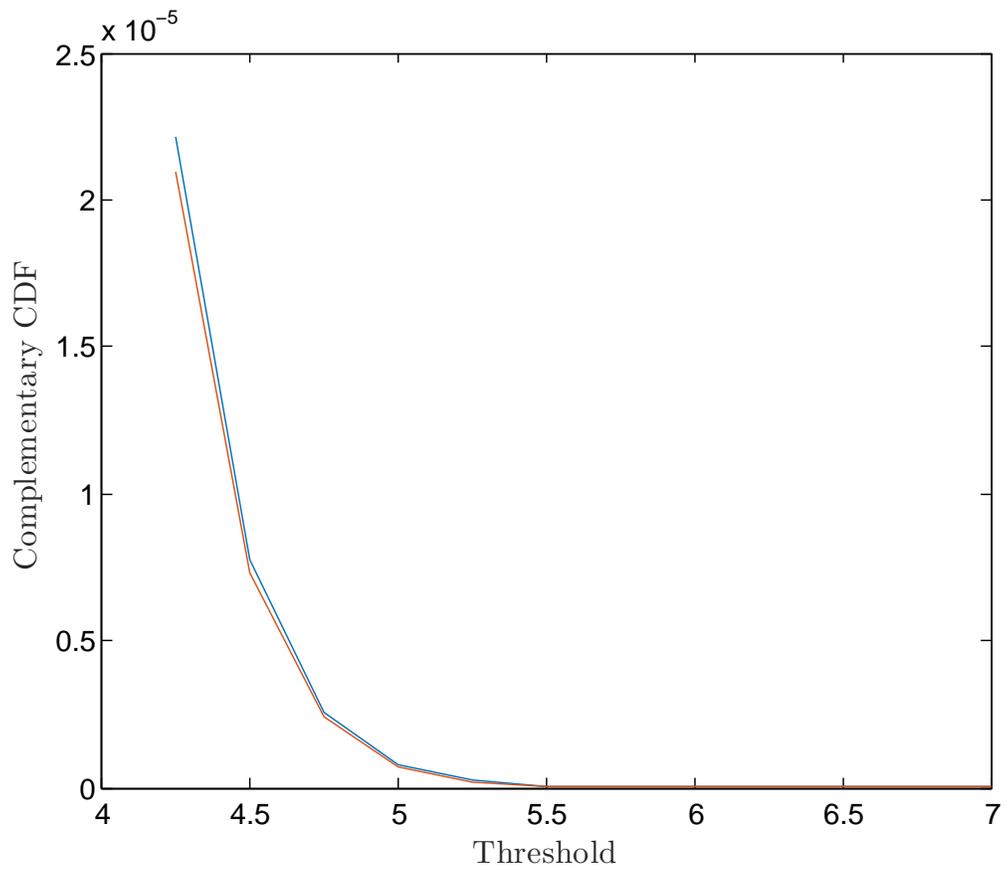
$$\begin{aligned} & \mathbb{P} \left[ \max_{i=1,\dots,n} X_i > b, \min_{i=1,\dots,n} X_i > 0 \right] \\ &= \mathbb{P}[X_1 > b, X_1 > 0] \\ &= 1 - \Phi(b). \end{aligned}$$

Next, we present a table given in 3.2 which summarizes the simulated versus true values, and also includes the sample variance and relative error calculations of our Monte Carlo estimates. As before, for the Monte Carlo parameters, we use  $10^6$  replications, and  $n = 16$  Gaussian marginals. We also give graphical summaries in 3.3 and 3.4.

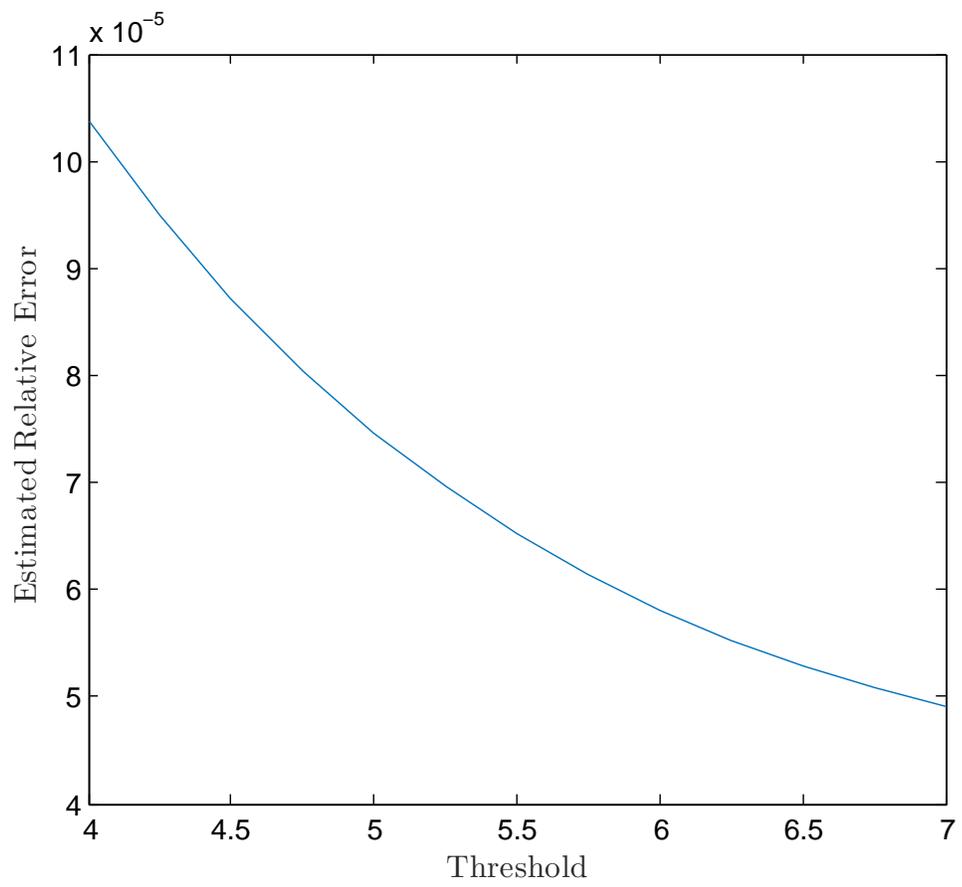
**Table 3.2:** Perfect Correlation Study: Monte Carlo Estimates vs. True Values

| Threshold | Sample Mean | True Values | Std. Error | Relative Error |
|-----------|-------------|-------------|------------|----------------|
| 4         | 3.167E-05   | 3.167E-05   | 3.235E-09  | 1.021E-04      |
| 4.5       | 3.397E-06   | 3.398E-06   | 2.886E-10  | 8.496E-05      |
| 5         | 2.866E-07   | 2.867E-07   | 2.053E-11  | 7.163E-05      |
| 5.5       | 1.899E-08   | 1.899E-08   | 1.161E-12  | 6.112E-05      |
| 6         | 9.866E-10   | 9.866E-10   | 5.199E-14  | 5.270E-05      |
| 6.5       | 4.016E-11   | 4.016E-11   | 1.842E-15  | 4.586E-05      |
| 7         | 1.280E-12   | 1.280E-12   | 5.150E-17  | 4.024E-05      |

Second, we present a graphical summary of the complementary CDF and relative error performance. As with the previous study, data from the tables and accompanying two graphs altogether are empirical evidence both that the MC estimates are highly accurate, and also that the relative error of the estimation procedure has stayed bounded, suggesting bounded relative error.



**Figure 3.3:** Perfect Correlation Study: Complementary CDF



**Figure 3.4:** Perfect Correlation Study: Relative Error vs. Thresholds

### 3.4 Appendix.

This appendix explains how our technique is related to a change-of-measure lemma from Dieker and Yakir [12]; we present this result below. The proof shown here is adapted from [12]. While that paper presented a claim in the context of fractional Brownian motion, the lemma used here has been modified specifically for the Gaussian vectors relevant to this research study.

**Lemma 3.4.1.** *Let  $\xi$  be a random vector in  $\mathbb{R}^n$  with a centered Gaussian distribution. Write  $\Xi_i = \xi_i - \text{Var}(\xi_i)/2$  and  $\Xi_i^j = \xi_i - \text{Var}(\xi_i - \xi_j)/2$  for  $i, j = 1, \dots, n$ . Then for any measurable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , we have for any  $j = 1, \dots, n$ , that*

$$\mathbb{E} [e^{\Xi_j} f(\Xi_1, \dots, \Xi_n)] = \mathbb{E} [f(\Xi_1^j + \text{Var}(\Xi_j)/2, \dots, \Xi_n^j + \text{Var}(\Xi_j)/2)].$$

*Proof.* To establish the equivalence of the two sides of the equality in the lemma as given here, we can show the equivalence of moment-generating functions. To do so, we compare the function obtained from the new, implied measure, versus that from the natural (original) probability measure, let us call it  $\mathbb{P}$ . For simplicity we let  $\mathbb{E} \equiv \mathbb{E}^{\mathbb{P}}$ .

We define, for an arbitrary event  $A \in \mathcal{F}$  on some measurable space  $(\Omega, \mathcal{F})$ , the probability measure  $\mathbb{Q}$  given by  $\mathbb{Q}(A) = \mathbb{E} [e^{\Xi_j} \mathbf{1}(A)]$ , with the random variable  $\Xi_j$  being as defined in the lemma. Let us denote by  $\mathbb{E}^{\mathbb{Q}}$  the expectation taken under  $\mathbb{Q}$ . For the same event  $A \in \mathcal{F}$ , we also have the original probability measure  $\mathbb{P}$ , where  $\mathbb{P}(A) = \mathbb{E} \mathbf{1}(A)$ .

Without loss of generality, we may establish the proof for the linear functional and arbitrary scaling vector  $t$ ,  $f(t\Xi) =: f(t_i \Xi_i, \dots, t_n \Xi_n) = \sum_{i=1}^n t_i \Xi_i$ . Taking for simplicity of presentation

$t \equiv 1$ , we find that

$$\begin{aligned}
\log \mathbb{E}^{\mathbb{Q}} [e^{f(\Xi)}] &= \log \mathbb{E} \left[ e^{\sum_{i=1}^n \Xi_i + \Xi_j} \right] \\
&= \mathbb{E} \left[ \sum_{i=1}^n \Xi_i + \Xi_j \right] + 1/2 \cdot \text{Var} \left( \sum_{i=1}^n \Xi_i + \Xi_j \right) \\
&= \mathbb{E} \left[ - \sum_{i=1}^n \text{Var}(\xi_i)/2 \right] + \text{Var}(\xi_j)/2 + 1/2 \cdot \text{Var} \left( \sum_{i=1}^n \xi_i + \xi_j \right) \\
&= - \sum_{i=1}^n (\text{Var}(\xi_i)/2 - \text{Cov}(\xi_i, \xi_j)) + 1/2 \cdot \text{Var} \left( \sum_{i=1}^n \xi_i \right) \\
&= \sum_{i=1}^n (\Xi_i^j + \text{Var}(\Xi_j)/2) + 1/2 \cdot \text{Var} \left( \sum_{i=1}^n \Xi_i^j \right) \\
&= \mathbb{E}[f(\Xi)] + 1/2 \cdot \text{Var}(f(\Xi)),
\end{aligned}$$

and this establishes the claim. □

This change of measure lemma is used in the following alternative proof of Lemma 3.2.2.

