

國立台灣大學電機資訊學院資訊工程學研究所
博士論文

Department of Computer Science and Information Engineering
College of Electrical Engineering and Computer Science
National Taiwan University
Doctoral Dissertation

蒙特卡羅方法之罕見事件及其應用於電路模擬
Rare Events in Monte Carlo Methods
with Applications to Circuit Simulation

麥陶德
Todd G. McKenzie

指導教授：許永真 博士
Advisor: Jane Yung-jen Hsu, Ph.D.
中華民國 108 年 7 月
July, 2019

Abstract

The simulation of rare events in Monte Carlo is of practical importance across many disciplines including circuit simulation, finance, and metrology to name a few. To make inferences about the behavior of distributions in systems which have a likelihood of 1 in billions, simulation by traditional Monte Carlo becomes impractical. Intuitively, Monte Carlo samples which are likely (i.e. samples drawn from regions of higher probability density) have little influence on the tail distribution which is associated with rare events. This work provides a novel methodology to directly sample rare events in input multivariate random vector space as a means to efficiently learn about the distribution tail in the output space. In addition, the true form of the Monte Carlo simulation is modeled by first linear and then quadratic forms. A systematic procedure is developed which traces the flow from the linear or quadratic modeling to the computation of distribution statistics such as moments and quantiles directly from the modeling form itself. Next, a general moment calculation method is derived based on the distribution quantiles where no underlying linear or quadratic model is assumed. Finally, each of the proposed methods is grounded in practical circuit simulation examples. Overall, the thesis provides several new methods and approaches to tackle some challenging problems in Monte Carlo simulation, probability distribution modeling, and statistical analysis.

Acknowledgements

First, I would like to thank my mentors for their constant encouragement and insight: my advisor at National Taiwan University, Jane Hsu, as well as Kishore Singhal, my manager at Synopsys. It is always hard to predict where exactly a research program will lead and I am grateful to have had extended time to explore many areas and gain that broadened experience. Jane was always very supportive throughout, even as my research drifted away from its initial direction. She encouraged me to work on ambitious problems, and to stick with them despite hardships and see them through to the end. I thank Kishore for spending considerable time to mentor me in many areas of statistics and in state-of-the-art Monte Carlo methods. The stream of shared insights (both practical and philosophical) supplied by Kishore over time through weekly technical discussions is something I now find difficult to imagine being without. I would like to thank Weidong Liu, who served as my previous manager at Synopsys, for invaluable guidance in many areas but most relevant to the thesis for suggesting the initial idea of focusing on research topics related to Monte Carlo simulation. I thank Frank Lee, VP of Engineering at Synopsys, for his strong support of the PhD research and for making it possible to utilize a portion of my time as well as other resources to conduct the research. Going back further, I also thank Scott Springer at IBM for his support of my studies and for granting an extended leave from my then role at IBM Microelectronics in Vermont to pursue

studies at NTU in Taipei during the early years of the PhD studies. Without the guidance and support of these folks, the thesis would not have been possible.

Though the thesis scope does not directly include machine learning, the inspiration behind the developed algorithms is certainly related to ML. I was fortunate enough to be exposed to memorable experiences in machine learning relatively early in my academic career with an automated VLSI design course taught by Sung-Kyu Lim at Georgia Tech as an undergraduate and later in a machine learning course taught by Andrew Ng at Stanford as a graduate student. It was not until my PhD studies were underway at National Taiwan University, though, that I was able to delve deeper into machine learning and become more proficient. Taking courses from Chih-Jen Lin, Hsuan-Tien Lin, and Shou-De Lin at NTU provided an excellent grounding in all things machine learning from Support Vector Machines and Convex Optimization to Probabilistic Graphical Models and Low Rank Approximation. Beyond the coursework, there were also international competitions such as the KDDCup where the theory could be immediately put to practice and these exercises led to many successful results[46, 32, 6]. Special thanks to Shou-De Lin for playing an important role in the early stages of my PhD research. In those first few years I explored a range of pure machine learning topics and though the final thesis topic ultimately departed from this earlier work, from a learning perspective the time with Shou-De was both valuable and fruitful[29].

When I initially decided to pursue the PhD degree in Taiwan as a non-native Chinese speaker, I knew there would be some interesting challenges both in and out of the classroom. Now that I find myself near the other end of this goal, I have

a small army of Chinese language teachers I wish to thank. To my first Chinese language teacher Xiao-Liang Li at Georgia Tech, who helped spark my passion for the Chinese language, and to my many teachers at the Mandarin Training Center at National Taiwan Normal University and at the International Chinese Language Program at National Taiwan University, I owe a great debt. At ICLP, I would especially like to thank Li-Yuan Chen, Zhou Chang-Zhen, Mei-Yuan Fan, Xin-Chun Chen, Rou-Yu Shen, Zhi-Cheng Xu, and Jing-Jing Chai.

I would like to sincerely thank my parents, Gordon and Linda McKenzie, who were supportive of me in my decision to pursue long term studies on the other side of the globe. I thank my ever understanding wife, Shu-Fen, who has shown me so much support over the course of my PhD studies, she is truly much more than I deserve. Finally I would like to thank my two boys, Connor and Logan, born in the 2nd and 4th years of my PhD study, for reminding me of what is really important in life and for being such wonderful sources of motivation and inspiration.

Contents

Abstract	i
Acknowledgements	iii
List of Figures	xii
List of Tables	xvi
List of Algorithms	xviii
Chapter 1 Introduction	1
1.1 Problem Statement	1
1.2 Motivation and Scope	1
1.3 Related Work	4
1.3.1 Importance Sampling	5
1.3.2 Statistical Blockade	6
1.3.3 Discussion	7

1.4 Thesis Organization	7
Chapter 2 Background : Random Variables and their Statistics 9	
2.1 Notation	9
2.2 Random Variable Probability Distribution	10
2.3 Expectations	12
2.3.1 Univariate Expectation	12
2.3.2 Multivariate Expectation	13
2.4 Moments	14
2.5 Order statistics	15
2.6 Expectation of Gaps between Order Statistics	15
2.6.1 Standard Moments	16
2.6.2 L-Moments	17
2.7 Sample Quantiles	20
2.8 Confidence Intervals for Binomial Proportions	20
2.9 QQ Plots	21
2.10 Error Function	24
2.11 Selected Probability Distributions	25
2.11.1 Normal distribution	26
2.11.2 Chi-squared distribution	27
2.11.3 Beta distribution	28
2.11.4 Log-normal distribution	29

2.11.5	Skew Normal distribution	30
2.11.6	Weibull distribution	32
2.11.7	Student's t distribution	33
Chapter 3	Monte Carlo Sampling Techniques	34
3.1	Monte Carlo Simulation	34
3.2	Motivation for Rare Sample Simulation	35
3.3	Linear Forms of Standard Normal Random Vectors	37
3.3.1	Empirical Verification	38
3.4	Statistics of the Linear Form	41
3.5	Monte Carlo Sampling for Machine Learning	41
3.5.1	Uniform Sampling in High Dimensions	44
3.5.2	Hyperspherical Uniform Volumetric Sampling	47
3.6	Rejection Sampling	50
3.7	Generating List of Random Numbers in Sorted Order	53
3.8	Inverse Transform Sampling	55
3.9	In-order Multivariate Sampling	56
3.10	Extensions	63
3.10.1	Correlated Sampling	63
3.10.2	Hyperspherical Non-Uniform Random Sampling	65
3.10.3	Directional Sampling	67
3.11	Discussion	71

Chapter 4 Quadratic Forms of the Multivariate Standard Normal	73
4.1 Quadratic Forms	74
4.2 Quadratic Modeling with Monomial Polynomials	74
4.3 Quadratic Modeling with Legendre Polynomials	75
4.4 Linearization via Eigenvector Decomposition	78
4.5 Raw Moments	79
4.6 Central Moments	80
4.7 Cumulants	80
4.8 Quantiles	81
4.9 Toy Example	82
4.10 Quadratic Modeling Examples	86
4.10.1 Univariate Linear	87
4.10.2 Univariate Quadratic	88
4.10.3 Multivariate Quartic	89
4.10.4 Multivariate Exponential and Quadratic	90
4.10.5 Quadratic Modeling Example Observations	91
4.11 Discussion	91
Chapter 5 Quantile-Based Moments	93
5.1 Computing Moments from Quantile Information	93
5.2 Experiments	96
5.3 Discussion	109

Chapter 6 Applications	110
6.1 Circuits	111
6.2 Monotonicity Verification	111
6.3 In-Order Multivariate Sampling Experiments	116
6.4 Linear and Quadratic Modeling with Direct Model-Based Quantile Computation	119
6.5 Quantile-Based Moment Experiments	120
6.6 Discussion	123
Chapter 7 Conclusions	125
7.1 Connections Between Our Proposed Methods	125
7.2 Connections of Our Proposed Methods with Related Work	126
7.2.1 Importance Sampling	126
7.2.2 Statistical Blockade	127
7.3 Thesis Contributions	127
7.4 Future Work	129
Bibliography	131
Appendix A Integral of High Order Error Function	1
Appendix B Operational Amplifier Netlist Listing	4

List of Figures

2.1	Standard Normal Distribution Density	11
2.2	Standard Normal Distribution Cumulative Density	12
2.3	Standard Normal Distribution QQ Plot	21
2.4	Example Production QQ Plot	23
2.5	Normal Distribution Density and QQ Plot	27
2.6	Chi-square Distribution Density and QQ Plot	28
2.7	Beta Distribution Density and QQ Plot	29
2.8	Log-normal Distribution Density and QQ Plot	30
2.9	Skew Normal Distribution Density and QQ Plot with $\xi = 0$ and $\omega = 1$	31
2.10	Weibull Distribution Density and QQ Plot	32
2.11	Student's t Distribution Density and QQ Plot	33
3.1	One and two dimensional density plots of 1M normal samples with mean=0 and standard deviation=1	35
3.2	Multivariate Normal Conditioned on Vector Norm	40

3.3	One and two dimensional density plots of 1M normal samples with mean=0 and standard deviation=2	42
3.4	One and two dimensional density plots of 1M uniform samples on the interval [-4,4]	43
3.5	2D Projections of 5D standard normal and uniform distributions . . .	44
3.6	One dimensional projections of the normal and uniform distributions drawn from a 300 dimensional space	45
3.7	The distance from the origin of the normal and uniform distributions drawn from a 300 dimensional space	46
3.8	The effective value of sigma of the normal and uniform distributions drawn from a 300 dimensional space	47
3.9	One and two dimensional density plots of 1M uniform samples on the interval [-4,4]	49
3.10	Two dimensional scatter plots of 1K samples of normal, uniform, and Hyperspherical Uniform Volumetric samples	50
3.11	Distance to Origin of 2D Normal Samples Unordered and Ordered by Distance to Origin	52
3.12	In-order Multivariate Sampling	58
3.13	In-order Multivariate Sampling	59
3.14	Hypersphere radius as a function of the number of IRVs and proba- bility in terms of σ	62

3.15 Correlated sampling with Hyperspherical Uniform Random Sampling and In-order Multivariate Sampling	65
3.16 Hyperspherical Non-uniform Volumetric Sampling distributions for k=4,0.5,0,-0.5	67
3.17 Samples from the von Mises-Fisher distribution on the unit circle for $\kappa = 1, 2, 5, 15$ in direction $\mu = (\sqrt{2}/2, \sqrt{2}/2)$	69
3.18 Angular densities from the von Mises-Fisher Distribution on the unit circle for $\kappa = 1, 2, 5, 15$ in direction $\mu = (\sqrt{2}/2, \sqrt{2}/2)$	70
4.1 Plots of the first six Legendre polynomials	76
4.2 Sample data of univariate linear form compared with analytic quantiles of modeled linear and quadratic forms	87
4.3 Sample data of univariate quadratic form compared with analytic quantiles of modeled linear and quadratic forms	88
4.4 Sample data of multivariate quartic form compared with analytic quantiles of modeled linear and quadratic forms	89
4.5 Sample data of multivariate exponential and quadratic form compared with analytic quantiles of modeled linear and quadratic forms .	90
5.1 Moment Trials Density : Normal ($\mu = 1$)	101
5.2 Moment Trials QQ Plot and Delta QQ Plot : Normal ($\mu = 1$)	102
5.3 Moment Trials Density : Chi-square ($k = 3$)	103
5.4 Moment Trials QQ Plot and Delta QQ Plot : Chi-square ($k = 3$) . . .	104