*Alternative proof of Lemma 3.2.2.* Choosing  $\xi_i = bX_i$  and

$$f(x) = \frac{\mathbf{1}[x \in A_b]}{\sum_{k=1}^n e^{bx_k - b^2/2}}$$

in Lemma 3.4.1, it follows that:

$$\begin{aligned}
\mathbb{P}[X \in A_b] &= \sum_{j=1}^n \mathbb{E} \left[ e^{bX_j - b^2/2} \frac{\mathbf{1}[X \in A_b]}{\sum_{k=1}^n e^{bX_k - b^2/2}} \right] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \frac{\mathbf{1}[X + bw^j \in A_b]}{\sum_{k=1}^n e^{b(X_k + bw_k^j) - b^2/2}} \right] \\
&= e^{-b^2/2} \sum_{j=1}^n \mathbb{E} \left[ \frac{\mathbf{1}[X + bw^j \in A_b]}{\sum_{k=1}^n e^{bX_k - b^2(1-w_k^j)}} \right] \\
&= e^{-b^2/2} \sum_{j=1}^n \mathbb{E} \left[ \frac{\mathbf{1}[X^{(j)} + X_j w^j + bw^j \in A_b]}{\sum_{k=1}^n e^{bX_k^{(j)} + bX_j w_k^j - b^2(1-w_k^j)}} \right].
\end{aligned}$$

From Lemma 3.2.2 we deduce that

$$\{X^{(j)} + X_j w^j + b w^j \in A_b\} = \{X_j + b > \xi_b^j(X^{(j)})\},$$

so that, since  $X^{(j)}$  and  $X_j$  are independent,

$$\begin{aligned} \mathbb{P}[X \in A_b] &= e^{-b^2/2} \sum_{j=1}^n \mathbb{E} \left[ \frac{\mathbf{1}[X_j + b > \xi_b^j(X^{(j)})]}{\sum_{k=1}^n e^{bX_k^{(j)} + bX_j w_k^j - b^2(1-w_k^j)}} \right] \\ &= e^{-b^2/2} \sum_{j=1}^n \mathbb{E} \left[ \int_{\xi_b^j(X^{(j)})-b}^{\infty} \frac{\phi(y)}{\sum_{k=1}^n e^{bX_k^{(j)} + b y w_k^j - b^2(1-w_k^j)}} dy \right] \\ &= e^{-b^2/2} \sum_{j=1}^n \mathbb{E} \left[ \int_{\xi_b^j(X^{(j)})}^{\infty} \frac{\phi(y+b)}{\sum_{k=1}^n e^{bX_k^{(j)} + b(y+b) w_k^j - b^2(1-w_k^j)}} dy \right], \end{aligned}$$

and it is readily verified that this agrees with the statement of the lemma. □

## CHAPTER 4

### GENERAL GAUSSIAN FUNCTIONALS

#### 4.1 Overview.

Our goal in this chapter is the construction of an efficient simulation algorithm to compute the quantity, for a level-crossing parameter  $b$ , and for a wide class of measurable functions  $g$ :

$$\mathbb{P}[g(X) > b]. \tag{4.1}$$

As before, we assume  $b$  to be large, making traditional Monte Carlo procedures inefficient.

Consistent with previous chapters, we denote the marginals of  $X$  by  $X_i$  for  $i = 1, \dots, n$ . Additionally, because here we consider general centered marginals, with arbitrary covariance structure, we set  $\text{Var}(X_i) =: \sigma_i^2$ . Our assumption that  $X$  has zero mean is without loss of generality. We could also have assumed that all marginals are standard normal without loss of generality, but we do not do so since we find it insightful to be able to work with arbitrary variances. Unlike in Chapter 3, we do not assume the covariance between arbitrary marginal components  $X_i$  and  $X_j$  to be strictly positive.

##### 4.1.1 Primary definitions.

Mathematically, as stated, we define  $X$  to be an  $n$ -dimensional, multivariate Gaussian random vector, where each of the marginals  $X_i$  has arbitrary variance  $\sigma_i^2$  for  $i = 1, \dots, n$ . In the absence of a subscript, we let  $X$  stand for a vector representation, i.e.  $X := (X_1, \dots, X_n)$ .

In addition to the Gaussian vector  $X$ , we use a random variable  $J$  with a uniform distribution on  $\{1, \dots, n\}$ , as well as a random variable  $\Lambda$  with a standard exponential distribution, mutually

independent of  $J$  and also independent of  $X$ .

With these observations, we then define:

$$X^{(j)} := X - X_j w^j / \sigma_j^2, \quad j = 1, \dots, n \quad (4.2)$$

where  $w^j$  is a  $n$ -dimensional vector with

$$w_k^j = \text{Cov}(X_j, X_k), \quad j, k = 1, \dots, n.$$

We then also define for  $x \in \mathbb{R}^n$ ,

$$\xi_b^j(x) = \inf \{ \xi \in \mathbb{R} : g(x + \xi w^j) > b \}, \quad (4.3)$$

where in addition, we require that  $g(x + \xi w^j)$  is nondecreasing in  $\xi$  for every  $j = 1, \dots, n$  and every  $x \in \mathbb{R}^n$ . We impose no other assumptions on the function  $g$ . The monotonicity requirement can be relaxed significantly in a relatively straightforward manner, at the expense of more complex estimators with additional indicators. For the sake of the presentation, we do not pursue this.

#### 4.1.2 Simulation set-up.

As in Chapter 3, we define a truncation function  $t$  by setting  $t(\xi) = \max(|\xi|, 1)$  for  $\xi \in \mathbb{R}$ . We also define, for  $j = 1, \dots, n$ ,  $x \in \mathbb{R}^n$ , and  $y \in \mathbb{R}$ ,

$$h_b^j(y; x) = \frac{\phi(y)}{\frac{1}{n} \sum_{k=1}^n e^{\xi_b^k(0)x_k - (\xi_b^j(0) - \xi_b^k(0)w_k^j/\sigma_j^2)\sigma_j y}}. \quad (4.4)$$

We first discuss how to estimate  $\int_{\xi}^{\infty} h(y) dy$  for a general function  $h$  with the ‘bulk’ close to 0. In our setting, the bulk location is inherited from the standard Gaussian density in the numerator, while the denominator acts as a (multiple of) an indicator function. As we see shortly, the key

complication relative to the setting of Chapter 3 is that  $\xi$  can be very negative and therefore far to the left of the ‘bulk’ of the function  $h$ , yet a good estimator must capture the bulk well.

**Estimating  $\int_{\xi}^{\infty} h(y)dy$ .** Our discussion of techniques for estimating  $\int_{\xi}^{\infty} h(y)dy$  builds on Chapter 2. As was shown in Chapter 2, it holds that for any  $\mu > 0$  and  $\xi \in \mathbb{R}$  and any function  $h$ ,

$$\int_{\xi}^{\infty} h(y)dy = \frac{1}{\mu} \mathbb{E} [e^{\Lambda} h(\Lambda/\mu + \xi)],$$

where  $\Lambda$  is a standard exponential. The best choice for  $\mu$  is  $\mu = \xi$  as  $\xi \rightarrow \infty$ ; however, this choice is problematic for  $\xi \leq 0$ .

As an additional potential problem, the bulk of the integral of  $h$  over  $(\xi, \infty)$  may lie far away from  $\xi$ , in which case the unbiased estimator

$$\frac{1}{\mu} e^{\Lambda} h(\Lambda/\mu + \xi), \tag{4.5}$$

underestimates the sought integral since  $\Lambda/\mu$  doesn’t hit the bulk of  $h$ . Thus, we seek an estimator that avoid such a rare event estimation problem.

We solve this problem by working with

$$\int_{\xi}^{\infty} h(y)dy = \int_{-\infty}^{\infty} h(y)dy - \int_{-\infty}^{\xi} h(y)dy$$

Both terms need to be estimated. If  $\xi < 0$  then we can use a similar estimator as in (4.5) for the second term, and there is no rare event estimation problem for this term. The first term may need to be estimated as well. For instance, since our functions  $h$  are ‘almost’ standard normal densities,

$$\frac{h(Z)}{\phi(Z)}$$

is a reasonable choice for an unbiased estimator of the integral

$$\int_{-\infty}^{\infty} h(y) dy,$$

where  $Z$  is a standard normal random variable.

Then, with the truncation function  $t(\xi)$  as defined above, we use the following estimator for  $\int_{\xi}^{\infty} h(y) dy$ :

$$\begin{cases} \frac{1}{t(\xi)} e^{\Lambda} h\left(\frac{\Lambda}{t(\xi)} + \xi\right) & \xi \geq 0 \\ \int_{-\infty}^{\infty} h_{-\infty}(y) dy - \frac{1}{t(\xi)} e^{\Lambda} h\left(-\frac{\Lambda}{t(\xi)} + \xi\right) & \xi < 0, \end{cases}$$

where we estimate the first term on the second line; more details are given in the proof of Proposition 4.2.1 below.

**Algorithm.** We now describe the steps for our algorithm to generate a single simulation replication.

**Algorithm 4.1.1.** *The algorithm takes the inputs defined below, with the main procedure and output following.*

*Inputs:* A Gaussian random vector  $X$  of dimension  $n$ , with  $\text{Var}(X_j) = \sigma_j^2$  for all  $j$ ; a standard exponential random variable  $\Lambda$ , which has unit mean; and a uniform random variable  $J$ , where  $J \sim \text{unif}(1, \dots, n)$ .

*Steps:* We generate sample realizations of  $X$  and  $\Lambda$  as well as an index  $j$  from the uniform variable  $J$ . From these inputs, we construct the sample vector for  $X^{(J)}$  as well as the related quantities  $\xi_b^J(x)$  and  $h_b^J(y; X^{(J)})$ .

*Output:* We then output our estimator  $\Theta_b$ :

$$\Theta_b = \begin{cases} \frac{1}{t(\sigma_J \xi_b^J(X^{(J)}))} e^\Lambda h_b^J \left( \frac{\Lambda}{t(\sigma_J \xi_b^J(X^{(J)}))} + \sigma_J \xi_b^J(X^{(J)}); X^{(J)} \right) & \xi_b^J(X^{(J)}) \geq 0 \\ \frac{e^{\xi_b^J(0)X_J}}{\frac{1}{n} \sum_{k=1}^n e^{\xi_b^k(0)X_k}} - \frac{1}{t(\sigma_J \xi_b^J(X^{(J)}))} e^\Lambda h_b^J \left( -\frac{\Lambda}{t(\sigma_J \xi_b^J(X^{(J)}))} + \sigma_J \xi_b^J(X^{(J)}); X^{(J)} \right) & \xi_b^J(X^{(J)}) < 0. \end{cases} \quad (4.6)$$

## 4.2 Theoretical results.

### 4.2.1 Estimator is unbiased.

The following lemma is the analog of Lemma 3.2.2.

**Lemma 4.2.1.** *Suppose the random variable  $J$  is uniformly distributed on  $\{1, \dots, n\}$ , and independent of  $X$ . For  $b$  in  $\mathbb{R}$ , we then have*

$$\mathbb{P}[g(X) > b] = \mathbb{E} \left[ \int_{\sigma_J \xi_b^J(X^{(J)})}^{\infty} h_b^J(y; X^{(J)}) dy \right],$$

where

$$h_b^j(y; X^{(j)}) = \frac{\phi(y)}{\frac{1}{n} \sum_{k=1}^n e^{\xi_b^k(0)X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0)w_k^j/\sigma_j^2)\sigma_j y}}.$$

*Proof.* For  $b \in \mathbb{R}$ , we have

$$\begin{aligned}
& \mathbb{P}[g(X) > b] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \frac{e^{\xi_b^j(0)X_j} \mathbf{1}[g(X) > b]}{\sum_{k=1}^n e^{\xi_b^k(0)X_k}} \right] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \frac{e^{\xi_b^j(0)X_j} \mathbf{1}[g(X^{(j)} + w^j X_j/\sigma_j^2) > b]}{\sum_{k=1}^n e^{\xi_b^k(0)X_k^{(j)} + \xi_b^k(0)w_k^j X_j/\sigma_j^2}} \right] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \frac{\mathbf{1}[g(X^{(j)} + w^j X_j/\sigma_j^2) > b]}{\sum_{k=1}^n e^{\xi_b^k(0)X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0))w_k^j/\sigma_j^2} X_j}} \right] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \int_{-\infty}^{\infty} \frac{\phi(y/\sigma_j)/\sigma_j}{\sum_{k=1}^n e^{\xi_b^k(0)X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0))w_k^j/\sigma_j^2} y} \mathbf{1}[g(X^{(j)} + w^j y/\sigma_j^2) > b] dy \right] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \int_{\sigma_j^2 \xi_b^j(X^{(j)})}^{\infty} \frac{\phi(y/\sigma_j)/\sigma_j}{\sum_{k=1}^n e^{\xi_b^k(0)X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0))w_k^j/\sigma_j^2} y} dy \right] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \int_{\sigma_j \xi_b^j(X^{(j)})}^{\infty} \frac{\phi(y)}{\sum_{k=1}^n e^{\xi_b^k(0)X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0))w_k^j/\sigma_j^2} \sigma_j y} dy \right] \\
&= \frac{1}{n} \sum_{j=1}^n \mathbb{E} \left[ \int_{\sigma_j \xi_b^j(X^{(j)})}^{\infty} h_b^j(y; X^{(j)}) dy \right],
\end{aligned}$$

as claimed. □

Here, proving that our estimator is unbiased is slightly more intricate compared to Chapter 3, since our estimator takes a different form depending on the sign of  $\xi_b^j(X^{(j)})$ .

**Proposition 4.2.1.** *For  $b \in \mathbb{R}$ , we have*

$$\mathbb{E}\Theta_b = \mathbb{P}(g(X) > b).$$

*Proof.* The discussion in Section 4.1.2 on estimating  $\int_{\xi}^{\infty} h(y)dy$ , in conjunction with Lemma 4.2.1,

shows that

$$\begin{cases} \int_{\sigma_J \xi_b^J(X^{(J)})}^{\infty} h_b^J(y; X^{(J)}) dy & \xi_b^J(X^{(J)}) \geq 0 \\ \int_{-\infty}^{\infty} h_b^J(y; X^{(J)}) dy - \int_{\sigma_J \xi_b^J(X^{(J)})}^{\infty} h_b^J(y; X^{(J)}) dy & \xi_b^J(X^{(J)}) < 0 \end{cases}$$

is an unbiased estimator for  $\mathbb{P}(g(X) > b)$ . Tracing the arguments in the proof of Lemma 4.2.1 in opposite order with  $\xi_b^j(X^{(j)}) \rightarrow -\infty$ , we find that

$$\frac{e^{\xi_b^J(0)X_J}}{\frac{1}{n} \sum_{k=1}^n e^{\xi_b^k(0)X_k}}$$

has the same mean as  $\int_{-\infty}^{\infty} h_b^J(y; X^{(J)}) dy$  conditional on  $X^{(J)}$  and  $J$ . □

#### 4.2.2 Estimator has desirable asymptotic properties.

For  $j = 1, \dots, n$  and  $b \in \mathbb{R}$ , we define the function  $f_b^j : \mathbb{R}^n \rightarrow \mathbb{R}_+$  by

$$f_b^j(x) = \frac{e^{\xi_b^j(0)x_j}}{\sum_{k=1}^n e^{\xi_b^k(0)x_k}}.$$

##### Conditions on the rare event set $A_b$ .

(A1) (Rare event condition) For every  $j$ , we have  $\lim_{b \rightarrow \infty} \xi_b^j(0) = \infty$ .

(A2) (Cone containment condition) If  $x \in A_b$ , then the cone  $x + \text{Conv}(\{w^j : j = 1, \dots, n\})$  also lies in  $A_b$ , where  $\text{Conv}(S)$  stands for the convex hull of the finite set  $S$ .

(A3) (Beam projection condition) For every  $j$  and for every  $K > 0$  large enough, we have

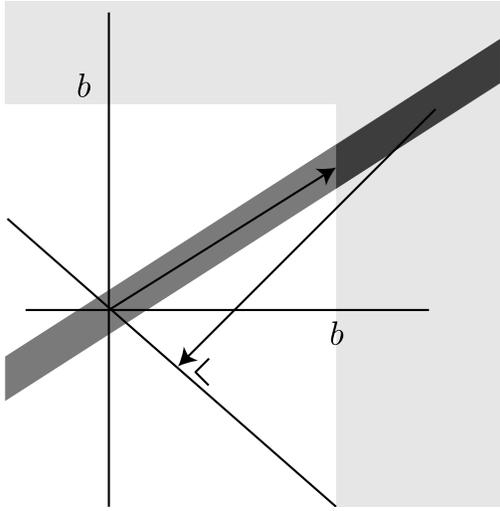
$$\lim_{b \rightarrow \infty} \sup_{x \in A_b \cap \{x : \|x - w^j x_j\| < K\}} |f_b^j(x) - 1| = 0.$$

(A4) (Straight hit condition) For every  $j$  and for every  $K > 0$  large enough, we have

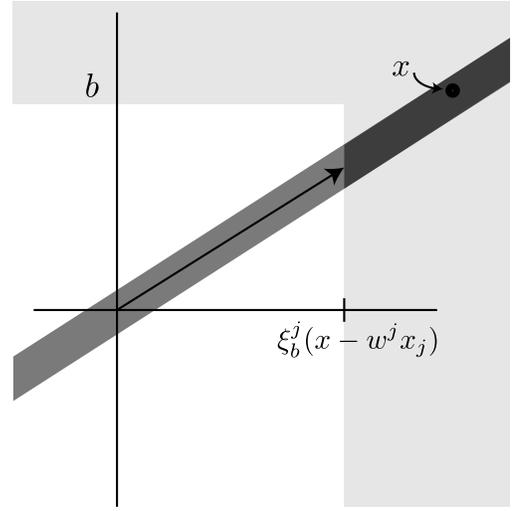
$$\lim_{b \rightarrow \infty} \sup_{x \in A_b \cap \{x: \|x - w^j x_j\| < K\}} |\xi_b^j(0)[\xi_b^j(x) - \xi_b^j(0)]| = 0.$$

We illustrate these assumptions in Figure 4.1 for the ‘standard’ case where

$$A_b = \{x : \max(x_1, x_2) > b\}.$$



(a) Beam projection condition. The darkest colored set is projected orthogonally onto a subset of the line  $\{x : x_1 + x_2 = 0\}$  for which  $f_b^1(x) \approx 1$ . Since  $\xi_b^j(0) = b$  for all  $j$ , the function  $f_b^j$  is constant when subtracting a multiple of  $(1, 1)$ .



(b) Straight hit condition. The darkest colored set is projected obliquely along the beam to a set with roughly the same  $\xi_b^j$  value (in fact, here, they have equal  $\xi_b^j$  value).

**Figure 4.1:** Illustrations of Assumption (A3), left, and Assumption (A4), right. The darkest colored set is  $A_b \cap \{x : \|x - w^j x_j\| < K\}$ .

The main result in this section is the following theorem.

**Theorem 4.2.2.** *Suppose that (A1)–(A4) hold. For the estimator  $\Theta_b$  defined in (4.6),*

$$\frac{\Theta_b}{\sum_{j=1}^n \frac{1}{\sigma_j \xi_b^j(0)} \phi(\sigma_j \xi_b^j(0))}$$

*converges to 1 in probability.*

The significance of this theorem is not so much the functional form of the denominator, but it is the fact that the estimator *after appropriate deterministic normalization* converges to a constant as the event becomes more rare. As such, this theorem highlights a property of our estimator that is consistent with a vanishing relative error. Strengthening this theorem to convergence of the first two moments, which immediately implies vanishing relative error, remains open and we leave this for further research.

We prove this theorem in a series of lemmas. It is an immediate consequence of the next lemma and (A1) that  $\xi_b^j(X^j) \rightarrow \infty$  as  $b \rightarrow \infty$ . In particular, to prove Theorem 4.2.2 it suffices to only consider the first row of (4.6). Our strategy is to rewrite parts of the estimator ‘up to a random error’, which vanishes as  $b \rightarrow \infty$ . We do so in a sequence of propositions (Proposition 4.2.2–Proposition 4.2.4), which together prove the theorem. Throughout, we assume that  $\sigma_j = 1$  for all  $j$  without loss of generality.

**Lemma 4.2.3.** *Under Assumption (A4),  $\xi_b^j(0)[\xi_b^j(X^{(j)}) - \xi_b^j(0)]$  converges to 0 in probability.*

*Proof.* The proof follows from arguments that we use repeatedly throughout this section. We have, for any  $\epsilon > 0$  and  $K > 0$ ,

$$P(\xi_b^j(0)[\xi_b^j(X^{(j)}) - \xi_b^j(0)] > \epsilon) \leq P(\|X^{(j)}\| \geq K) + P(\xi_b^j(0)[\xi_b^j(X^{(j)}) - \xi_b^j(0)] > \epsilon, \|X^{(j)}\| < K).$$

From Assumption (A4) and dominated convergence we deduce that the second term converges to

0 as  $b \rightarrow \infty$ . Consequently, we establish the claim by first letting  $b \rightarrow \infty$  and then  $K \rightarrow \infty$  in the preceding display.  $\square$

**Proposition 4.2.2.** *Under Assumptions (A1) and (A4),  $t(\xi_b^j(X^{(j)}))/\xi_b^j(0)$  converges to 1 in probability.*

*Proof.* It follows from the bound, for every  $\epsilon > 0$ ,

$$P(\xi_b^j(X^{(j)}) > \xi_b^j(0)(1 + \epsilon)) \leq P(\xi_b^j(X^{(j)}) - \xi_b^j(0) > \epsilon/\xi_b^j(0)),$$

which is valid by (A1) and converges to 0 by Lemma 4.2.3. From a similar lower bound we conclude that  $\xi_b^j(X^{(j)})/\xi_b^j(0)$  converges to 1 in probability. For large enough  $b$ , by (A1) and (A4) we have  $\xi_b^j(X^{(j)}) \geq 1$  and therefore  $t(\xi_b^j(X^{(j)})) = \xi_b^j(X^{(j)})$ .  $\square$

**Lemma 4.2.4.** *Suppose (A2) and (A3) hold. For  $x$  with  $x_j = 0$  and  $\|x\| < K$ , we have*

$$\lim_{b \rightarrow \infty} \sup_{x \in \mathbb{R}^n: x_j=0, \|x\| < K} \sup_{\lambda > 0} \left| \frac{h_b^j \left( \frac{\lambda}{t(\xi_b^j(x))} + \xi_b^j(x); x \right)}{n\phi \left( \frac{\lambda}{t(\xi_b^j(x))} + \xi_b^j(x) \right)} - 1 \right| = 0.$$

*Proof.* By definition of  $h_b^j$  and  $f_b^j$ , we have

$$\frac{h_b^j \left( \frac{\lambda}{t(\xi_b^j(x))} + \xi_b^j(x); x \right)}{n\phi \left( \frac{\lambda}{t(\xi_b^j(x))} + \xi_b^j(x) \right)} = f_b^j \left( x + w^j \left( \frac{\lambda}{t(\xi_b^j(x))} + \xi_b^j(x) \right) \right).$$

Noting that  $x + w^j \xi_b^j(x)$  lies in the (closure of)  $A_b$ , then by Assumption (A2)  $x + w^j \left( \frac{\lambda}{t(\xi_b^j(x))} + \xi_b^j(x) \right)$  lies in  $A_b$  as well. Moreover, we have

$$\sup_{x \in \mathbb{R}^n: x_j=0, \|x\| < K} \sup_{\lambda > 0} \left| f_b^j \left( x + w^j \left( \frac{\lambda}{t(\xi_b^j(x))} + \xi_b^j(x) \right) \right) - 1 \right| = \sup_{x \in A_b \cap \{x: \|x - w^j x_j\| < K\}} |f_b^j(x) - 1|.$$

Therefore, the lemma follows from Assumption (A3).  $\square$

As a corollary, we obtain a convergence in probability (or, since the limit is a constant, in distribution).

**Proposition 4.2.3.** *Suppose (A2) and (A3) hold. For every  $j$ , the random variable*

$$\frac{h_b^j \left( \frac{\Lambda}{t(\xi_b^j(X^{(j)}))} + \xi_b^j(X^{(j)}); X^{(j)} \right)}{n\phi \left( \frac{\Lambda}{t(\xi_b^j(X^{(j)}))} + \xi_b^j(X^{(j)}) \right)} = f_b^j \left( X^{(j)} + w^j \left( \frac{\Lambda}{t(\xi_b^j(X^{(j)}))} + \xi_b^j(X^{(j)}) \right) \right)$$

*converges to 1 in probability.*

*Proof.* The argument is similar to the one in the proof of Lemma 4.2.3. Writing  $Y_b^j$  for the random variable from the statement, then  $Y_b^j$  is bounded by 1 from above. For every  $\epsilon > 0$ , we find that

$$P(Y_b^j < 1 - \epsilon) \leq P(\|X^{(j)}\| \geq K) + P(Y_b^j < 1 - \epsilon, \|X^{(j)}\| < K). \quad (4.7)$$

Lemma 4.2.4 and dominated convergence imply that  $\lim_{b \rightarrow \infty} P(Y_b^j < 1 - \epsilon, \|X^{(j)}\| < K) = 0$ . Thus, first letting  $b \rightarrow \infty$  and then  $K \rightarrow \infty$  in (4.7) yields the claim.  $\square$

**Proposition 4.2.4.** *Suppose (A4) holds. For every  $j$ , the random variable*

$$\frac{\phi \left( \frac{\Lambda}{t(\xi_b^j(X^{(j)}))} + \xi_b^j(X^{(j)}) \right)}{e^{-\Lambda}\phi(\xi_b^j(0))}$$

*converges to 1 in probability.*

*Proof.* We give the proof with  $t(\xi_b^j(X^{(j)}))$  replaced by  $\xi_b^j(X^{(j)})$ , noting that we may do so in view of Proposition 4.2.2. We have

$$\frac{\phi \left( \frac{\Lambda}{\xi_b^j(X^{(j)})} + \xi_b^j(X^{(j)}) \right)}{e^{-\Lambda}\phi(\xi_b^j(0))} = \exp \left( -\frac{\Lambda^2}{2\xi_b^j(X^{(j)})^2} \right) \times \frac{\phi(\xi_b^j(X^{(j)}))}{\phi(\xi_b^j(0))}.$$

The first term converges to 1 in probability by Lemma 4.2.3. For the second term, we note that

$$\xi_b^j(X^{(j)})^2 - \xi_b^j(0)^2 = \xi_b^j(0)[\xi_b^j(X^{(j)}) - \xi_b^j(0)] \times [\xi_b^j(X^{(j)})/\xi_b^j(0) - 1],$$

which converges to 0 in probability by Lemma 4.2.3. As a result, the second term also converges to 1 in probability and this proves the claim.  $\square$

We are now in a position to establish the main theorem.