5.5	Moment Trials Density : Student's t ($\nu = 4$)	105
5.6	Moment Trials QQ Plot and Delta QQ Plot : Student's t ($\nu = 4$) . .	106
5.7	Moment Trials Density : Weibull ($\lambda = 1, k = 1$)	107
5.8	Moment Trials QQ Plot and Delta QQ Plot : Weibull ($\lambda = 1, k = 1$) .	108
6.1	Monotonicity plots of delay measurement of an inverter circuit in 16nm technology	113
6.2	Monotonicity plots of delay measurement of a MAJ circuit in 16nm technology	114
6.3	Monotonicity plots of delay measurement of a operational amplifier circuit in 0.2 μ m technology	115
6.4	Origin Distance versus Measure Percentile : 7nm Inverter Delay . . .	117
6.5	Origin Distance versus Measure Percentile : 16nm AOI Delay	118
6.6	Origin Distance versus Measure Percentile : 16nm CKGTPLT Measure	119
6.7	QQ plots of MC sample data versus quantiles of linear and quadratic models extracted directly : operational amplifier in 0.2 μ m technology	120
6.8	QQ plots of two delay measurements of an inverter circuit in 7nm technology	122

List of Tables

2.1	Symbol Definitions	9
2.2	Variable Notation	10
3.1	The effect of non-uniform distribution radial density parameter r on resulting sample distribution	66
4.1	Sample versus analytic raw moments : Booth function	84
4.2	Sample versus analytic central moments : Booth function	84
4.3	Normalized central moments and cumulants : Booth function	85
4.4	Sample versus analytic quantiles : Booth function	86
4.5	Univariate Linear Example True and Modeled Forms	87
4.6	Univariate Quadratic Example True and Modeled Forms	88
4.7	Multivariate Quartic Example True and Modeled Forms	89
4.8	Multivariate Exponential and Quadratic Example True and Modeled Forms	90
5.1	Sample vs. Quantile Based Moment Computation : Mean	97

5.2	Sample vs. Quantile Based Moment Computation : Standard Deviation	98
5.3	Sample vs. Quantile Based Moment Computation : Skewness	99
5.4	Sample vs. Quantile Based Moment Computation : Kurtosis	100
6.1	Circuit Characteristics	111
6.2	Comparison of sample and quantile-based moments for a 7nm inverter circuit	121

List of Algorithms

1	Hyperspherical Uniform Volumetric Sampling Steps (d+1 RVs per sample)	48
2	Volumetric Uniform Sampling Steps (Revised, d RVs per sample)	48
3	Generate Sorted Uniform Random Numbers by Sorting	53
4	Generate Sorted Uniform Random Numbers in Two Passes	54
5	Generate Sorted Uniform Random Numbers in One Pass, Version 1	54
6	Generate Sorted Uniform Random Numbers in One Pass, Version 2	55
7	In-order Multivariate Normal Sampling Steps (Original, d+1 RVs per sample)	57
8	In-order Multivariate Normal Sampling Steps (Revised, d RVs per sample)	57
9	Volumetric Non-uniform Sampling Steps with radial density parameter r	66

Chapter 1

Introduction

1.1 Problem Statement

The problems addressed by this thesis are as follows:

1. How to investigate rare Monte Carlo events given a limited simulation budget?
2. How to model and characterize quadratic forms of multivariate standard normal random variables?
3. How to calculate mean, standard deviation, skewness and kurtosis efficiently from quantile information alone?

1.2 Motivation and Scope

Stochastic process variation plays an ever increasing role in the reliability of semiconductor circuit as scaling descends further into the nanoscale regime advanced by the

two plus year transistor count doubling cadence in dense integrated circuits[35, 5]. Process variation accounts for the naturally occurring variation during the semiconductor manufacturing process such as gate length or width, fin height, and oxide thickness[33]. Also important are sources of systematic variation such as random dopant fluctuation (RDF), random polysilicon crystal orientation, and other like physical effects[16]. The effect of such variation are intensified by transistor scaling due to the fact that the variation impact is roughly inversely proportional to the square root of the transistor area[40].

In this context, two central questions often posed by circuit designers are :

1. Given a percentile of component yield, what is the measure value at the yield boundary? (the “quantile question”)
2. Given a measure specification limit, what percentage of components satisfy that criterion? (the “yield question”)

As these questions must be answered iteratively during the design process by Monte Carlo methods, providing accurate answers is an important part of the circuit design flow. Highly replicated systems and systems that are difficult or impossible to replace such as space applications require extremely low failure rates. The automotive industry, driven by stringent safety standards of ISO 26262[25], also requires the hardware and hardware development systems to meet strict high-sigma requirements. The high-sigma circuit simulation which is needed for these applications can easily become a bottleneck in the circuit design process due to the high number of simulation samples required.

1.2. MOTIVATION AND SCOPE

With the high replication of cell instances for circuits such as static random access memories(SRAMs) and flip-flops, the occurrence of extremely rare events becomes more likely. For a cell which is replicated 10 million times (for example, in a 10-Mb SRAM array), each manufactured instance is slightly different due to manufacturing variation though they are identical by design. Because there is a built in design margin, most of the cell instances will meet specifications despite this variation. If the desired chip yield is 99%, the cell yield requirement would be 99.999998995%, which is roughly 6-sigma. In order to simulate just 5 failing samples at such a yield rate, a Monte Carlo simulation of 4.48 billion samples would be required. If the simulation of one sample takes just 1 second, then 51.9 days would be required on 1000 cores to complete the full simulation. As is evident, it is difficult to accurately determine yield rates under such conditions in a reasonable amount of time. This thesis explores several methods to simulate rare samples, including a novel method called In-Order Multivariate Sampling.

Another area of research covered in the thesis is the modeling of vectors of standard normal random numbers. Such vectors are used extensively in semiconductor simulation systems to model process and physical variation. While linear forms of multivariate standard normals are well understood and often used, the quadratic form is used much less frequently, despite known treatment in computational finance[20] which has not yet become widespread in circuit simulation. The thesis aims to bridge this gap by integrating and extending relevant techniques into a coherent flow for quadratic form utilization in the circuit simulation domain. The flow includes the fitting of the quadratic form utilizing sparse orthogonal polyno-

mial models through the extraction of statistical properties such as moments and quantiles directly from the quadratic form without sampling.

The calculation of moments is a common task when studying complex systems. When the linear or quadratic form is available, the moments can be directly calculated from the model as discussed. Moreover, in the case when sample data is given, then the moments can be calculated from the sample data in the standard way. However, if the CDF of the given probability function is theoretically derived, only the information of a QQ plot is provided, or the sample data is otherwise unavailable, can quantiles be used to calculate moments? The thesis provides a solution in the form of a novel method for moment calculation given quantile information only.

In the thesis, the topics outlined above are first studied from a theoretical perspective and with the results of simplified examples to build intuition. These findings are then grounded in actual circuit simulation results which show the applicability of the methods to real world problems.

1.3 Related Work

High Sigma Monte Carlo is an important area of research related to rare events in Monte Carlo simulation. Examples of High Sigma Monte Carlo methods include, for example, Importance Sampling and Statistical Blockade. In Sections 1.3.1 and 1.3.2, we will describe these methods then later in Sections 7.2.1 and 7.2.2 the connection between the methods proposed in the thesis and these related works.

1.3. RELATED WORK

1.3.1 Importance Sampling

Importance Sampling[41] is a method of variance reduction that can be used in the Monte Carlo method. In the context of simulation, Importance Sampling is often used in conjunction with the multivariate normal distribution as a way to scale or translate the input distribution.

For the scaling approach, the random variable X is multiplied by a constant c with $c > 1$. The alternative density function f_{scaled} evaluated at a sample x from X is given by

$$f_{scaled}(x) = \frac{1}{c} f\left(\frac{x}{c}\right) \quad (1.1)$$

For the translation approach, the amount of shift represented by the constant t is chosen as to minimize the variance of the importance sampling estimator.

$$f_{translated}(x) = f(x - t), t > 0 \quad (1.2)$$

where t is chosen as to minimize the variance of the importance sampling estimator.

In either case, deciding the constant which minimizes the variance of the importance sampling estimator becomes a sub-problem which itself requires training data to determine effectively. These approaches also lack the ability to be verified easily using a full Monte Carlo simulation as the samples become different.

1.3.2 Statistical Blockade

The idea of filtering samples based on a surrogate model is captured well in the work of Singhee and Rutenbar known as Statistical Blockade[43]. The basic idea is focus simulation in the tail regions of the measure distribution, avoiding the simulation of the majority of Monte Carlo samples which do not lie in the tail regions. An outline of the steps of Statistical Blockade method is summarized as follows:

1. Simulate training samples
2. Build surrogate model based on training samples
3. Predict results of large Monte Carlo sample set
4. Filter predictions of large Monte Carlo sample set (retain only predictions in the tail)
5. Simulate predicted tail samples
6. Summarize tail sample statistics (quantile extraction or yield computation)

Though the Statistical Blockade method can be useful to investigate the distribution tails, there are several challenges associated with the method. First, an accurate surrogate model is critically important for the results to be accurate. If the surrogate model is not accurate, then the identification of the tail samples for the final simulation step will be incorrect. If the correct tails samples are not correctly identified, then the results yield or quantile statistics will be erroneous. Another challenge associated with the Statistical Blockade method is the need to predict all of the Monte Carlo samples in the large sample set. For high sigma, the number of required samples can be easily be in the billions creating a bottleneck in the

1.4. THESIS ORGANIZATION

surrogate model prediction step.

1.3.3 Discussion

The relationships between the related work and the In-Order Multivariate Sampling technique proposed in Chapter 3 are given in Sections 7.2.1 and 7.2.2. The related works associated the Quadratic Form flow defined in Chapter 4 are comprised of the sub-steps of the flow itself and provided in that chapter. The quantile-based moment computation detailed in Chapter 5 has as its reference flow the sample-based moment computation method described in Section 2.6.1.

1.4 Thesis Organization

In Chapter 2, a foundation is established which covers random variable expectation, moments, quantiles, probability distributions, and other topics. The main theme of the thesis centers on the statistics and sampling of random variables, so this chapter lays the groundwork for subsequent chapters.

In Chapter 3, we propose a new algorithm In-Order Multivariate Sampling (IMS) to efficiently study rare events in Monte Carlo simulation associated with distribution tails. Both the “yield question” and the “quantile question” discussed in the previous section are intrinsically linked to the tail distribution of the circuit measures which represent rare events. The IMS method is first motivated by describing the details and challenges of Monte Carlo simulation at higher sigma.

In Chapter 4, the linear representation of the true Monte Carlo function which

was analyzed in Chapter 3 is extended to the quadratic domain. A flow for modeling and characterizing quadratic forms of multivariate normal distributions is presented and reviewed. The results of the characterization flow include the computation of moments, cumulants, and quantiles.

In Chapter 5, a method for calculating moments from quantile information is presented. The method is compared against the sample moment method for several common probability distributions.

In Chapter 6, applications to circuit simulation are presented which tie together and reveal the practical value of methods developed in earlier chapters.

In Chapter 7, the contributions of the thesis are outlined and future work is proposed.