*Proof of Theorem 4.2.2.* Fix  $j$ . Propositions 4.2.3 and 4.2.4 show that

$$\frac{h_b^j\left(\frac{\Lambda}{t(\xi_b^j(X^{(j)}))} + \xi_b^j(X^{(j)}); X^{(j)}\right)}{ne^{-\Lambda}\phi(\xi_b^j(0))}$$

converges to 1 in probability. Together with Proposition 4.2.2, we thus find that

$$\frac{\frac{1}{t(\xi_b^j(X^{(j)}))}e^{\Lambda}h_b^j\left(\frac{\Lambda}{t(\xi_b^j(X^{(j)}))} + \xi_b^j(X^{(j)}); X^{(j)}\right)}{n\phi(\xi_b^j(0))/\xi_b^j(0)}$$

converges to 1 in probability.  $\square$

### 4.3 Numerical experiments.

#### 4.3.1 Overview.

In this section, we present three numerical simulation studies, focused on two specific instances of the problem form we introduced at the beginning of this chapter, that is  $\mathbb{P}[X \in A_b]$ , equivalently,

$$\mathbb{P}\left[\max_{i=1,\dots,n} X_i > b\right].$$

Similar to Chapter 3, the first problem instances are chosen by changing the covariance structure of the Gaussian vector  $X$ , which also have the benefit of allowing closed-form solutions to which

we can compare our method for accuracy. For an additional experiment, we also consider a cosine correlation structure. These instances are:

1. Perfectly correlated marginals  $X_i$ . Here the pairwise correlation is set approximately equal to 1.
2. A cosine correlation structure.

We note finally that we have not included an IID study in the numerical section for this chapter. This exclusion is due to added complications in the calculation of our method for  $\xi_b^j(x)$  in the IID case of Chapter 4, which is unlike in the other two studies for this section. These computational differences make the IID case here less comparable to the others.

#### 4.3.2 Perfect correlation study.

For this case study, because the marginals are perfectly correlated, the problem has the simple solution,

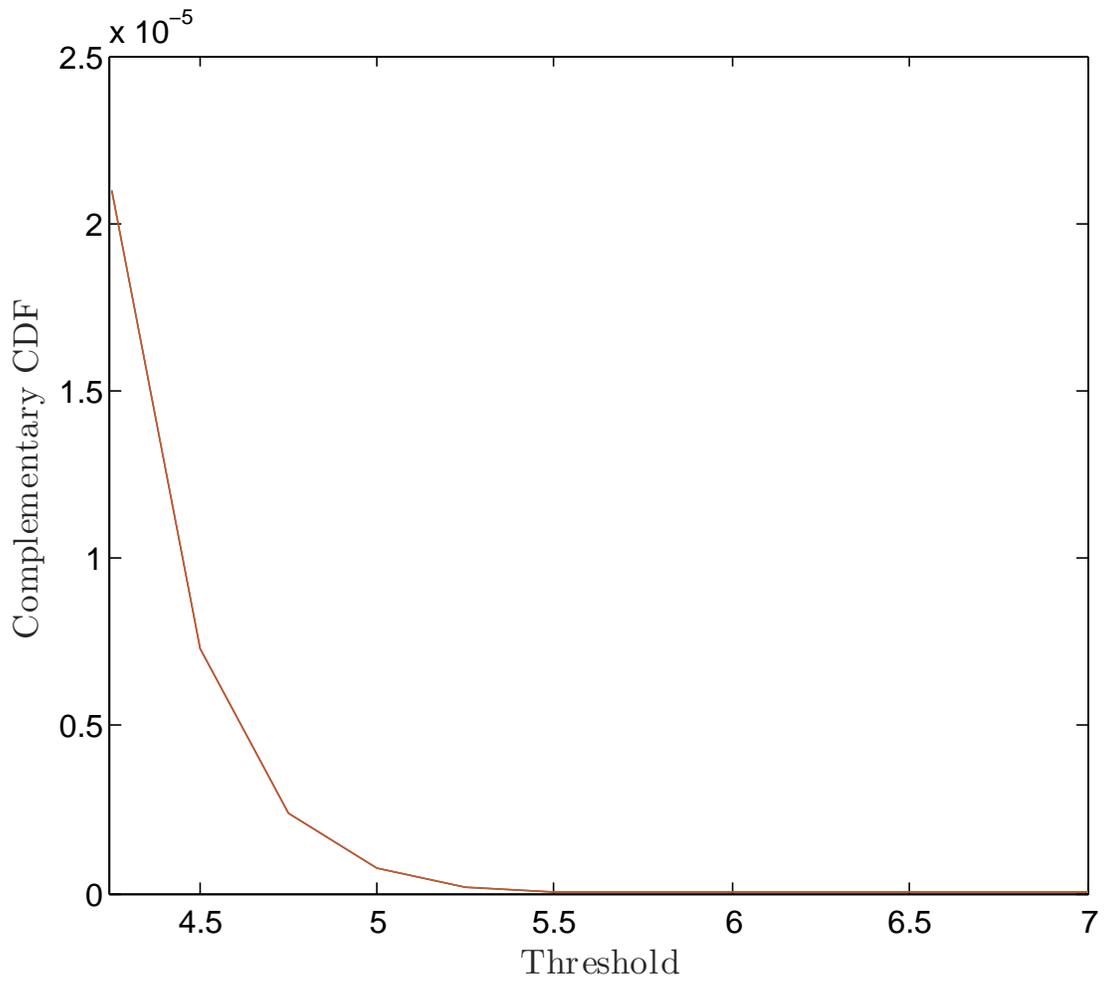
$$\mathbb{P} \left[ \max_{i=1, \dots, n} X_i > b \right] = \mathbb{P}[X_1 > b] = 1 - \Phi(b).$$

First, we present a table given in 4.1 which summarizes the simulated versus true values, and also includes the sample variance and relative error calculations of our Monte Carlo estimates. For the Monte Carlo parameters, we use  $10^6$  replications, and  $n = 16$  Gaussian marginals. Following this we also give graphical summaries in 4.2 and 4.3 of the data.

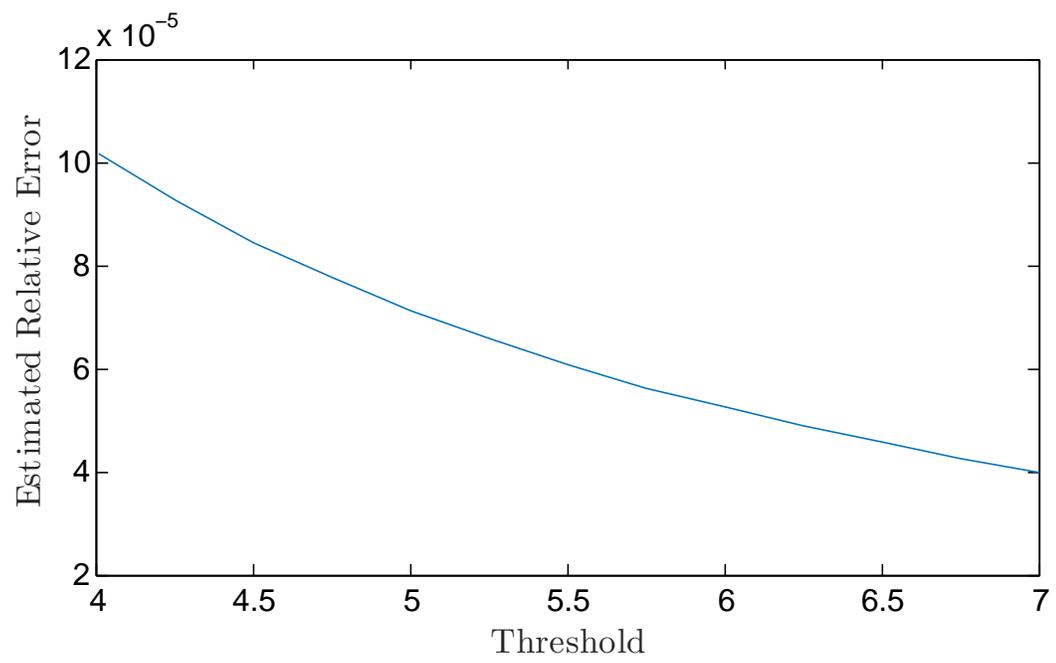
Second, we present a graphical summary of the complementary CDF and relative error performance. As with the previous study, data from the tables and accompanying two graphs altogether are empirical evidence both that the MC estimates are highly accurate, and also that the relative error of the estimation procedure has stayed bounded, suggesting computational efficiency.

**Table 4.1:** Perfect Correlation Study: Monte Carlo Estimates vs. True Values

| Threshold | Sample Mean | True Values | Std. Error | Relative Error |
|-----------|-------------|-------------|------------|----------------|
| 4         | 3.167E-05   | 3.167E-05   | 3.217E-09  | 1.016E-04      |
| 4.5       | 3.398E-06   | 3.398E-06   | 2.870E-10  | 8.45E-05       |
| 5         | 2.867E-07   | 2.867E-07   | 2.041E-11  | 7.12E-05       |
| 5.5       | 1.899E-08   | 1.899E-08   | 1.154E-12  | 6.07E-05       |
| 6         | 9.867E-10   | 9.866E-10   | 5.167E-14  | 5.24E-05       |
| 6.5       | 4.016E-11   | 4.016E-11   | 1.830E-15  | 4.56E-05       |
| 7         | 1.280E-12   | 1.280E-12   | 5.117E-17  | 4.00E-05       |



**Figure 4.2:** Perfect Correlation Study: Complementary CDF



**Figure 4.3:** Perfect Correlation Study: Relative Error vs. Thresholds

### 4.3.3 Cosine correlation study.

For this numerical test, we choose a cosine covariance function, where closed-form formulae exist to evaluate the answer exactly.

For this purpose, we represent the Gaussian process as a cosine field, given by the parameterization

$$f(t) = Z_1 \cos t + Z_2 \sin t, \quad (4.8)$$

where  $t$  stands for time, with  $Z_1$  and  $Z_2$  representing independent standard normals. As can be shown, the cosine process from (4.8) is a classical example of a stationary Gaussian process, where it can further be shown that the covariance between any two points is given by the cosine of the difference between those two points, i.e.

$$\mathbb{C}\text{ov}(f(s), f(t)) = \cos(t - s), \quad s, t \in \mathbb{R}.$$

Using this set-up, the result is known for the time interval  $t \in [0, .75]$  that

$$\mathbb{P} \left[ \sup_{t \in [0, .75]} f(t) > x \right] = 1 - \Phi(x) + \frac{3}{8\pi} e^{-x^2/2}, \quad (4.9)$$

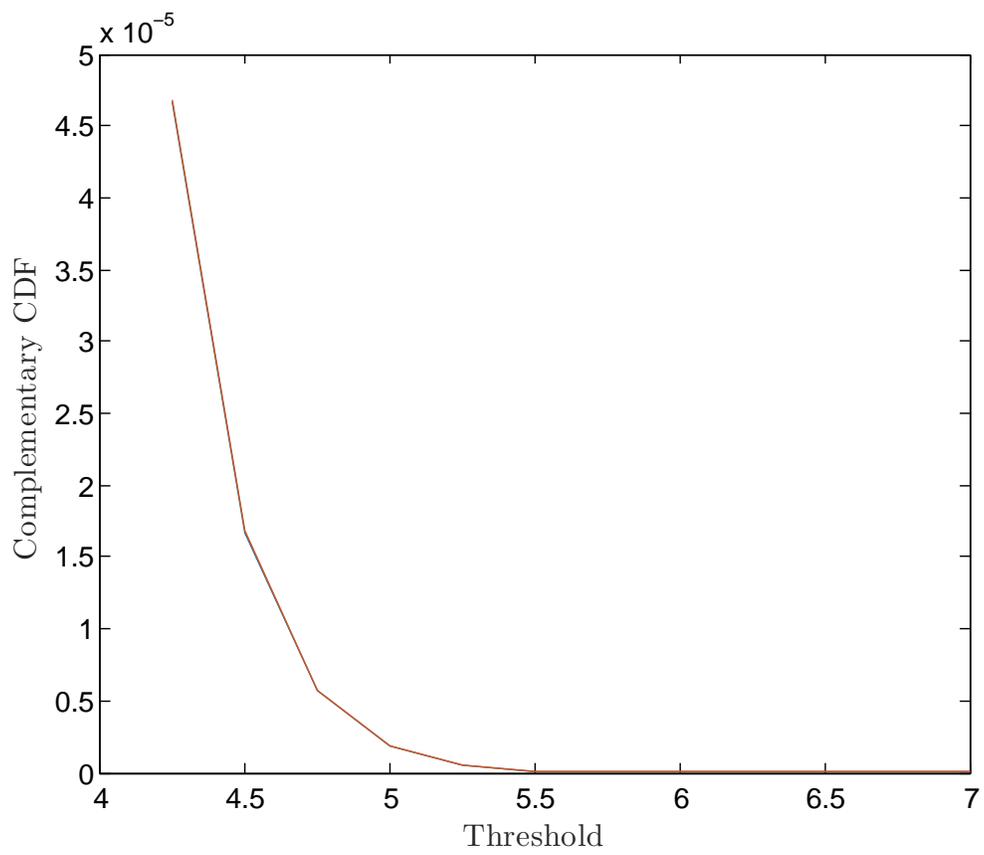
where  $\Phi(\cdot)$  represents the standard normal cumulative distribution function.