Chapter 2

Background : Random Variables and their Statistics

This chapter aims to provide an introduction to probability distributions and their statistics which will provide a foundation for the rest of the thesis.

2.1 Notation

Table 2.1: Symbol Definitions

Notation	Meaning
$\mathbb{E}[X]$	the expectation of a random variable X
\mathbb{N}	the set of natural numbers $\{1, 2, 3, \dots\}$
$Pr[E]$	probability of event E

Table 2.2: Variable Notation

Notation	Deterministic Variable	Random Variable
\mathbf{X}	-	collection of multivariate random vectors
\mathbf{x}	constant vector	multivariate random vector
X	constant matrix	random scalar
x	constant scalar	-

2.2 Random Variable Probability Distribution

A continuous random variable X has a probability density function (PDF) given by f_X if

$$\Pr[a \leq X \leq b] = \int_a^b f_X(x) \, dx \quad (2.1)$$

where $\Pr[\cdot]$ denotes event probability. The PDF has the property that it is non-negative for all x and the integral under its curve on its support is 1. As an example, the probability density plot for the standard normal distribution (discussed later in Section 2.11.1) is shown in Figure 2.1.

2.2. RANDOM VARIABLE PROBABILITY DISTRIBUTION

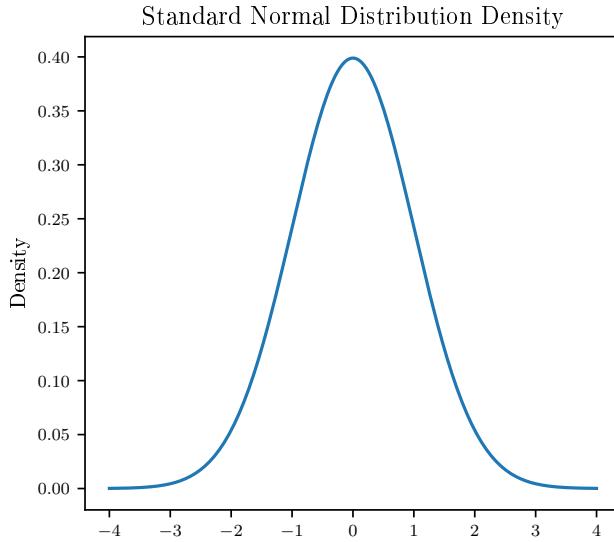


Figure 2.1: Standard Normal Distribution Density

The cumulative probability distribution (CDF) of X is given by

$$F_X(x) = \Pr[X \leq x] = \int_{-\infty}^x f_X(t) dt. \quad (2.2)$$

thus if at x the density f_X is continuous, then

$$f_X(x) = \frac{d}{dx} F_X(x). \quad (2.3)$$

$F_X(x)$ is a weakly monotonically increasing function of X with $0 \geq F(x) \geq 1$ for all x and $F(-\infty) = 0$, $F(\infty) = 1$. The CDF plot for the standard normal distribution is shown in Figure 2.2.

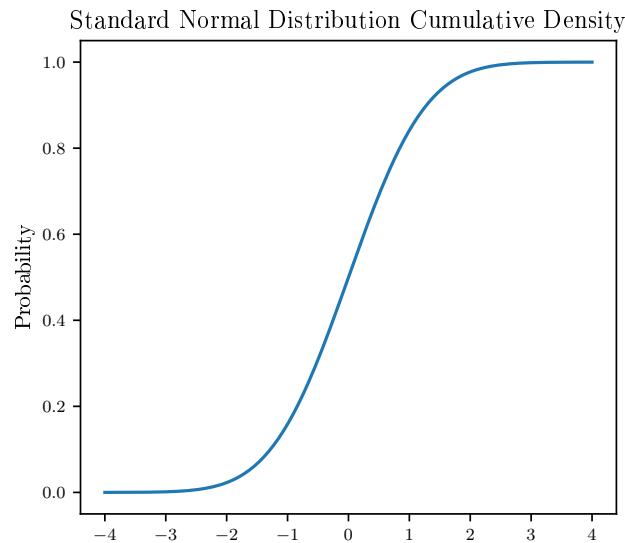


Figure 2.2: Standard Normal Distribution Cumulative Density

2.3 Expectations

2.3.1 Univariate Expectation

The expectation of a random variable X is denoted as $\mathbb{E}(X)$ and given for discrete X by

$$\mathbb{E}(X) = \sum_x f(x) \quad (2.4)$$

and for continuous X by

2.3. EXPECTATIONS

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x) dx \quad (2.5)$$

where $f(x)$ is the PDF of X .

The expectation of a function $g(X)$ can be inferred from the probability distribution of X and the function for discrete X by

$$\mathbb{E}[g(X)] = \sum_x g(x)f(x) \quad (2.6)$$

and for continuous X by

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x) dx . \quad (2.7)$$

2.3.2 Multivariate Expectation

Extending the expectations in the previous section to the multivariate case, the expectation of a multivariate random vector \mathbf{X} is given by

$$\mathbb{E}(\mathbf{X}) = \begin{bmatrix} \mathbb{E}(X_1) \\ \mathbb{E}(X_2) \\ \vdots \\ \mathbb{E}(X_n) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} . \quad (2.8)$$

Given a function $g : \mathbb{R}^n \rightarrow \mathbb{R}$, for discrete \mathbf{X} the expectation may be written as

$$\mathbb{E}[g(\mathbf{X})] = \sum_{\mathbf{x}} g(\mathbf{x})f(\mathbf{x}) \quad (2.9)$$

or for continuous \mathbf{X} the expectation is given by

$$\begin{aligned}\mathbb{E}[g(\mathbf{X})] &= \int_{\mathbb{R}^n} g(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \dots, x_n) f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n.\end{aligned}\tag{2.10}$$

2.4 Moments

Moments, used extensively in physics and mathematics, provide quantitative measures of the shapes of functions. Within the realm of physics, the function may represent a physical density where the zeroth moment is the total mass, the first moment divided by the zeroth moment is the center of mass, and the second moment is the rotational inertia. In a statistical context, probability density functions are considered. In this case, the total probability (always exactly one) is the zeroth moment, the first moment is the mean, the second central moment is the variance, the third standardized moment is the skewness, and the fourth standardized moment is the kurtosis. It can be shown that on bounded intervals, the total set of all moments uniquely determine the distribution[18, 42], however in practice the first four moments are the most widely used.

Moment computation on discrete distributions can be expensive if the dataset size is large. While streaming moment computation is possible in some cases, it is useful to have alternative moment computation methods. Cases where streaming moment computation may not be practical include when datasets or streams are no longer available. In this chapter, both standard and L-moments (also known as

linear moment) will be described. We will derive methods for calculating standard moments from quantile information with some important assumptions in Chapter 5.

2.5 Order statistics

The k th order statistic of a size n sample from \mathbf{x} , denoted $X_{(k)}$, is the k th smallest of X_1, X_2, \dots, X_n iid random variables. Thus $X_{(1)}$ and $X_{(n)}$ are the smallest and largest X , respectively. Meaning that

$$X_{(1)} = \min(X_1, X_2, \dots, X_n) \quad (2.11)$$

$$X_{(n)} = \max(X_1, X_2, \dots, X_n) . \quad (2.12)$$

2.6 Expectation of Gaps between Order Statistics

The general formula for the average distance between the k th and $(k+1)$ th order statistic of a sample drawn from a random distributed variable x with cumulative probability density function $F(x)$ of size n is given by

$$\frac{n}{(n-k)!k!} \int_{-\infty}^{\infty} F(x)^{n-k} (1-F(x))^k dx . \quad (2.13)$$

This formulation by Karl Pearson[39] was a generalized solution to a problem earlier posed by Galton[12].

2.6.1 Standard Moments

Standard moments are commonly defined in three ways : raw, central, and normalized. The n th raw moment is given by

$$\mu'_n = \mathbb{E}[\mathbf{X}^n] = \int_{-\infty}^{\infty} x^n f(x) dx . \quad (2.14)$$

The n th central moments are given by

$$\mu_n = \mathbb{E}[(\mathbf{X} - \mu)^n] = \int_{-\infty}^{\infty} (x - \mu)^n f(x) dx, \mu = \mathbb{E}[\mathbf{X}] . \quad (2.15)$$

The n th normalized (also known as standardized) moment is the n th central moment divided by σ^n (where σ is the standard deviation) as given by

$$\frac{\mu_n}{\sigma^n} = \frac{\mathbb{E}[(\mathbf{X} - \mu)^n]}{\sigma^n} . \quad (2.16)$$

The named properties of distributions including the mean, standard deviation, skewness and kurtosis can also be defined in terms of the raw, central, and normalized moments as follows :

$$\mu = \mu'_0 \quad (\text{mean}) \quad (2.17)$$

$$\sigma = \sqrt{\mu_1} \quad (\text{standard deviation}) \quad (2.18)$$

$$\gamma = \frac{\mu_2}{\sigma^2} \quad (\text{skewness}) \quad (2.19)$$

$$\kappa = \frac{\mu_3}{\sigma^3} \quad (\text{kurtosis}) . \quad (2.20)$$

2.6.2 L-Moments

Introduced by Hosking in 1990, L-moments are so named as they are the construction of linear combinations of order statistics[21, 22] and are measures of the location, scale, and shape of probability distributions or data samples. L-moments are generally known to be more robust than classical moments as they are less sensitive to outliers and often yield more efficient parameter estimates than the maximum likelihood estimates[45]. Data drawn from a wide variety of probability distributions can be effectively characterized by L-moments and unlike product moments, the sampling properties for L-moments statistics are nearly unbiased, even in small samples, and are near normally distributed.

L-moments have various theoretical advantages over the classical moments. For example, for L-moments of a probability distribution to be meaningful, it is only required that the distribution have finite mean, whereas no higher-order moments need be finite. Similarly, in order for the standard errors of L-moments to be finite, only the distribution is required to have finite variance and no higher-order moments need be finite. L-moments also give asymptotic approximations to sampling distributions better than classical moments and provide superior identification of the parent distribution which generated a particular data sample.

First 4 L-moments are defined as follows :

*CHAPTER 2. BACKGROUND : RANDOM VARIABLES AND THEIR
STATISTICS*

$$\lambda_1 = \mathbb{E}[\mathbf{X}] \quad (2.21)$$

$$\lambda_2 = (\mathbb{E}[\mathbf{X}_{2:2}] - \mathbb{E}[\mathbf{X}_{1:2}])/2 \quad (2.22)$$

$$\lambda_3 = (\mathbb{E}[\mathbf{X}_{3:3}] - 2\mathbb{E}[\mathbf{X}_{2:3}] + \mathbb{E}[\mathbf{X}_{1:3}])/3 \quad (2.23)$$

$$\lambda_4 = (\mathbb{E}[\mathbf{X}_{4:4}] - 3\mathbb{E}[\mathbf{X}_{3:4}] + 3\mathbb{E}[\mathbf{X}_{2:4}] - \mathbb{E}[\mathbf{X}_{1:4}])/4 \quad (2.24)$$

where $\mathbf{X}_{k:n}$ denotes the k th order statistic in a independent sample of size n from the distribution of \mathbf{X} .