In our experiments, we estimate an approximation of this probability by discretizing the interval over which we take the supremum. In fact, we divide the interval into  $n = 16$  equal pieces and run  $10^6$  simulation replications.

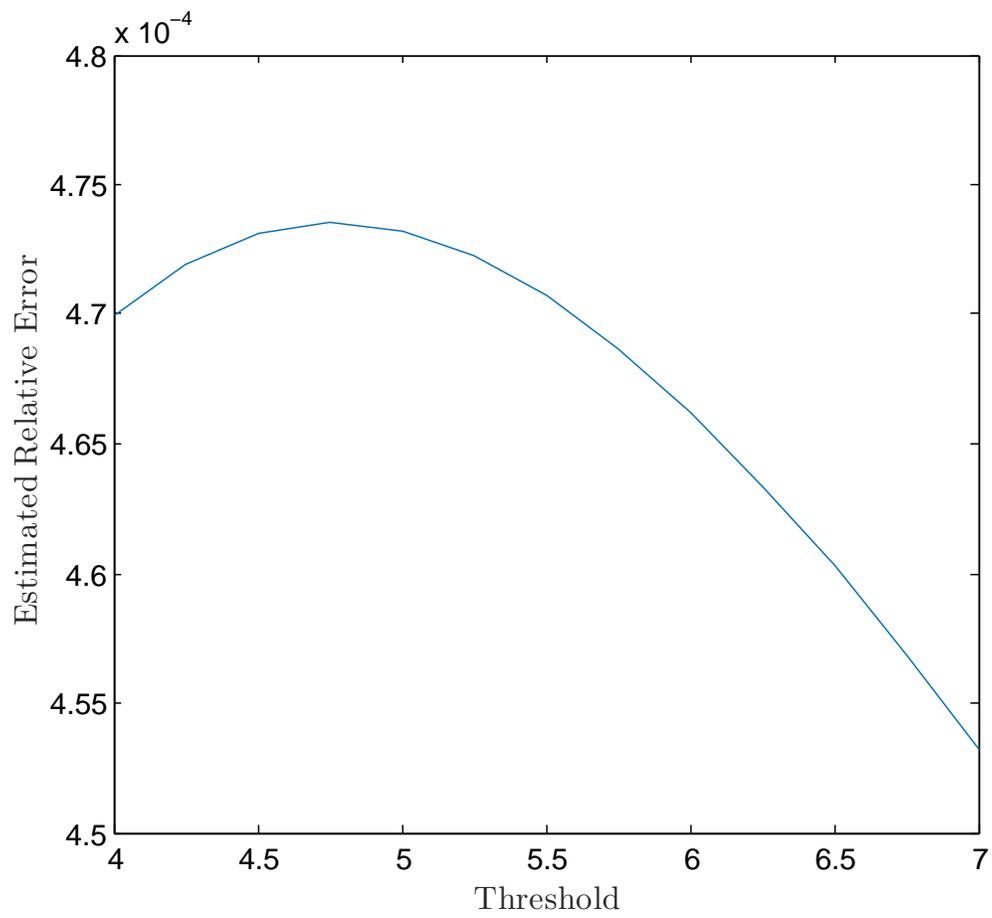
Below, we present a table in 4.2 which summarizes the simulated versus true values, and also includes the sample variance and relative error calculations. We also give graphical summaries following the table in 4.4 and 4.5.

**Table 4.2:** Cosine Study: Monte Carlo Estimates vs. True Values

| Threshold | Sample Mean | True Values | Std. Error | Relative Error |
|-----------|-------------|-------------|------------|----------------|
| 4         | 7.159E-05   | 7.171E-05   | 3.364E-08  | 4.699E-04      |
| 4.5       | 8.164E-06   | 8.180E-06   | 3.862E-09  | 4.731E-04      |
| 5         | 7.299E-07   | 7.315E-07   | 3.454E-10  | 4.732E-04      |
| 5.5       | 5.108E-08   | 5.121E-08   | 2.405E-11  | 4.707E-04      |
| 6         | 2.797E-09   | 2.805E-09   | 1.304E-12  | 4.662E-04      |
| 6.5       | 1.196E-10   | 1.200E-10   | 5.507E-14  | 4.603E-04      |
| 7         | 3.998E-12   | 4.013E-12   | 1.812E-15  | 4.532E-04      |



**Figure 4.4:** Cosine Study: Complementary CDF



**Figure 4.5:** Cosine Study: Relative Error vs. Thresholds

#### 4.4 Appendix.

This appendix works out the ‘standard maximum’ case in more detail, where  $g(X) = \max_{i=1,\dots,n} X_i$  and the  $X_i$  are standard normal. Throughout this appendix, we assume that  $0 < \text{Cov}(X_i, X_j) < 1$  for all  $i \neq j$ . We show that in this case, we can improve our estimator in the regime  $b \rightarrow -\infty$ . Recall that we chose our estimator in search of desirable properties as  $b \rightarrow \infty$ , yet it is unbiased for every  $b \in \mathbb{R}$ . In this appendix we are working towards an estimator that is efficient in both tails.

As  $b \rightarrow -\infty$  the first term on the second line in (4.6) becomes  $n\mathbf{1}[X_J = \min(X_1, \dots, X_n)]$ , which has a high variance if  $n$  is large. This is undesirable and therefore we introduce a control variate. This construction builds on the fact that this term, conditionally on  $X^{(j)}$ , has the same mean as

$$\int_{-\infty}^{\infty} h_b^j(y; X^{(j)}) dy,$$

which converges as  $b \rightarrow -\infty$  by dominated convergence to

$$\int_{-\infty}^{\infty} h_{-\infty}^j(y; X^{(j)}) dy$$

where, for  $y \in \mathbb{R}$ ,

$$h_{-\infty}^j(y; X^{(j)}) := \lim_{b \rightarrow -\infty} h_b^j(y; X^{(j)}) = n\phi(y)\mathbf{1}\left(y \leq \min_{k \neq j} \frac{X_k^{(j)}}{1 - w_k^j}\right).$$

As a result,

$$\int_{-\infty}^{\infty} h_{-\infty}^j(y; X^{(j)}) dy = n\Phi\left(\min_{k \neq j} \frac{X_k^{(j)}}{1 - w_k^j}\right),$$

has, conditionally on  $X^{(j)}$ , the same mean as  $n\mathbf{1}[X_j = \min(X_1, \dots, X_n)]$ .

Therefore, the first expression on the second line of (4.6) has, conditionally on  $J$  and  $X^{(J)}$ , the

same mean as

$$n\Phi\left(\min_{k \neq J} \frac{X_k^{(J)}}{1 - w_k^J}\right) - n\mathbf{1}[X_j = \min(X_1, \dots, X_n)] + \frac{e^{\xi_b^J(0)X_J}}{\frac{1}{n} \sum_{k=1}^n e^{\xi_b^k(0)X_k}} - \frac{1}{t(\sigma_J \xi_b^J(X^{(J)}))} e^{\Lambda h_b^J} \left( -\frac{\Lambda}{t(\sigma_J \xi_b^J(X^{(J)}))} + \sigma_J \xi_b^J(X^{(J)}); X^{(J)} \right),$$

so by replacing the second line of (4.6) by this expression, we have an estimator with lower variance as  $b \rightarrow -\infty$ .

## CHAPTER 5

### CONDITIONAL FUNCTIONALS

#### 5.1 Overview.

Our goal in this chapter is the construction of an efficient simulation algorithm to compute the quantity, for a level-crossing parameter  $b$ , and for any measurable functions  $g$ :

$$\mathbb{E}[f(X) \mid g(X) > b]. \tag{5.1}$$

As before, we assume  $b$  to be large, making traditional Monte Carlo procedures inefficient.

Problems involving a conditional expectation are readily incorporated into the existing methodology we have established in this thesis. To see the connection, we need only to note that the conditional expectation from the above display can be reduced to the ratio of two unconditional expectations, given by

$$\frac{\mathbb{E}[f(X); g(X) > b]}{\mathbb{P}[g(X) > b]}.$$

This chapter is intended as an expository treatment, similar to Chapter 3; thus, for the construction of the conditional estimator, we focus on linear functionals  $f(X)$ , given a vector  $c \in \mathbb{R}^n$  of the form

$$f(X) := \langle c, X \rangle.$$

This simplification is to allow a concise form of the resulting estimator, without additional terms which may arise in more complicated functions. Examples of other relevant functions would be a sum of squared marginals  $\sum_i X_i^2$ , or sums involving cross-product terms  $X_i \cdot X_j$ .

Having developed theory for the conditional estimator, we then discuss a wide array of problem applications. Particularly, these areas of application may be understood as motivated by observa-

tions concerning financial tail risk, which often incorporate conditional statements.

## 5.2 Applications in finance.

Problems involving conditional expectations of the form  $\mathbb{E}[f(X)|g(X) > b]$ , as we explore here in this chapter, are particularly relevant to work done in insurance and financial risk. We focus in this chapter on financial applications involving the computation of risk measures.

Risk measures are closely related to the work done in Chapter 4 because they involve tail probabilities of form  $\mathbb{P}[g(X) > b]$ . For this instance, the Gaussian vector  $X$  may be comprised of components  $X_1, X_2, \dots, X_n$  which correspond to  $n$  assets in a hypothetical portfolio. The computation of the tail probability  $\mathbb{P}[g(X) > b]$  is then used to assess the firm's level of financial risk. The choice of  $g$  will vary depending on how a practitioner measures loss. In some situations, total loss may be most important, in which case a linear functional summing the losses across portfolio assets  $X_i$  may be used. In other cases, where the size of a specific loss in an individual asset is important, a functional such as  $\max_i X_i$ , rather than total loss, may be used instead.

Mathematically, risk measures can be understood as a probability mapping, often based on calculating a quantile  $q$ , such that  $q = \mathbb{P}[g(X) > b]$ , or instead directly using the underlying tail probability, to which a risk value is assigned. In the classical setting of VaR, the risk metric is the extreme quantile itself, but related metrics are also often used, frequently alongside VaR. These additional measures may condition on certain tail events, such as a VaR which exceeds a certain loss threshold.

Financial risk measures have received considerable attention in the literature, particularly following severe market volatility in 2007 and 2008. VaR and its associated measures such as expected shortfall, also known as conditional VaR or average VaR, can be categorized within a much broader class of risk functionals, with differing attributes including coherence versus non-coherence, or convexity, which are summarized, for example, in [16] as well as [14] and [15].

For assessing financial risk, one important area of study looks at fluctuations in the market

value of a portfolio's total assets. The total assets may be represented in a simple setting by the linear sum  $\sum_i X_i$ , with a risk measure computed conditional on a loss level given by a threshold  $b$ . In the notation of this chapter, taking  $g(X) := \max_i X_i$ , the relevant quantity to compute would be

$$\mathbb{E} \left[ \sum_i X_i \mid \max_i X_i > b \right].$$

Because these problem involve tail probabilities for extreme thresholds  $b$ , the underlying probabilities are rare events to which the present line of research is well-suited. Moreover, given the rise in complicated portfolio structures in the financial industry, there is a need for computationally-powerful simulation algorithms which, in turn, are flexible across different choices of risk functionals. We believe the flexibility of our method to incorporate various functions  $f$  and  $g$  is a key benefit, although we show only a simple setting in this chapter.

### 5.2.1 Primary definitions.

Consistent with prior chapters, we define  $X$  to be an  $n$ -dimensional, multivariate Gaussian random vector. As in Chapter 4, we allow each of the marginals  $X_i$  to have arbitrary variance  $\sigma_i^2$  for  $i = 1, \dots, n$ , and we define  $w_i^j = \text{Cov}(X_i, X_j)$ . Also consistent with prior usage, we let  $J$  signify a uniformly-distributed random variable on  $\{1, \dots, n\}$ , and let  $\Lambda$  be a standard, exponentially-distributed random variable, mutually independent of  $J$  and also independent of  $X$ .

Our definitions of the associated variables  $X^{(j)}$  and  $\xi_b^j$  correspond to the previous formulas (4.2) and (4.3), respectively. To assist the reader, we have included these definitions here below, with

$$X^{(j)} := X - X_j w^j / \sigma_j^2, \quad j = 1, \dots, n$$

and for  $x \in \mathbb{R}^n$ ,

$$\xi_b^j(x) = \inf \{ \xi \in \mathbb{R} : g(x + \xi w^j) > b \},$$

where  $g(x + \xi w^j)$  is nondecreasing in  $\xi$  for every  $j = 1, \dots, n$  and every  $x \in \mathbb{R}^n$ .

Furthermore, as previously done, we use a truncation function  $t$  by setting  $t(\xi) = \max(|\xi|, 1)$  for  $\xi \in \mathbb{R}$ . We also define the function  $h_b^j(y; x)$ , given for  $j = 1, \dots, n$ ,  $x \in \mathbb{R}^n$ , and  $y \in \mathbb{R}$ , by

$$h_b^j(y; x) = \frac{\phi(y)}{\frac{1}{n} \sum_{k=1}^n e^{\xi_b^k(0)x_k - (\xi_b^j(0) - \xi_b^k(0)w_k^j/\sigma_j^2)\sigma_j y}}.$$

New to this chapter, we also define the related function, which we shall call  $u_b^j(y; x)$ , given by

$$u_b^j(y; x) := y \cdot h_b^j(y; x). \quad (5.2)$$

As mentioned in the Overview, we focus here on linear functionals  $f(X) = \sum_{i=1}^n c_i X_i$ . Here we let the  $X_i$  correspond to market value losses in a portfolio with  $n$  assets in total, where  $X_i$  represents the market value loss on asset  $i$ . (A positive  $X_i$  thus corresponds to a loss, while a negative  $X_i$  corresponds to a gain.) Additionally, we let the constants  $c_1, c_2, \dots, c_n$  correspond to arbitrary weights.

We then define the total portfolio market loss  $L$  to be given by

$$L := \sum_{i=1}^n c_i X_i, \quad c_i > 0, \quad i = 1, \dots, n. \quad (5.3)$$

We also define a “loss level” which we shall denote  $b$ .

Given the loss level  $b$ , the risk quantity which we will simulate is the expected loss  $L = \langle c, X \rangle$  given the occurrence of the loss event  $b$ . In this chapter, we will focus specifically on the max functional. The mathematical quantity for the conditional expected loss is then given by:

$$\mathbb{E} \left[ \langle c, X \rangle \mid \max_i X_i > b \right] = \frac{\mathbb{E} [\langle c, X \rangle ; \max_i X_i > b]}{\mathbb{P} [\max_i X_i > b]}. \quad (5.4)$$

### 5.2.2 Simulation set-up.

Our simulation goal is the construction of an unbiased Monte Carlo estimator  $\eta_b$  such that

$$\mathbb{E}\eta_b = \mathbb{E} \left[ \langle c, X \rangle ; \max_i X_i > b \right], \quad (5.5)$$

so that an estimator for the sought quantity (5.4) can be found upon taking the ratio of  $\eta_b$  and the estimator  $\Theta_b$  for the tail probability  $\mathbb{P}(g(X) > b)$  from Chapter 4. Thus, the estimator described in Chapter 5 uses the output from the Monte Carlo estimator in Chapter 4 as input.

**Algorithm.** We now describe the steps for our algorithm that generates our estimator for

$$\mathbb{E}[\langle c, X \rangle | g(X) > b].$$

**Algorithm 5.2.1.** *The algorithm takes the inputs defined below, with the main procedure and output following.*

*Inputs:* A Gaussian random vector  $X$  of dimension  $n$ , with  $\text{Var}(X_j) = \sigma_j^2$  for all  $j$ ; a standard exponential random variable  $\Lambda$ , which has unit mean; and a uniform random variable  $J$ , where  $J \sim \text{unif}(1, \dots, n)$ .

*Steps:* We generate sample realizations of  $X$  and  $\Lambda$  as well as an index  $j$  from the uniform variable  $J$ . From these inputs, we construct the sample vector for  $X^{(J)}$  as well as the related quantities  $\xi_b^J(x)$  and  $h_b^J(y; X^{(J)})$ . By the methods of Chapter 4, we also compute and store the value for the simulated tail probability, again to be denoted  $\Theta^b$ , given by:

$$\Theta_b = \begin{cases} \frac{1}{t(\sigma_J \xi_b^J(X^{(J)}))} e^\Lambda h_b^J \left( \frac{\Lambda}{t(\sigma_J \xi_b^J(X^{(J)}))} + \sigma_J \xi_b^J(X^{(J)}); X^{(J)} \right) & \xi_b^J(X^{(J)}) \geq 0 \\ \frac{e^{\xi_b^J(0)X_J}}{\frac{1}{n} \sum_{k=1}^n e^{\xi_b^k(0)X_k}} - \frac{1}{t(\sigma_J \xi_b^J(X^{(J)}))} e^\Lambda h_b^J \left( -\frac{\Lambda}{t(\sigma_J \xi_b^J(X^{(J)}))} + \sigma_J \xi_b^J(X^{(J)}); X^{(J)} \right) & \xi_b^J(X^{(J)}) < 0. \end{cases}$$

*Output:* We have the collective output:

$$\frac{\eta_{b,1} \mathbf{1}[\xi_b^J(X^{(J)}) \geq 0] + \eta_{b,2} \mathbf{1}[\xi_b^J(X^{(J)}) < 0]}{\Theta_b}, \quad (5.6)$$

where

$$\begin{aligned} \eta_{b,1} = & \langle c, X^{(J)} \rangle \left[ \frac{1}{t(\sigma_J \xi_b^J(X^{(J)}))} e^{\Lambda h_b^J} \left( \frac{\Lambda}{t(\sigma_J \xi_b^J(X^{(J)}))} + \sigma_J \xi_b^J(X^{(J)}) X^{(J)} \right) \right] \\ & + \frac{1}{t(\sigma_J \xi_b^J(X^{(J)}))} \langle c, w^J \rangle e^{\Lambda l_b^J} \left( \frac{\Lambda}{t(\sigma_J \xi_b^J(X^{(J)}))} + \sigma_J \xi_b^J(X^{(J)}) X^{(J)} \right) \end{aligned}$$

and

$$\begin{aligned} \eta_{b,2} = & \langle c, X^{(J)} \rangle \left[ \frac{e^{\xi_b^J(0) X^J}}{\frac{1}{n} \sum_{k=1}^n e^{\xi_b^k(0) X^k}} - \frac{1}{t(\sigma_J \xi_b^J(X^{(J)}))} e^{\Lambda h_b^J} \left( -\frac{\Lambda}{t(\sigma_J \xi_b^J(X^{(J)}))} + \sigma_J \xi_b^J(X^{(J)}); X^{(J)} \right) \right] \\ & - \frac{1}{t(\sigma_J \xi_b^J(X^{(J)}))} \langle c, w^J \rangle e^{\Lambda l_b^J} \left( -\frac{\Lambda}{t(\sigma_J \xi_b^J(X^{(J)}))} + \sigma_J \xi_b^J(X^{(J)}); X^{(J)} \right) \end{aligned}$$

We write  $\eta_b = \eta_{b,1} \mathbf{1}[\xi_b^J(X^{(J)}) \geq 0] + \eta_{b,2} \mathbf{1}[\xi_b^J(X^{(J)}) < 0]$ . If multiple Monte Carlo replications are run, our output is the sample average of the numerator in (5.6) over the sample average of the denominator in (5.6).

### 5.3 Theoretical results.

#### 5.3.1 Numerator of estimator is unbiased.

The following lemma is the analog of Lemma 4.2.1.

**Lemma 5.3.1.** *Suppose the random variable  $J$  is uniformly distributed on  $\{1, \dots, n\}$ , and inde-*

pendent of  $X$ . For  $b$  in  $\mathbb{R}$ , we then have

$$\mathbb{E}[\langle c, X \rangle; g(X) > b] = \mathbb{E} \left[ \int_{\sigma_j \xi_b^j(X^{(j)})}^{\infty} [\langle c, X^{(j)} \rangle + \langle c, w^j y \rangle] h_b^j(y; X^{(j)}) dy \right],$$

where

$$h_b^j(y; X^{(j)}) = \frac{\phi(y)}{\frac{1}{n} \sum_{k=1}^n e^{\xi_b^k(0)X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0)w_k^j/\sigma_j^2)\sigma_j y}}.$$

*Proof.* For  $b \in \mathbb{R}$ , we then may construct our estimator using the steps to follow, proceeding similar to the proof of unbiasedness we gave in Chapter 4:

$$\begin{aligned} & \mathbb{E}[\langle c, X \rangle; g(X) > b] \\ &= \sum_{j=1}^n \mathbb{E} \left[ \sum_{i=1}^n c_i \cdot X_i \cdot \frac{e^{\xi_b^j(0)X_j} \mathbf{1}[g(X) > b]}{\sum_{k=1}^n e^{\xi_b^k(0)X_k}} \right] \\ &= \sum_{j=1}^n \mathbb{E} \left[ \sum_{i=1}^n c_i (X_i^{(j)} + X_j w_i^j) \cdot \frac{e^{\xi_b^j(0)X_j} \mathbf{1}[g(X^{(j)} + w^j X_j/\sigma_j^2) > b]}{\sum_{k=1}^n e^{\xi_b^k(0)X_k^{(j)} + \xi_b^k(0)w_k^j X_j/\sigma_j^2}} \right] \\ &= \sum_{j=1}^n \mathbb{E} \left[ [\langle c, X^{(j)} \rangle + X_j \langle c, w^j \rangle] \cdot \frac{\mathbf{1}[g(X^{(j)} + w^j X_j/\sigma_j^2) > b]}{\sum_{k=1}^n e^{\xi_b^k(0)X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0)w_k^j/\sigma_j^2)X_j}} \right] \\ &= \sum_{j=1}^n \mathbb{E} \left[ [\langle c, X^{(j)} \rangle + X_j \langle c, w^j \rangle] \cdot \frac{\mathbf{1}[g(X^{(j)} + w^j X_j/\sigma_j^2) > b]}{\sum_{k=1}^n e^{\xi_b^k(0)X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0)w_k^j/\sigma_j^2)X_j}} \right]. \end{aligned}$$

From here we keep track of each component of the estimator separately, according to the components of the sum given by: (i)  $\langle c, X^{(j)} \rangle$  and (ii)  $X_j \langle c, w^j \rangle$ .