The named properties of L-moments include Mean, L-Cv, L-Skewness, and L-Kurtosis, are defined as follows :

$$\mu_L = \lambda_1 \quad (\text{mean}) \quad (2.25)$$

$$\sigma_L = \lambda_2 \lambda_1 \quad (\text{L-Cv}) \quad (2.26)$$

$$\gamma_L = \frac{\lambda_3}{\lambda_2} \quad (\text{L-skewness}) \quad (2.27)$$

$$\kappa_L = \frac{\lambda_4}{\lambda_2} \quad (\text{L-Kurtosis}) . \quad (2.28)$$

In Hosking's work, the continuous L-moments are defined in terms of the CDF as follows

$$\lambda_1 = \int_0^1 x(F) dF = \int_{-\infty}^{\infty} xf(x) dx \quad (2.29)$$

$$\lambda_2 = \int_0^1 x(F)(2F - 1) dF = \int_{-\infty}^{\infty} x(2F(x) - 1) f(x) dx \quad (2.30)$$

$$\lambda_3 = \int_0^1 x(F)(6F^2 - 6F + 1) dF \quad (2.31)$$

$$\lambda_4 = \int_0^1 x(F)(20F^3 - 30F^2 + 12F - 1) dF . \quad (2.32)$$

In terms of the independent random variable x , these can be redefined as follows

$$\lambda_1 = \int_{-\infty}^{\infty} xf(x) dx \quad (2.33)$$

$$\lambda_2 = \int_{-\infty}^{\infty} x(2F(x) - 1) f(x) dx \quad (2.34)$$

$$\lambda_3 = \int_{-\infty}^{\infty} x(6F(x)^2 - 6F(x) + 1) f(x) dx \quad (2.35)$$

$$\lambda_4 = \int_{-\infty}^{\infty} x(20F(x)^3 - 30F(x)^2 + 12F(x) - 1) f(x) dx . \quad (2.36)$$

The sample L-moments L_i , estimates of λ_i , can be computed as follows

$$L_1 = \frac{1}{n} \sum_{j=1}^n x_j \quad (2.37)$$

$$L_2 = \frac{1}{n} \sum_{j=2}^n x_j \frac{j-1}{n-1} \quad (2.38)$$

$$L_3 = \frac{1}{n} \sum_{j=3}^n x_j \frac{(j-1)(j-2)}{(n-1)(n-2)} \quad (2.39)$$

$$L_4 = \frac{1}{n} \sum_{j=4}^n x_j \frac{(j-1)(j-2)(j-3)}{(n-1)(n-2)(n-3)} . \quad (2.40)$$

2.7 Sample Quantiles

There are many different definitions for the sample quantiles in use in the literature and across software packages[23]. In this thesis, the definition provided by Hazen[19] is used as follows:

$$p_k = \frac{k - \frac{1}{2}}{n} \quad (2.41)$$

where p_k is the quantile value such that $(\Phi^{-1}(p_k), X_{(k)})$ for $1 \leq k \leq n$ comprise the points drawn on a standard QQ plot.

2.8 Confidence Intervals for Binomial Proportions

The two-sided Clopper-Pearson interval[7, 44] is an inversion of the equal-tailed binomial test for a proportion p . Hence for a confidence level α , the interval contains all values of p which are not rejected by the test. The lower and upper limits of the interval, denoted p_L and p_U respectively, can be computed in terms of the beta distribution:

$$p_L = \beta(\alpha/2, k, n - k + 1) \quad (2.42)$$

$$p_U = \beta(1 - \alpha/2, k + 1, n - k) . \quad (2.43)$$

Applying this confidence interval to the uniform sample distribution of the same size of the sample set drawn from any distribution X allows for the construction of

2.9. QQ PLOTS

confidence intervals around the order statistics of sample of X , which will be shown in the next section.

2.9 QQ Plots

Another useful representation of the cumulative distribution function is the Quantile-Quantile (or QQ) plot. The QQ plot shows the data quantile versus a reference distribution quantile, usually the standard normal distribution quantile. For this reason, normally distributed data appears as a straight line on a QQ plot, see Figure 2.3. Because the standard normal quantile axis is linear in sigma, there is an emphasis on the data quantile tails and a de-emphasis on the region far from the tails.

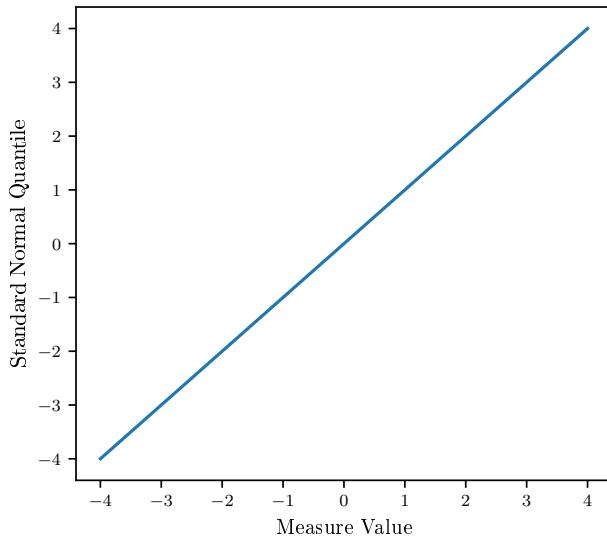


Figure 2.3: Standard Normal Distribution QQ Plot

*CHAPTER 2. BACKGROUND : RANDOM VARIABLES AND THEIR
STATISTICS*

It is also useful to show an axis opposite the standard normal axis which shows the probability as shown in Figure 2.4. Since the standard normal quantile axis is linear, the probability axis has a probit scale. From another perspective, the QQ plot is the CDF plot with a probit probability scale replacing the linear probability scale.

2.9. QQ PLOTS

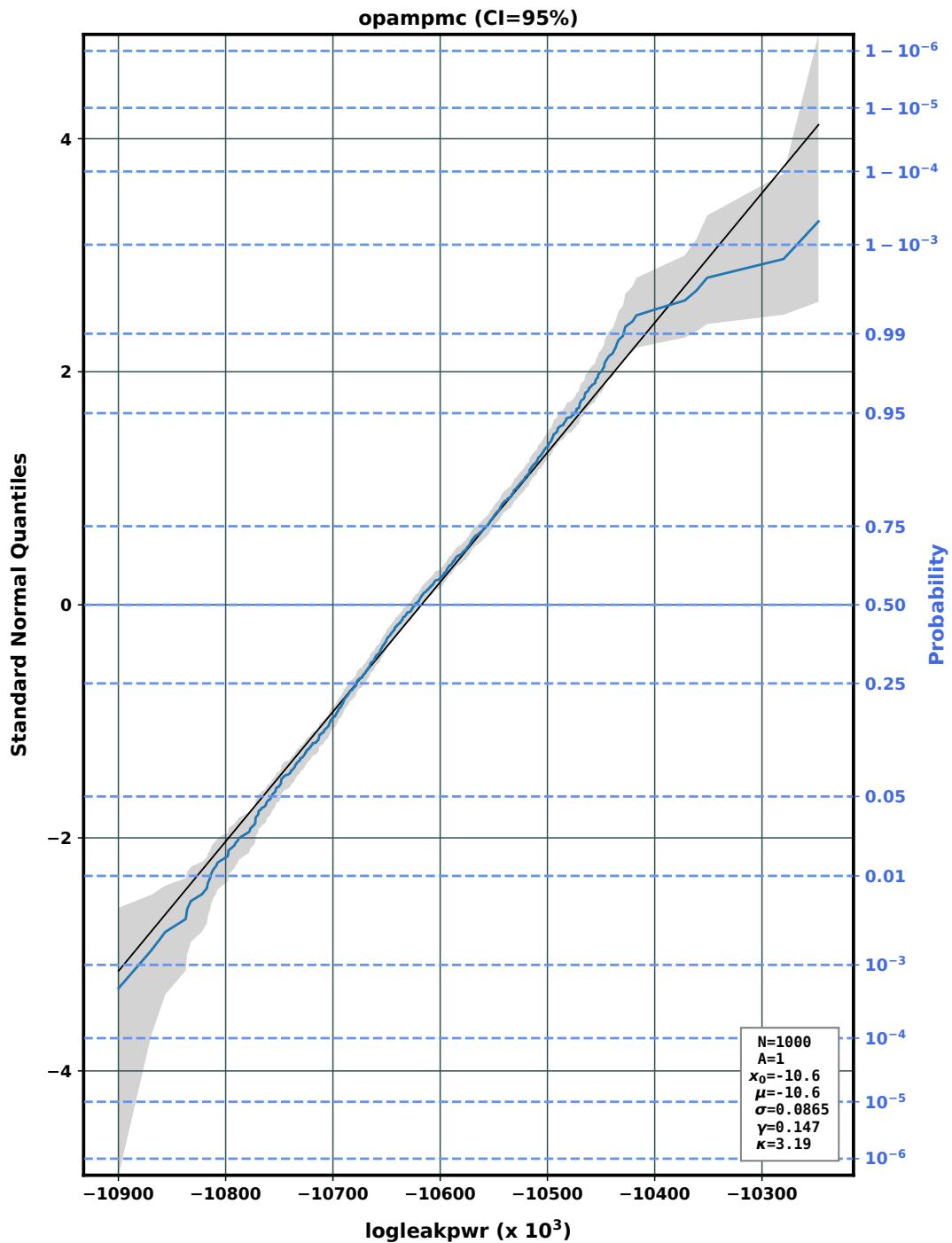


Figure 2.4: Example Production QQ Plot

According to style, the data quantile may appear on either the X or Y axis. Our convention is to use the X axis for the data quantile as it is then directly comparable to the corresponding probability density function plot, in which the data quantile also appears on the X axis. If n samples are drawn from an arbitrary random distribution, the confidence interval of the k th order statistic is described in Section 2.8 which is shown in the grey shaded region around the blue QQ curve in Figure 2.4.

2.10 Error Function

Because the error function is linked strongly to the normal distribution and the normal distribution will be used in the formulation of quantile based moment calculation in Chapter 5, a brief review of the error function is provided in this section. The error function, denoted by $\text{erf}(\cdot)$ is a non-elementary function utilized by the cumulative distribution function of the normal distribution given by

$$\Phi = \frac{1}{2} \left[1 + \text{erf} \left(\frac{x - \mu}{\sigma\sqrt{2}} \right) \right]. \quad (2.44)$$

Though the concept of the error function was known since at least 1808 when Adrain and Gauss independently developed the formula for the normal distribution, it was J. W. L. Glaisher who first proposed the name for the error function. In 1871 he wrote, “As it is necessary that the function should have a name, and as I do not know that any has been suggested, I propose to call it the Error-function, on account of its earliest and still most important use being in connexion[sic] with the

2.11. SELECTED PROBABILITY DISTRIBUTIONS

theory of Probability and notably the theory of Errors..." [13]. The error function, in its standardized form, is defined as

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt . \quad (2.45)$$

In Glashier's work, many of the properties of the error function such as its first order derivatives and integrals as well as related functions were provided. Much more has been described of the error function since this time and higher order derivatives as well as some higher order integrals are available in common mathematical texts[1, 15]. For example, the k th order derivative of the error function is given in closed form by

$$\operatorname{erf}^{(k)}(x) = \frac{2(-1)^{k-1}}{\sqrt{\pi}} H_{k-1}(x) e^{-x^2} = \frac{2}{\sqrt{\pi}} \frac{d^{k-1}}{dx^{k-1}} \left(e^{-x^2} \right), k \in \mathbb{N} \quad (2.46)$$

where $H_k(x)$ is a Hermite polynomial. On the other hand, no general formulae are readily available for integrals of the error function above the third order. A proposal for a general formula for the integral of powers of the error function is given in Appendix A.

2.11 Selected Probability Distributions

The probability distributions discussed in this section are utilized in subsequent theory discussion and/or are used as the basis for theoretical examples in following chapters. In particular, the normal and chi-square distributions is later discussed in the context of formulating a novel sampling method in Chapter 3. The normal

**CHAPTER 2. BACKGROUND : RANDOM VARIABLES AND THEIR
STATISTICS**

distribution is again used in the proposal of a method of moment calculation using quantile information in Chapter 5. The beta distribution will be utilized in its relation to the margin distribution of the multivariate standard normal distribution conditioned on origin distance. The remaining distributions outlined in this section are used for empirical experiments, where differing moment computation methods are compared across selected probability distributions.