We start with the second component, (ii)  $X_j \langle c, w^j \rangle$ .

$$\begin{aligned}
& \sum_{j=1}^n \mathbb{E} \left[ X_j \langle c, w^j \rangle \cdot \frac{\mathbf{1}[g(X^{(j)} + w^j X_j / \sigma_j^2) > b]}{\sum_{k=1}^n e^{\xi_b^k(0) X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0) w_k^j / \sigma_j^2) X_j}} \right] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \int_{-\infty}^{\infty} \frac{\langle c, w^j y \rangle \phi(y / \sigma_j) / \sigma_j}{\sum_{k=1}^n e^{\xi_b^k(0) X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0) w_k^j / \sigma_j^2) y}} \mathbf{1}[g(X^{(j)} + w^j y / \sigma_j^2) > b] dy \right] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \int_{\sigma_j^2 \xi_b^j(X^{(j)})}^{\infty} \frac{\langle c, w^j y \rangle \phi(y / \sigma_j) / \sigma_j}{\sum_{k=1}^n e^{\xi_b^k(0) X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0) w_k^j / \sigma_j^2) y}} dy \right] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \int_{\sigma_j \xi_b^j(X^{(j)})}^{\infty} \frac{\langle c, w^j y \rangle \phi(y)}{\sum_{k=1}^n e^{\xi_b^k(0) X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0) w_k^j / \sigma_j^2) \sigma_j y}} dy \right] \\
&= \frac{1}{n} \sum_{j=1}^n \mathbb{E} \left[ \int_{\sigma_j \xi_b^j(X^{(j)})}^{\infty} \langle c, w^j y \rangle h_b^j(y; X^{(j)}) dy \right].
\end{aligned}$$

Next, we address the first component, (i)  $\langle c, X^{(j)} \rangle$ .

$$\begin{aligned}
& \sum_{j=1}^n \mathbb{E} \left[ \langle c, X^{(j)} \rangle \frac{\mathbf{1}[g(X^{(j)} + w^j X_j / \sigma_j^2) > b]}{\sum_{k=1}^n e^{\xi_b^k(0) X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0) w_k^j / \sigma_j^2) X_j}} \right] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \langle c, X^{(j)} \rangle \int_{-\infty}^{\infty} \frac{\phi(y / \sigma_j) / \sigma_j}{\sum_{k=1}^n e^{\xi_b^k(0) X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0) w_k^j / \sigma_j^2) y}} \mathbf{1}[g(X^{(j)} + w^j y / \sigma_j^2) > b] dy \right] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \langle c, X^{(j)} \rangle \int_{\sigma_j^2 \xi_b^j(X^{(j)})}^{\infty} \frac{\phi(y / \sigma_j) / \sigma_j}{\sum_{k=1}^n e^{\xi_b^k(0) X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0) w_k^j / \sigma_j^2) y}} dy \right] \\
&= \sum_{j=1}^n \mathbb{E} \left[ \langle c, X^{(j)} \rangle \int_{\sigma_j \xi_b^j(X^{(j)})}^{\infty} \frac{\phi(y)}{\sum_{k=1}^n e^{\xi_b^k(0) X_k^{(j)} - (\xi_b^j(0) - \xi_b^k(0) w_k^j / \sigma_j^2) \sigma_j y}} dy \right] \\
&= \frac{1}{n} \sum_{j=1}^n \mathbb{E} \left[ \langle c, X^{(j)} \rangle \int_{\sigma_j \xi_b^j(X^{(j)})}^{\infty} h_b^j(y; X^{(j)}) dy \right].
\end{aligned}$$

Finally, merging the two derivations above, we arrive at a single expression for the modified estimator, which is

$$\frac{1}{n} \sum_{j=1}^n \mathbb{E} \left[ \int_{\sigma_j \xi_b^j(X^{(j)})}^{\infty} [\langle c, X^{(j)} \rangle + \langle c, w^j y \rangle] h_b^j(y; X^{(j)}) dy \right].$$

□

Here we show the final steps for the unbiasedness of the numerator of our estimator, according to the sign of  $\xi_b^J(X^{(J)})$ .

**Proposition 5.3.1.** *For  $b \in \mathbb{R}$ , we have*

$$\mathbb{E}\eta_b = \mathbb{E}[\langle c, X \rangle; g(X) > b].$$

*Proof.* Here we follow the same reasoning used in estimating  $\int_{\xi}^{\infty} h(y)dy$  as given previously in Section 4.1.2, but modified for this chapter, so that the function  $h(y)$  inside the integral is scaled by the terms  $\langle c, X^{(J)} \rangle$  and  $\langle c, w^J y \rangle$ . Using these results in conjunction with Lemma 5.3.1, shows that

$$\begin{cases} \int_{\sigma_J \xi_b^J(X^{(J)})}^{\infty} [\langle c, X^{(J)} \rangle + \langle c, w^J y \rangle] h_b^J(y; X^{(J)}) dy & \xi_b^J(X^{(J)}) \geq 0 \\ \int_{-\infty}^{\infty} [\langle c, X^{(J)} \rangle + \langle c, w^J y \rangle] h_b^J(y; X^{(J)}) dy - \int_{\sigma_J \xi_b^J(X^{(J)})}^{\infty} [\langle c, X^{(J)} \rangle + \langle c, w^J y \rangle] h_b^J(y; X^{(J)}) dy & \xi_b^J(X^{(J)}) < 0 \end{cases}$$

is an unbiased estimator for  $\mathbb{E}[\langle c, X \rangle; g(X) > b]$ . □

## 5.4 Numerical experiments.

### 5.4.1 Overview.

Here we take  $X$  a general Gaussian random vector and assess the joint expectation which is the numerator of our estimator, with the numerator represented by the expression  $\eta_b$ , using a linear functional  $f(X)$  and also a max functional for  $g(X)$ . This expression is

$$\begin{aligned} \mathbb{E}\eta_b &= \mathbb{E}[f(X); g(X) > b] \\ &= \mathbb{E}\left[\langle c, X \rangle; \max_i X_i > b\right]. \end{aligned}$$

For purposes of studying the expected shortfall, we vary the constants  $c$  as well as change the correlation structure between the marginals  $X_i$ . In turn, we can also compare risk measures computed using our method to traditional methods. For these comparisons, we take  $g(X)$  to be the max functional, while traditional methods take  $g(X)$  to be a linear function of the marginals  $X_i$ , for which analytical solutions exist.

For validation purposes, we use the known values for the univariate case where  $X \sim \mathcal{N}(0, 1)$ , given by

$$\mathbb{E}[X; X > b] = \phi(b).$$

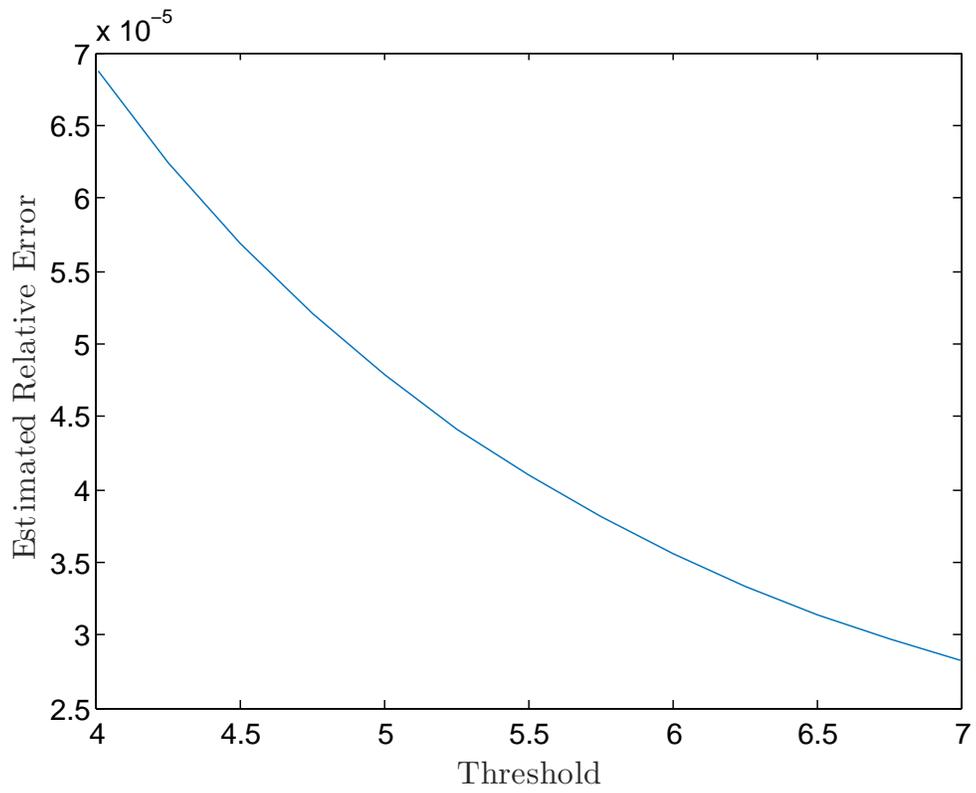
Below, we present some initial tests of our algorithm for varying covariance structures.

#### 5.4.2 Perfect correlation study.

In this first example, we take a simple case where all constants  $c_i$  are set equal to 1. We use Monte Carlo parameters with  $n = 10^6$  replications and  $m = 16$  Gaussian marginals, assuming pairwise correlations. Because we assume perfect correlation between the marginals  $X_i$ , a multi-dimensional setting collapses to a univariate setting for which the test threshold is known.

**Table 5.1:** Perfect Correlation Study:  $\mathbb{E}[X_1; \max(X_1, \dots, X_{16}) > b]$

| Threshold | Sample Mean | True Values | Std. Error | Relative Error |
|-----------|-------------|-------------|------------|----------------|
| 4         | 1.339E-04   | 1.338E-04   | 9.102E-09  | 6.796E-05      |
| 4.5       | 1.60E-05    | 1.60E-05    | 8.969E-10  | 5.61E-05       |
| 5         | 1.49E-06    | 1.49E-06    | 6.989E-11  | 4.70E-05       |
| 5.5       | 1.08E-07    | 1.08E-07    | 4.298E-12  | 3.99E-05       |
| 6         | 6.08E-09    | 6.08E-09    | 2.082E-13  | 3.42E-05       |
| 6.5       | 2.67E-10    | 2.67E-10    | 7.933E-15  | 2.97E-05       |
| 7         | 9.15E-12    | 9.13E-12    | 2.375E-16  | 2.60E-05       |



**Figure 5.1:** Perfect Correlation Study: Estimated Relative Error

### 5.4.3 Bivariate study.

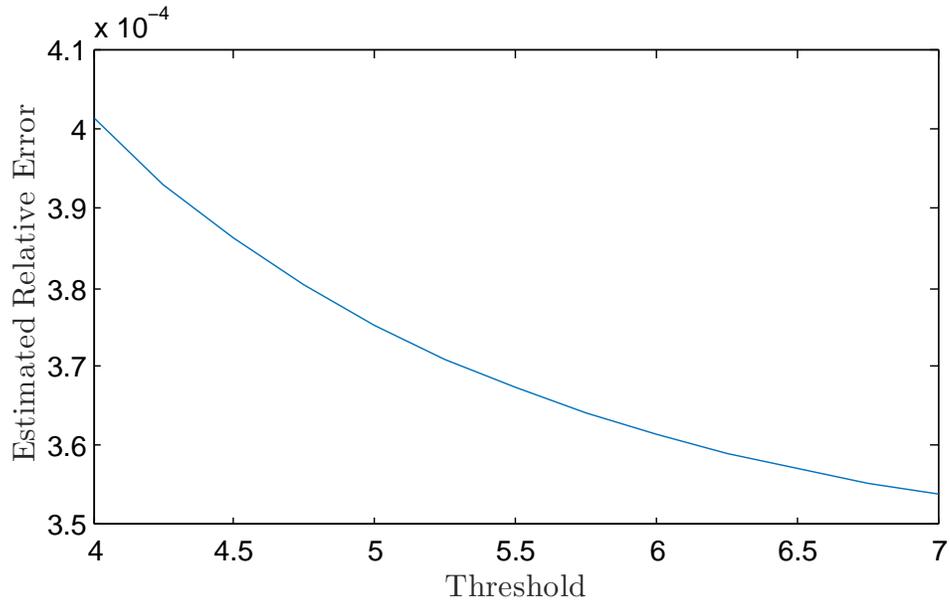
In this new case, we take a hypothetical portfolio  $L = \sum_{i=1}^2 c_i X_i$ , with  $c_i = 1$ , but assume the correlation between the marginals is equal to 0.5. We then compute the new conditional loss  $L$  for thresholds  $b \in \{4, 4.5, 5, \dots, 6.5\}$ , using Monte Carlo parameters  $n = 10^6$  replications and  $m = 2$  Gaussian marginals. For a bivariate normal vector, with standard marginals, the true value may be computed according to the formula,

$$\mathbb{E}[X_1; X_1 + X_2 > b] = \phi(b) \cdot [1 + \text{Cov}(X_1, X_2)].$$

For the graph in 5.3 depicting how the joint expectation changes for varying covariances, a threshold of  $b = 6$  is used.