2.11.1 Normal distribution

The process variation in modern semiconductor device models is nearly exclusively based on the normal distribution (also known as the Gaussian distribution). For this reason, the normal distribution will be the probability distribution of focus for much of the thesis, though other distributions will be used in empirical experiments or for describing certain statistical relationships with respect to the normal.

The normal distribution has a cumulative distribution function given by

$$\Phi(x) = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x - \mu}{\sigma\sqrt{2}}\right) \right] \quad (2.47)$$

where μ is the mean, σ is the standard deviation, and $\operatorname{erf}(\cdot)$ is the error function, discussed in Section 2.10. The probability density function of the normal distribution is given by

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (2.48)$$

The density and QQ plots of the normal distribution are shown in Figure 2.5.

2.11. SELECTED PROBABILITY DISTRIBUTIONS

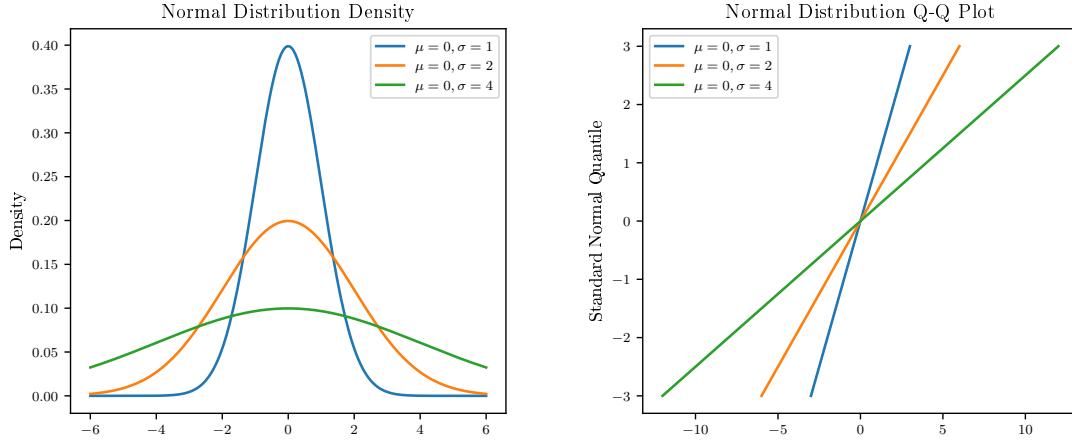


Figure 2.5: Normal Distribution Density and QQ Plot

2.11.2 Chi-squared distribution

This chi-square distribution is important as it is the sum of squares of normal variates. Its probability density function is given by

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}. \quad (2.49)$$

The cumulative density function is given by

$$F(x) = \frac{1}{\Gamma(k/2)} \gamma\left(\frac{k}{2}, \frac{x}{2}\right) \quad (2.50)$$

where $\gamma(s, t)$ is the lower incomplete gamma function given by

$$\gamma(s, x) = \int_0^x t^{s-1} e^{-t} dt. \quad (2.51)$$

The density and QQ plots of the chi-square distribution are shown in Figure 2.6.

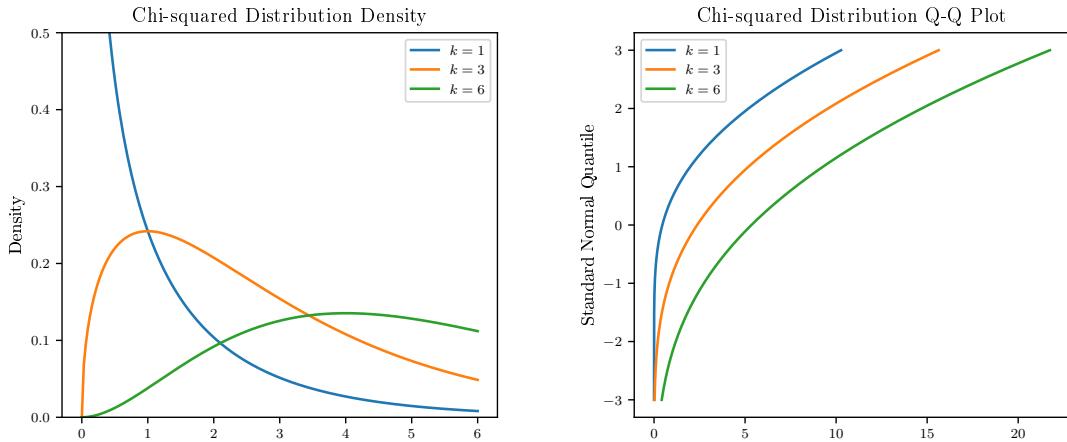


Figure 2.6: Chi-square Distribution Density and QQ Plot

2.11.3 Beta distribution

The probability density function is given by

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \quad (2.52)$$

where

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \quad (2.53)$$

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt. \quad (2.54)$$

The cumulative density function is given by the regularized incomplete beta function I_x as

2.11. SELECTED PROBABILITY DISTRIBUTIONS

$$F(x) = I_x(\alpha, \beta) = \frac{B(x; \alpha, \beta)}{B(\alpha, \beta)} \quad (2.55)$$

where $B(x; a, b)$ is the incomplete beta function given by

$$B(x; a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt. \quad (2.56)$$

The density and QQ plots of the beta distribution are shown in Figure 2.7.

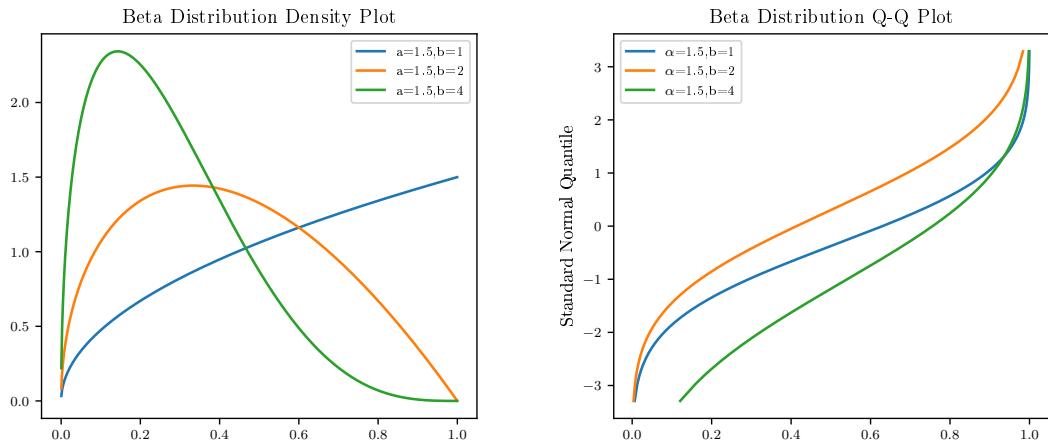


Figure 2.7: Beta Distribution Density and QQ Plot

2.11.4 Log-normal distribution

The log-normal distribution is used quite extensively across many fields[2, 30, 34], having probability density function given by

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right). \quad (2.57)$$

The cumulative density function is given by

$$F(x) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left[\frac{\ln x - \mu}{\sqrt{2}\sigma} \right]. \quad (2.58)$$

The density and QQ plots of the log-normal distribution are shown in Figure 2.8.

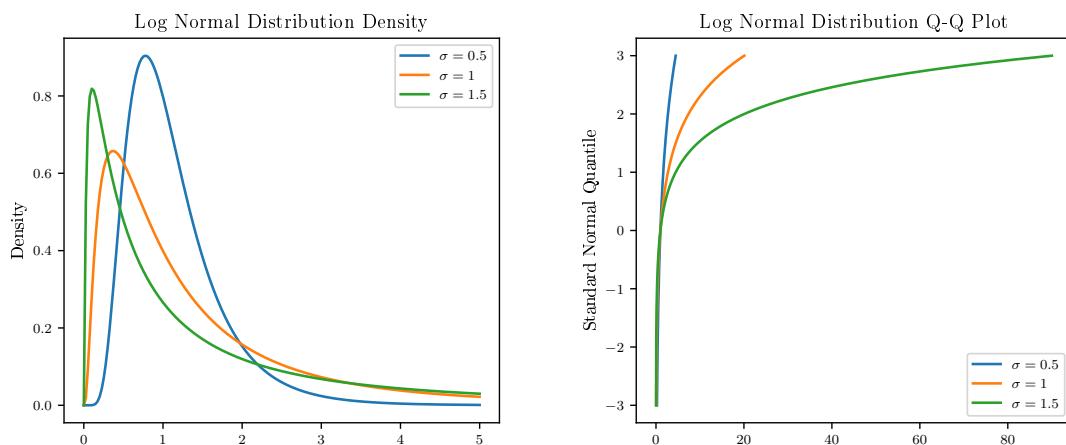


Figure 2.8: Log-normal Distribution Density and Q-Q Plot

2.11.5 Skew Normal distribution

The skew normal distribution[3] has the probability density function given by

$$f(x) = \frac{2}{\omega\sqrt{2\pi}} e^{-\frac{(x-\xi)^2}{2\omega^2}} \int_{-\infty}^{\alpha(\frac{x-\xi}{\omega})} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (2.59)$$

The cumulative density function is given by

2.11. SELECTED PROBABILITY DISTRIBUTIONS

$$F(x) = \Phi\left(\frac{x - \xi}{\omega}\right) - 2T\left(\frac{x - \xi}{\omega}, \alpha\right) \quad (2.60)$$

where $T(h, a)$ is the Owen's T function[37]. The density and QQ plots of the skew normal distribution are shown in Figure 2.9.

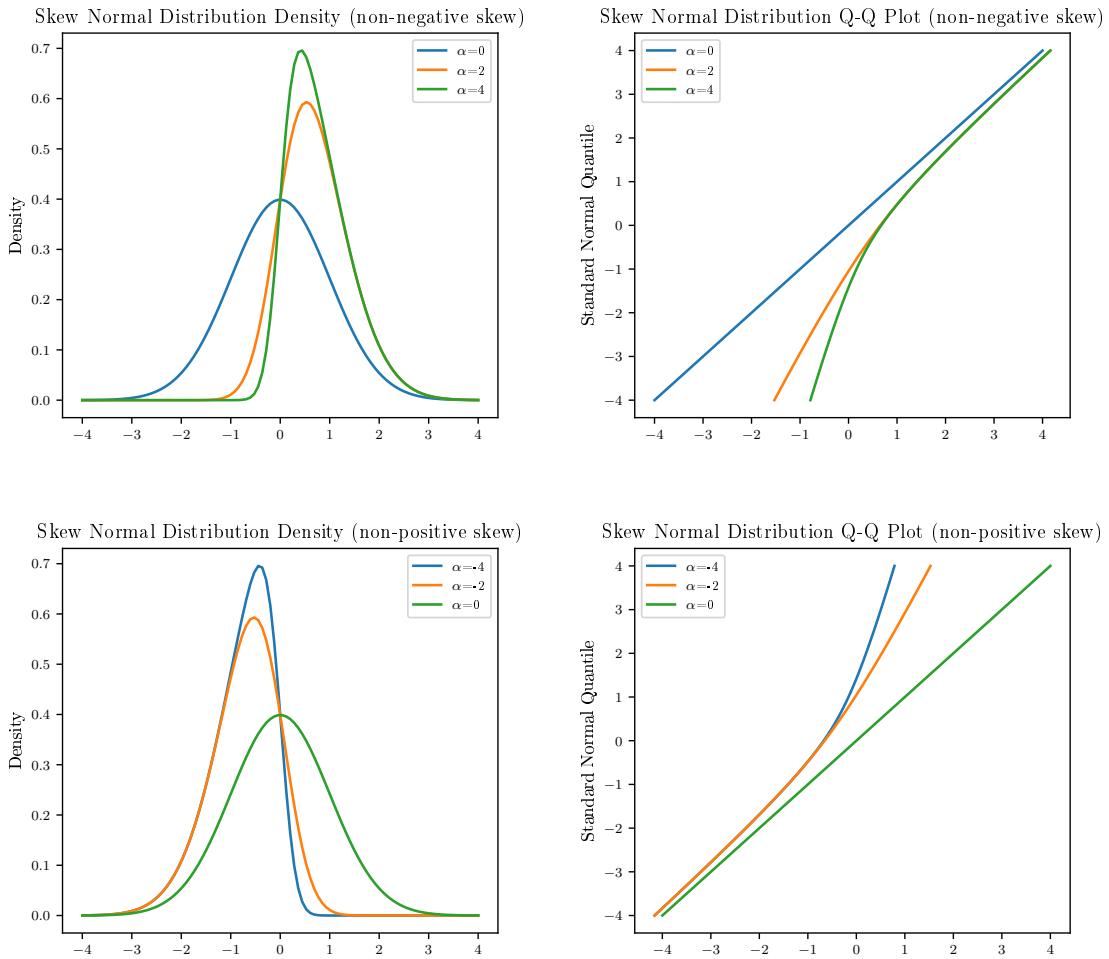


Figure 2.9: Skew Normal Distribution Density and QQ Plot with $\xi = 0$ and $\omega = 1$

2.11.6 Weibull distribution

The probability density function is given by

$$f(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & x < 0 \end{cases}. \quad (2.61)$$

The cumulative density function is given by

$$F(x) = \begin{cases} 1 - e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & x < 0 \end{cases}. \quad (2.62)$$

The density and QQ plots of the Weibull distribution are shown in Figure 2.10.

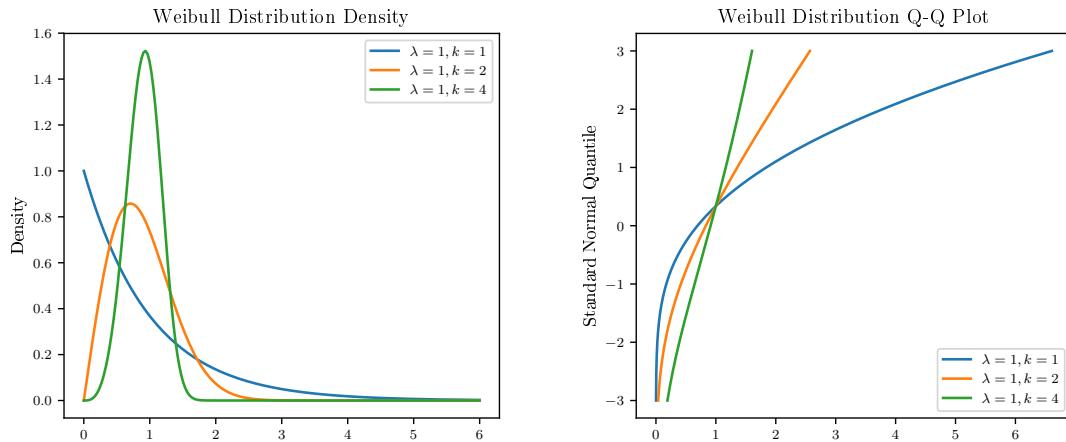


Figure 2.10: Weibull Distribution Density and QQ Plot

2.11.7 Student's t distribution

The probability density function is given by

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}. \quad (2.63)$$

The cumulative density function is given by

$$F(x) = \frac{1}{2} + x\Gamma\left(\frac{\nu+1}{2}\right) \frac{{}_2F_1\left(\frac{1}{2}, \frac{\nu+1}{2}; \frac{3}{2}; -\frac{x^2}{\nu}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \quad (2.64)$$

where ${}_2F_1$ is the hypergeometric function.

The density and QQ plots of the Student's t distribution are shown in Figure 2.11.

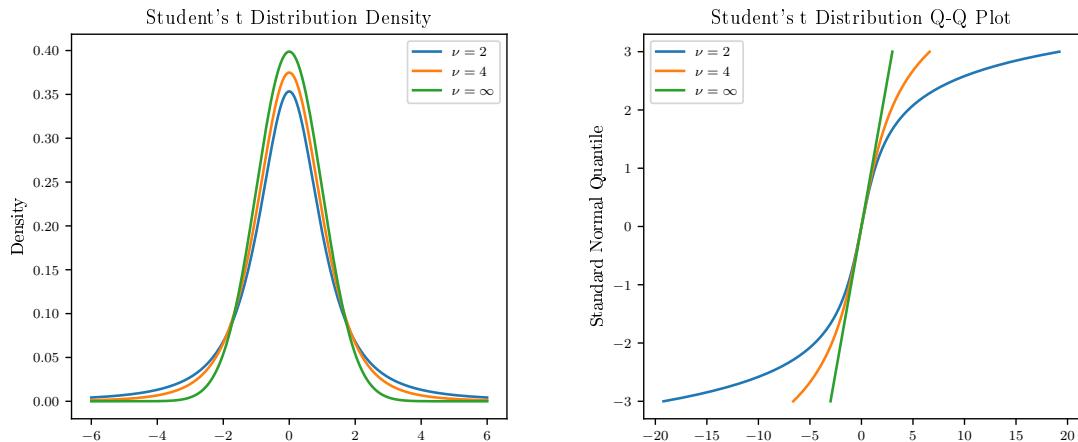


Figure 2.11: Student's t Distribution Density and QQ Plot

Chapter 3

Monte Carlo Sampling Techniques

3.1 Monte Carlo Simulation

A simple random sample is meant to be an unbiased representation of a group. Specifically, a simple random sample is a subset of a statistical population in which each member of the subset has an equal probability of being chosen. The process of choosing simple random samples is called Simple Random Sampling (SRS). For nearly all Monte Carlo simulation performed in the realm of circuit simulation, SRS is employed drawing samples from the multivariate standard normal distribution. Simple Random Sampling with $N(0,1)$, which denotes a normal distribution with mean 0 and variance 1, produces a majority of samples near the origin as shown in Figure 3.1.

3.2. MOTIVATION FOR RARE SAMPLE SIMULATION

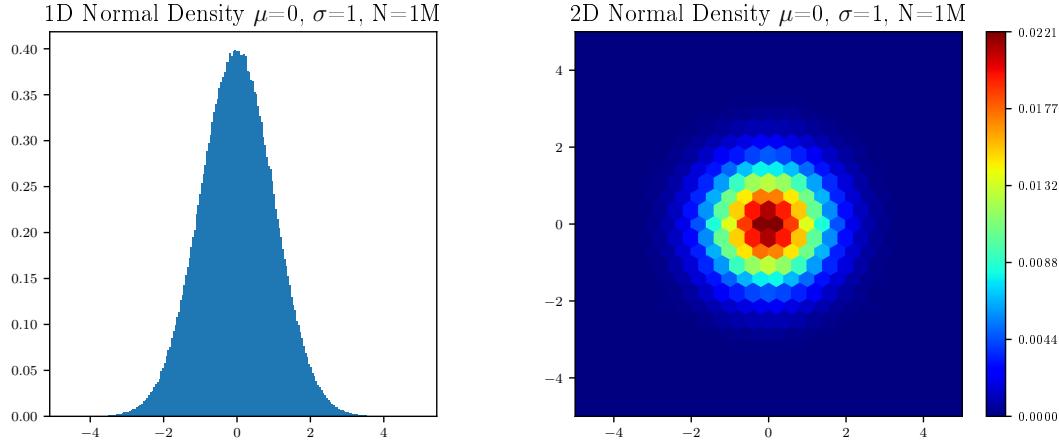


Figure 3.1: One and two dimensional density plots of 1M normal samples with mean=0 and standard deviation=1

The right plot of Figure 3.1 is called a hexbin plot. Hexbin plots are useful for providing a graphical representation of the 2D densities or 2D histograms. This hexbin plot shows the density of the bivariate standard normal distribution, where each axis represents one dimension of that distribution.

3.2 Motivation for Rare Sample Simulation

Rare samples in the input space for the multivariate standard normal are samples which are located far from the origin, as probability density for this distribution is inversely proportional to origin distance. Rare samples in the output space, where the output can be considered a single measure response, can be considered to be the tail samples of that response.

For arbitrary functions of normal random variables, it does not follow that rare samples in the input space tend to correspond with rare samples in the output space. As a counter example, consider the function $Y = X^2$ where $X \sim N(0, 1)$. The smallest values of Y (left tail of its distribution) occur when X is near zero where the PDF of X is most dense.

However, for monotone functions of normal random variables, there is a correspondence between rare samples in the input and rare samples in the output. Specifically, for a non-decreasing monotone function $g(x)$ it is true that

$$x \leq x' \Rightarrow g(x) \leq g(x')$$

while for a non-increasing monotone function $h(x)$ it can also be said that

$$x \geq x' \Rightarrow h(x) \geq h(x') .$$

To extend this notion to higher dimension multivariate functions, let \preceq be the element-wise inequality on \mathbb{R}^n ($x \preceq y$ if and only if $x_i \leq y_i \forall i$). Using this notation, a function $g : \mathbb{R}^n \mapsto \mathbb{R}$ is monotone non-decreasing with respect to this partial ordering if

$$x \preceq y \Rightarrow f(x) \leq f(y) .$$

This is equivalent to requiring

$$f(x) \leq f(x + \gamma e_i) \quad \forall x \in \mathbb{R}^n, \gamma \geq 0, \text{ and } i \in \{1, 2, \dots, n\},$$

3.3. LINEAR FORMS OF STANDARD NORMAL RANDOM VECTORS

where $\{e_1, \dots, e_n\}$ is the standard basis for \mathbb{R}^n . The monotone non-increasing case follows analogously.

For monotone functions, the rare samples in the input space correspond to rare samples in the output space. Measures used in circuit simulation often satisfy this multivariate monotonicity property, as shown later in Section 6.2. If not only the monotone property holds, but we also assume linearity (which is a reasonable approximation in many cases) then additional properties of the output distribution can be derived as shown in the following section.

3.3 Linear Forms of Standard Normal Random Vectors

In the previous section, motivation for rare sample simulation was provided based on the commonality of monotonic behavior of the circuit response. For the linear response, a special case of the monotonic form, the nature of the posterior probability density given the distance of the multivariate normal input from the origin can be described exactly.