**Table 5.2:** Bivariate Study:  $\mathbb{E}[X_1; \max(X_1, X_2) > b]$

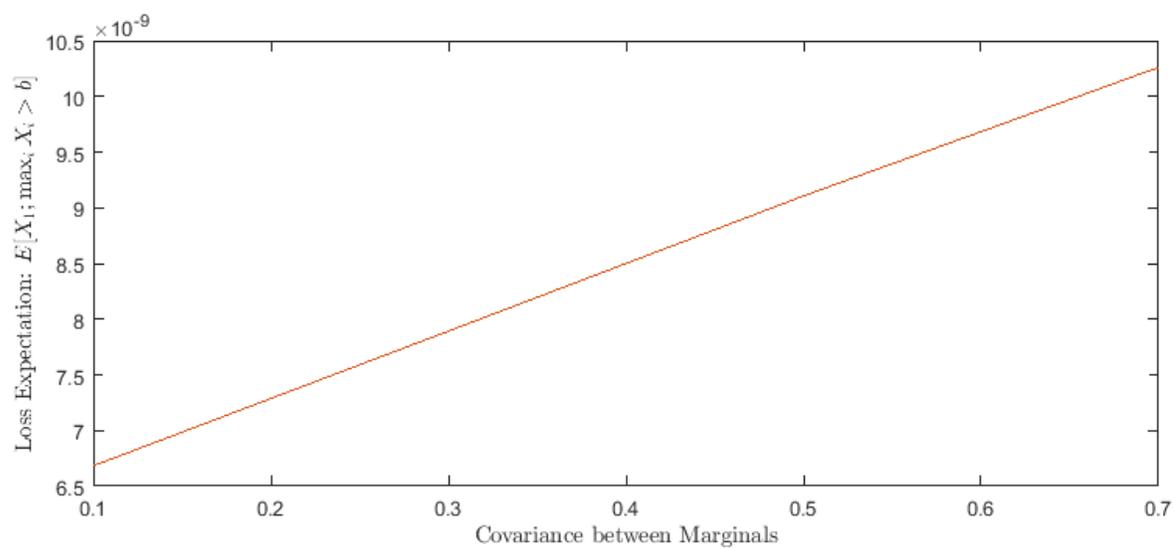
| Threshold | Sample Mean | True Values | Std. Error | Relative Error |
|-----------|-------------|-------------|------------|----------------|
| 4         | 1.984E-04   | 2.007E-04   | 2.514E-07  | 1.268E-03      |
| 4.5       | 2.38E-05    | 2.40E-05    | 2.900E-08  | 1.22E-03       |
| 5         | 2.22E-06    | 2.23E-06    | 2.632E-09  | 1.18E-03       |
| 5.5       | 1.61E-07    | 1.62E-07    | 1.874E-10  | 1.16E-03       |
| 6         | 9.10E-09    | 9.11E-09    | 1.040E-11  | 1.14E-03       |
| 6.5       | 4.00E-10    | 4.00E-10    | 4.516E-13  | 1.13E-03       |
| 7         | 1.37E-11    | 1.37E-11    | 1.532E-14  | 1.12E-03       |



**Figure 5.2:** Bivariate Study, 0.5 Correlation: Estimated Relative Error

**Table 5.3:** Covariance Table for Monte Carlo Estimates

| Cases:                 | Problem: $\mathbb{E}[X_1; \max(X_1, X_2) > 6]$ |             |            |                |
|------------------------|--|-------------|------------|----------------|
| $\text{Cov}(X_1, X_2)$ | Sample Means                                   | True Values | Std. Error | Relative Error |
| 0.1                    | 6.707E-09                                      | 6.683E-09   | 1.785E-11  | 2.662E-03      |
| 0.2                    | 7.283E-09                                      | 7.291E-09   | 1.598E-11  | 2.195E-03      |
| 0.3                    | 7.898E-09                                      | 7.899E-09   | 1.414E-11  | 1.791E-03      |
| 0.4                    | 8.485E-09                                      | 8.506E-09   | 1.226E-11  | 1.445E-03      |
| 0.5                    | 9.099E-09                                      | 9.114E-09   | 1.047E-11  | 1.151E-03      |
| 0.6                    | 9.694E-09                                      | 9.721E-09   | 8.538E-12  | 8.807E-04      |
| 0.7                    | 1.027E-08                                      | 1.033E-08   | 6.678E-12  | 6.501E-04      |



**Figure 5.3:** Monte Carlo Estimates versus Covariance, Threshold  $b = 6$

## BIBLIOGRAPHY

- [1] ADLER, R. J., BLANCHET, J. H., AND LIU, J. Efficient Monte Carlo for high excursions of Gaussian random fields. Ann. Appl. Probab. 22, 3 (2012), 1167–1214.
- [2] ADLER, R. J., SAMORODNITSKY, G., AND TAYLOR, J. E. Excursion sets of three classes of stable random fields. Adv. in Appl. Probab. 42, 2 (2010), 293–318.
- [3] ADLER, R. J., AND TAYLOR, J. E. Random fields and geometry. Springer Monographs in Mathematics. Springer, New York, 2007.
- [4] ALBRECHT, P. Bericht zur Prüfung im Oktober 2007 über Grundprinzipien der Versicherungs- und Finanzmathematik (Grundwissen). Bl. DGVFM 29, 1 (2008), 131–136.
- [5] ASMUSSEN, S., AND GLYNN, P. W. Stochastic simulation: algorithms and analysis, vol. 57 of Stochastic Modelling and Applied Probability. Springer, New York, 2007.
- [6] BARDOU, O., FRIKHA, N., AND PAGÈS, G. Computing VaR and CVaR using stochastic approximation and adaptive unconstrained importance sampling. Monte Carlo Methods Appl. 15, 3 (2009), 173–210.
- [7] BARDOU, O., FRIKHA, N., AND PAGÈS, G. Recursive computation of value-at-risk and conditional value-at-risk using MC and QMC. In Monte Carlo and quasi-Monte Carlo methods 2008. Springer, Berlin, 2009, pp. 193–208.
- [8] BLANCHET, J., GLYNN, P., AND LEDER, K. On Lyapunov inequalities and subsolutions for efficient importance sampling. ACM Trans. Model. Comput. Simul. 22, 3 (2012), Art. 13, 27.
- [9] BOTEV, Z. I., AND KROESE, D. P. Efficient Monte Carlo simulation via the generalized splitting method. Stat. Comput. 22, 1 (2012), 1–16.
- [10] DIEKER, A. B., AND MANDJES, M. On spectral simulation of fractional Brownian motion. Probab. Engrg. Inform. Sci. 17, 3 (2003), 417–434.
- [11] DIEKER, A. B., AND MANDJES, M. On asymptotically efficient simulation of large deviation probabilities. Adv. in Appl. Probab. 37, 2 (2005), 539–552.
- [12] DIEKER, A. B., AND YAKIR, B. On asymptotic constants in the theory of Gaussian processes. Bernoulli 20 (2014), 1600–1619.
- [13] FABOZZI, F. J., STOYANOV, S. V., AND RACHEV, S. T. Computational aspects of portfolio risk estimation in volatile markets: a survey. Stud. Nonlinear Dyn. Econom. 17, 1 (2013), 103–120.

- [14] FABOZZI, F. J., AND TUNARU, R. On risk management problems related to a coherence property. Quant. Finance 6, 1 (2006), 75–81.
- [15] FÖLLMER, H., AND KNISPEL, T. Entropic risk measures: coherence vs. convexity, model ambiguity, and robust large deviations. Stoch. Dyn. 11, 2-3 (2011), 333–351.
- [16] FUH, C.-D., HU, I., HSU, Y.-H., AND WANG, R.-H. Efficient simulation of value at risk with heavy-tailed risk factors. Oper. Res. 59, 6 (2011), 1395–1406.
- [17] GIESECKE, K., KAKAVAND, H., MOUSAVI, M., AND TAKADA, H. Exact and efficient simulation of correlated defaults. SIAM J. Financial Math. 1, 1 (2010), 868–896.
- [18] GLASSERMAN, P. Risk horizon and rebalancing horizon in portfolio risk measurement. Math. Finance 22, 2 (2012), 215–249.
- [19] GLASSERMAN, P., HEIDELBERGER, P., AND SHAHABUDDIN, P. Portfolio value-at-risk with heavy-tailed risk factors. Math. Finance 12, 3 (2002), 239–269.
- [20] GLASSERMAN, P., HEIDELBERGER, P., SHAHABUDDIN, P., AND ZAJIC, T. A large deviations perspective on the efficiency of multilevel splitting. IEEE Trans. Automat. Control 43, 12 (1998), 1666–1679.
- [21] GLASSERMAN, P., HEIDELBERGER, P., SHAHABUDDIN, P., AND ZAJIC, T. Multilevel splitting for estimating rare event probabilities. Oper. Res. 47, 4 (1999), 585–600.
- [22] GLASSERMAN, P., AND KANG, W. OR forum—design of risk weights. Oper. Res. 62, 6 (2014), 1204–1221.
- [23] GLASSERMAN, P., KANG, W., AND SHAHABUDDIN, P. Large deviations in multifactor portfolio credit risk. Math. Finance 17, 3 (2007), 345–379.
- [24] GLASSERMAN, P., KANG, W., AND SHAHABUDDIN, P. Fast simulation of multifactor portfolio credit risk. Oper. Res. 56, 5 (2008), 1200–1217.
- [25] GLASSERMAN, P., AND WANG, Y. Counterexamples in importance sampling for large deviations probabilities. Ann. Appl. Probab. 7, 3 (08 1997), 731–746.
- [26] GLASSERMAN, P., AND XU, X. Robust risk measurement and model risk. Quant. Finance 14, 1 (2014), 29–58.
- [27] GLYNN, P. W., AND IGLEHART, D. L. Importance sampling for stochastic simulations. Management Sci. 35, 11 (1989), 1367–1392.
- [28] GLYNN, P. W., AND WHITT, W. The asymptotic efficiency of simulation estimators. Oper. Res. 40, 3 (1992), 505–520.
- [29] HASHORVA, E., AND HÜSLER, J. Extremes of Gaussian processes with maximal variance near the boundary points. Methodol. Comput. Appl. Probab. 2, 3 (2000), 255–269.

- [30] HASHORVA, E., AND HÜSLER, J. The neighbourhood of bivariate maxima. Rend. Circ. Mat. Palermo (2) Suppl., 70, part I (2002), 361–376. IV International Conference in “Stochastic Geometry, Convex Bodies, Empirical Measures & Applications to Engineering Science”, Vol. I (Tropea, 2001).
- [31] HASHORVA, E., AND HÜSLER, J. On multivariate Gaussian tails. Ann. Inst. Statist. Math. 55, 3 (2003), 507–522.
- [32] HASHORVA, E., AND HÜSLER, J. Multiple maxima in multivariate samples. Statist. Probab. Lett. 75, 1 (2005), 11–17.
- [33] HONG, L. J., HU, Z., AND LIU, G. Monte Carlo methods for value-at-risk and conditional value-at-risk: a review. ACM Trans. Model. Comput. Simul. 24, 4 (2014), Art. 22, 37.
- [34] HONG, L. J., HU, Z., AND ZHANG, L. Conditional value-at-risk approximation to value-at-risk constrained programs: a remedy via Monte Carlo. INFORMS J. Comput. 26, 2 (2014), 385–400.
- [35] HUANG, D., FABOZZI, F. J., AND FUKUSHIMA, M. Robust portfolio selection with uncertain exit time using worst-case VaR strategy. Oper. Res. Lett. 35, 5 (2007), 627–635.
- [36] HUANG, D., ZHU, S.-S., FABOZZI, F. J., AND FUKUSHIMA, M. Portfolio selection with uncertain exit time: a robust CVaR approach. J. Econom. Dynam. Control 32, 2 (2008), 594–623.
- [37] IYENGAR, S. Evaluation of normal probabilities of symmetric regions. SIAM J. Sci. Statist. Comput. 9, 3 (1988), 418–423.
- [38] KIM, J. H., KIM, W. C., AND FABOZZI, F. J. Recent developments in robust portfolios with a worst-case approach. J. Optim. Theory Appl. 161, 1 (2014), 103–121.
- [39] L’ECUYER, P., MANDJES, M., AND TUFFIN, B. Importance sampling in rare event simulation. In Rare event simulation using Monte Carlo methods. Wiley, Chichester, 2009, pp. 17–38.
- [40] LIU, J., AND XU, G. Some asymptotic results of Gaussian random fields with varying mean functions and the associated processes. Ann. Statist. 40, 1 (2012), 262–293.
- [41] LIU, J., AND XU, G. On the density functions of integrals of Gaussian random fields. Adv. in Appl. Probab. 45, 2 (2013), 398–424.
- [42] LIU, J., AND XU, G. On the conditional distributions and the efficient simulations of exponential integrals of Gaussian random fields. Ann. Appl. Probab. 24, 4 (2014), 1691–1738.
- [43] NGUYEN, Q. H., AND ROBERT, C. Y. New efficient estimators in rare event simulation with heavy tails. J. Comput. Appl. Math. 261 (2014), 39–47.

- [44] TAYLOR, J. E., AND ADLER, R. J. Gaussian processes, kinematic formulae and Poincaré's limit. Ann. Probab. 37, 4 (2009), 1459–1482.
- [45] VANDEN-EIJNDEN, E., AND WEARE, J. Rare event simulation of small noise diffusions. Comm. Pure Appl. Math. 65, 12 (2012), 1770–1803.
- [46] YAMAI, Y., AND YOSHIBA, T. Comparative analyses of expected shortfall and value-at-risk. J. Oper. Res. Soc. Japan 45, 4 (2002), 490–506. Theory, methodology and applications in financial engineering.