Let Z be a sample from an uncorrelated multivariate normal distribution

$$Z \sim N(\mathbf{0}, I_d) .$$

The sum of squares of the elements of Z are distributed according to the χ^2 distribution

$$X = \sum_{i=1}^d Z_i^2 \implies X \sim \chi_d^2 .$$

Define the linear map Y of Z with vector \mathbf{w} and constant μ as

$$Y = Z\mathbf{w} + \mu .$$

If a beta-distributed random variable B is defined as

$$B \sim \beta\left(\frac{d-1}{2}, \frac{d-1}{2}\right)$$

then the marginal distribution Y given a fixed value x for X is given by

$$(Y|X=x) = x\|\mathbf{w}\|_2 \left(B - \frac{1}{2}\right) + \|\mu\|_2 .$$

3.3.1 Empirical Verification

In these experiments, 10M samples are drawn from a multivariate normal distributions of dimension d , for $d \in [2, 3, 4, 5, 50, 500]$ and the samples are then multiplied by the simple linear map I_d

$$Y = ZI_d, Z \sim N(0, I_d) .$$

From Section 3.3, the normalizing factor m was calculated as

$$m = 2\sqrt{d}F_{\chi^2}^{-1}\left(\frac{1}{2}, d\right) .$$

3.3. LINEAR FORMS OF STANDARD NORMAL RANDOM VECTORS

In Figure 3.2, the red curves show the probability density functions of $\beta(\frac{d-1}{2}, \frac{d-1}{2})$, while the blue histograms of \hat{B} represent empirical samples

$$\hat{B} = \frac{Y + m}{2m} .$$

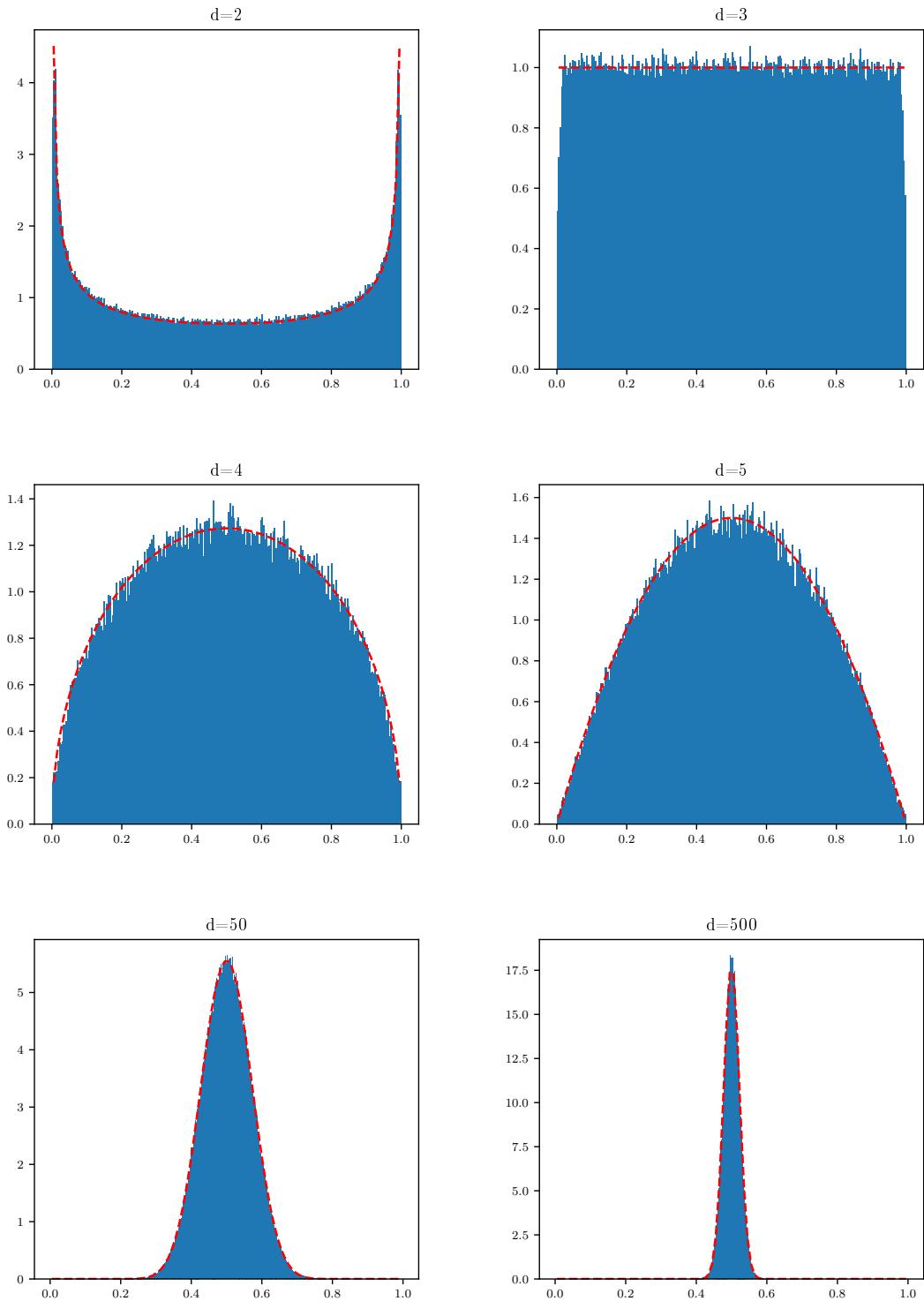


Figure 3.2: Multivariate Normal Conditioned on Vector Norm

3.4 Statistics of the Linear Form

It is well known that a linear form of independent normal random variables is itself normally distributed. Consider a linear form of n independent standard normal random variables as follows:

$$\mathbf{y} = \sum_{i=1}^n a_i \mathbf{x}_i \quad (3.1)$$

where a_1, a_2, \dots, a_n are constants. If the means and variances of \mathbf{x}_i are given by μ_i and σ_i^2 , respectively then the mean and variance of \mathbf{y} are given by

$$\mu_{\mathbf{y}} = \sum_{i=1}^n a_i \mu_i , \quad (3.2)$$

$$\sigma_{\mathbf{y}}^2 = \sum_{i=1}^n a_i^2 \sigma_i^2 . \quad (3.3)$$

Further, if each \mathbf{x}_i is standard normal ($\mu_i = 0$ and $\sigma_i = 1 \forall i$), then the mean and variance of \mathbf{y} is given by

$$\mu_{\mathbf{y}} = 0 , \quad (3.4)$$

$$\sigma_{\mathbf{y}}^2 = \sum_{i=1}^n a_i^2 . \quad (3.5)$$

3.5 Monte Carlo Sampling for Machine Learning

Samples drawn from $N(0, k)$ with $k > 1$ are shifted further from the origin than the standard normal, however the sample distribution is still concentrated at the origin

as shown in Figure 3.3.

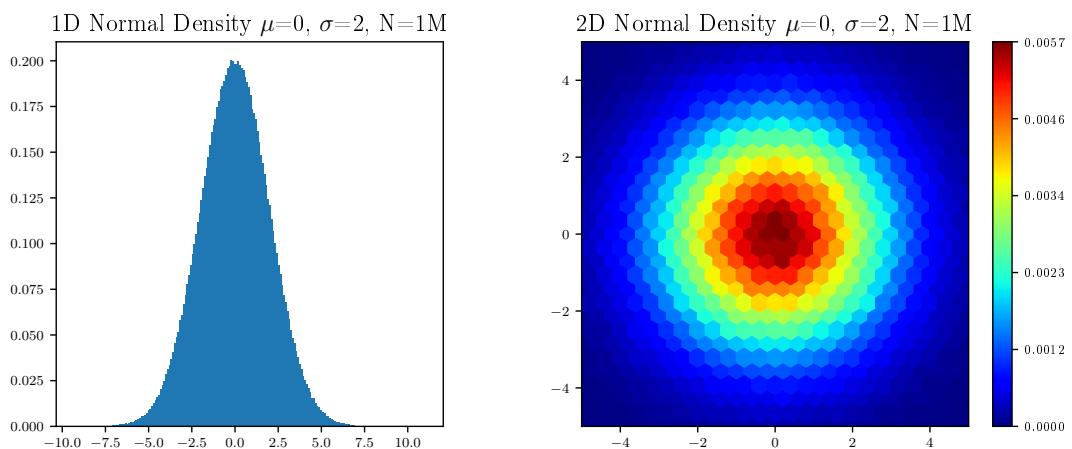


Figure 3.3: One and two dimensional density plots of 1M normal samples with mean=0 and standard deviation=2

Using a uniform distribution, as shown in Figure 3.4,

3.5. MONTE CARLO SAMPLING FOR MACHINE LEARNING

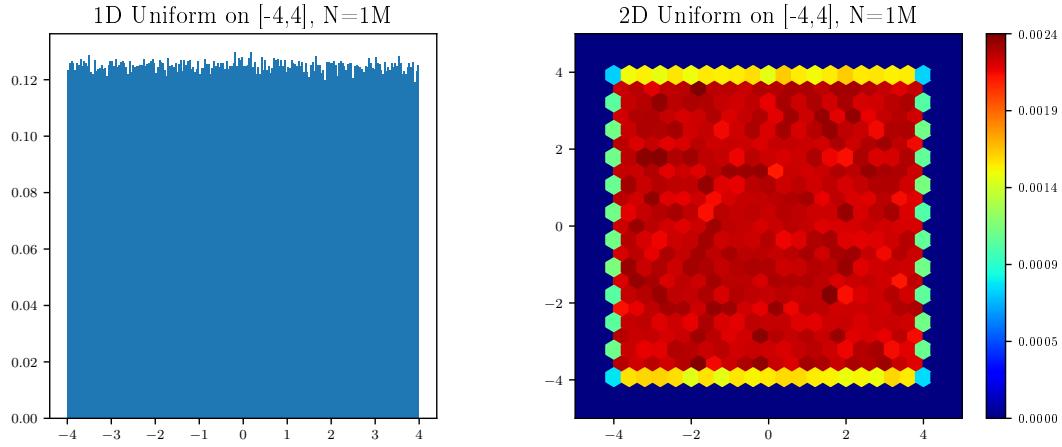


Figure 3.4: One and two dimensional density plots of 1M uniform samples on the interval $[-4,4]$

Figure 3.5 shows the 2D projection of the 5D multivariate standard normal and uniform distributions, respectively. Because these distributions are joint normal and joint uniform, respectively, these figures closely resemble the 2D densities of Figures 3.1 and 3.4.

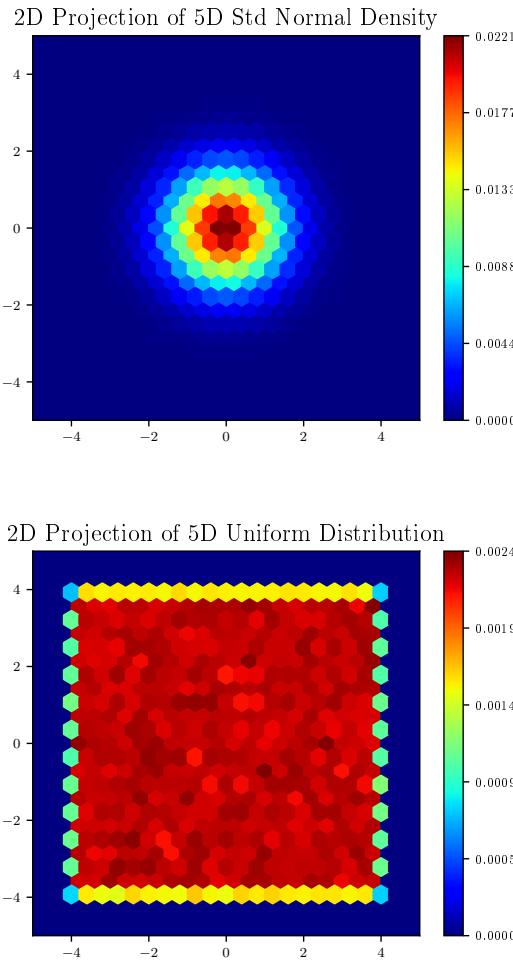


Figure 3.5: 2D Projections of 5D standard normal and uniform distributions

3.5.1 Uniform Sampling in High Dimensions

Drawing from the uniform distribution on the interval $[-5,5]$ in a high dimension d can also be thought of as drawing from a uniform hypercubical volumetric probability density distribution of dimensions 10^d centered at the origin. This uniform

3.5. MONTE CARLO SAMPLING FOR MACHINE LEARNING

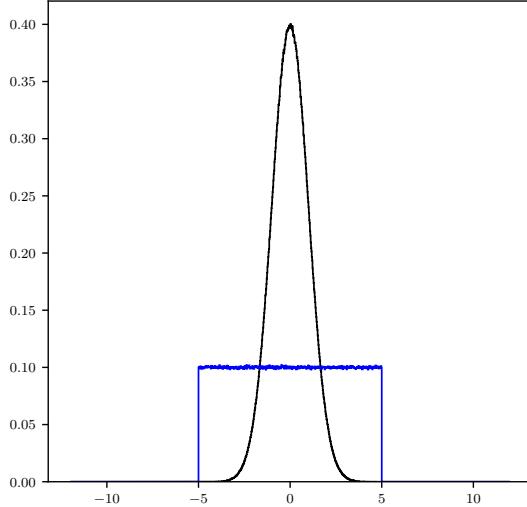


Figure 3.6: One dimensional projections of the normal and uniform distributions drawn from a 300 dimensional space

distribution, when sampled in high dimensions, draws samples which are often too improbable for practical statistical purpose (though importantly, they still may be useful for certain machine learning applications). To see the statistical challenges in sampling from the uniform in high dimensions for probability distributions of the multivariate normal, suppose samples are drawn in 300 dimensional space from both the uniform on $[-5, 5]$ and normal distribution with 1D density projections shown in Figure 3.6.

Figures 3.7 and 3.8 illustrate the distance from the origin and effective sigma value for normal and uniform distributions sampled in this 300 dimensional space, respectively. Note that the majority of the probability density exceeds 30 sigma for the uniform distribution in blue found in Figure 3.8. For reference, a 30 sigma events

occur at a probability level of 1 in 2×10^{197} . In comparison, the estimated number of elementary particles in the known universe is 10^{86} . For most practical purposes, such low probability events need not be considered and this is certainly true for circuit simulation where typically 6-7 sigma or perhaps even 8 sigma is considered the upper end of high sigma simulation.

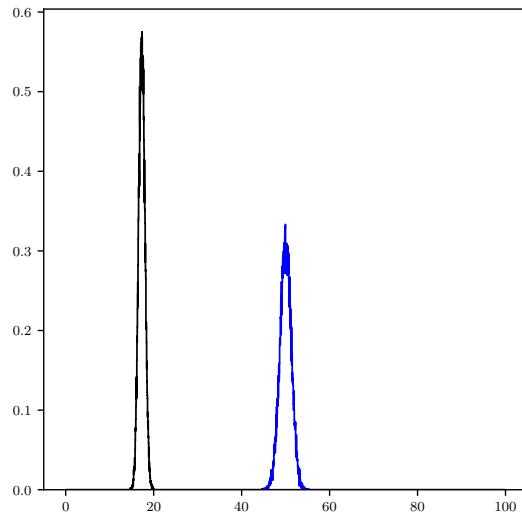


Figure 3.7: The distance from the origin of the normal and uniform distributions drawn from a 300 dimensional space

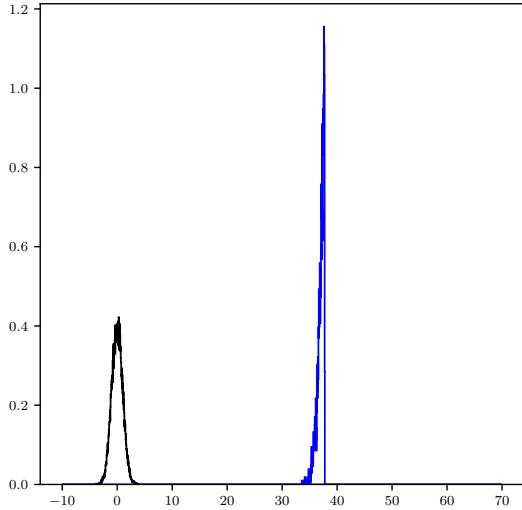


Figure 3.8: The effective value of sigma of the normal and uniform distributions drawn from a 300 dimensional space

3.5.2 Hyperspherical Uniform Volumetric Sampling

To avoid the issue of drawing highly improbable samples, samples can be drawn[17] uniformly from the volume of a hypersphere with radius r . In one and two dimensions, as shown in Figure 3.9, this equates to drawing from a uniform interval $[-r, r]$ and from the area contained within the circle of radius r , respectively. In 3 or more dimensions, samples are drawn uniformly from the volume enclosed in the sphere or hypersphere, respectively. Algorithm 1 shows the steps of drawing samples from a hypersphere of dimension d uniformly.

An improvement of Algorithm 1 is given by Algorithm 2, where one fewer random

Algorithm 1 Hyperspherical Uniform Volumetric Sampling Steps (d+1 RVs per sample)

1. Draw N joint standard normal random vectors each of length d , call each vector \mathbf{x}_i
 2. For each vector \mathbf{x}_i , let s_i be the square of its distance from the origin
 3. Draw N uniform random variables on $(0, 1)$, call values u_i
 4. Hyperspherical Uniform Volumetric samples \mathbf{Z} are given by $\mathbf{z}_i = \mathbf{x}_i u_i^{\frac{1}{d}} / \sqrt{s_i}$
-

variable is required per sample. This is because s_i can be utilized not only for normalization as in Algorithm 1, but its value can be again leveraged as a new random variable from which u_i can be calculated (by F_{χ^2} , i.e. the CDF of the chi-squared distribution) rather than drawn independently.

Algorithm 2 Volumetric Uniform Sampling Steps (Revised, d RVs per sample)

1. Draw N joint standard normal random vectors each of length d , call each vector \mathbf{x}_i
 2. For each vector \mathbf{x}_i , let s_i be the square of its distance from the origin
 3. Calculate \mathbf{u} by $u_i = \gamma(\frac{d}{2}, \frac{s_i}{2}) / \Gamma(\frac{d}{2}) = F_{\chi^2}(s_i, d)$
 4. Volumetric uniform samples \mathbf{Z} are given by $\mathbf{z}_i = \mathbf{x}_i u_i^{\frac{1}{d}} / \sqrt{s_i}$
-

3.5. MONTE CARLO SAMPLING FOR MACHINE LEARNING

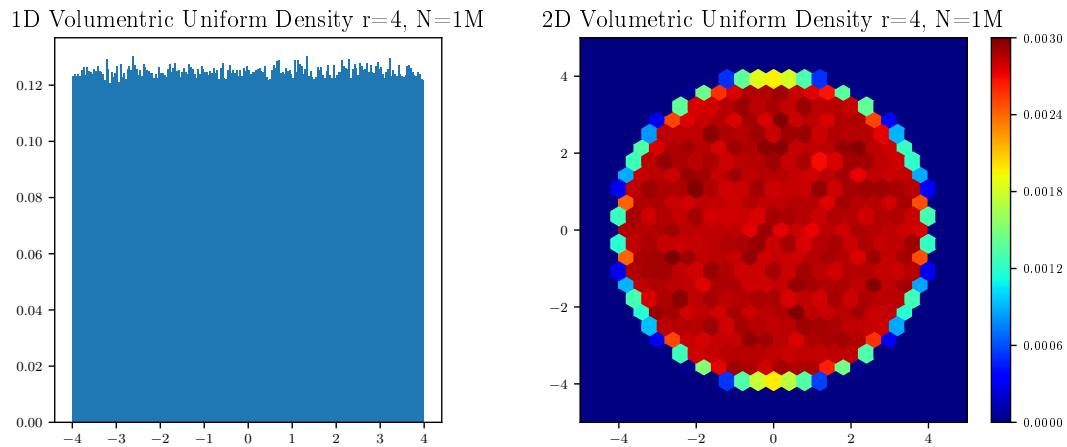


Figure 3.9: One and two dimensional density plots of 1M uniform samples on the interval $[-4,4]$

Scatter plots of the four distributions discussed are shown in Figure 3.10.

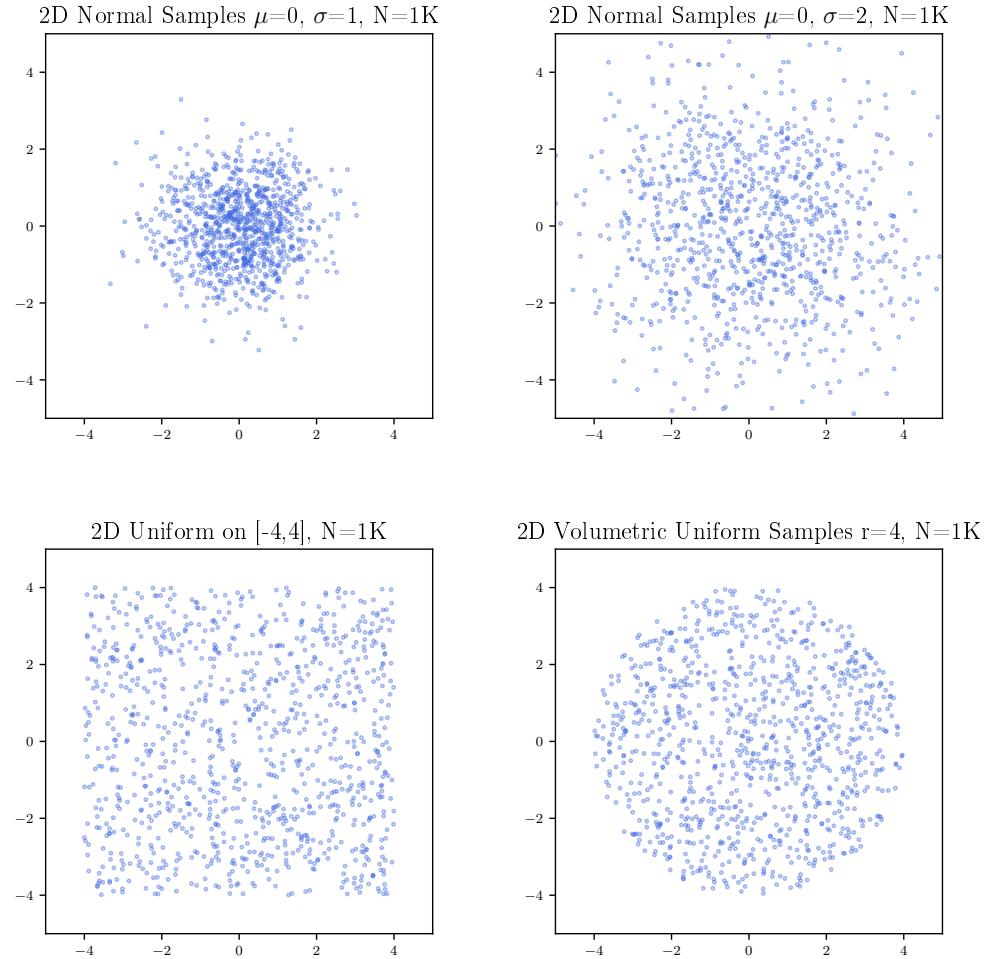


Figure 3.10: Two dimensional scatter plots of 1K samples of normal, uniform, and Hyperspherical Uniform Volumetric samples

3.6 Rejection Sampling

Rejection sampling is a method to simulate random samples from a target distribution (which is unknown or difficult from which to sample) by using random samples

3.6. REJECTION SAMPLING

from a similar, more convenient distribution. A random subset of the generated samples are rejected while the rest are accepted, by the so-called accept-reject criteria. Often the end goal is for the accepted samples to be distributed as if they were from the target distribution.

The canonical example of rejection sampling is to estimate the value of π . This is done by sampling uniformly on the unit square, accepting the values which meet the criteria $x^2 + y^2 \leq 1$. Then, noting that the ratio of the area of the circle to the square is $\pi/4$, the value of π is estimated by 4 times the number of accepted samples divided by the total samples. A similar method can be used to estimate closed integrals of arbitrary functions. While the accept-reject criteria may be easy to evaluate, the analytic integral may be difficult or impossible to compute directly.

For standard area/volume estimation, the rejection criteria is fixed. However, there are important situations where the rejection criteria is dependent on the samples drawn. Consider the acceptance criteria that accepts exactly the k largest samples from a random subset of size n . This problem is similar to the scenario of interest in many engineering problems.

Figure 3.11 shows the example of rejection sampling in 2 dimensions. The lower and upper left plots show the 3K standard normal samples and their corresponding origin distances. In the upper left plot, the origin distances are sorted and all but the furthest 300 samples are rejected, yielding the plot in the lower right.

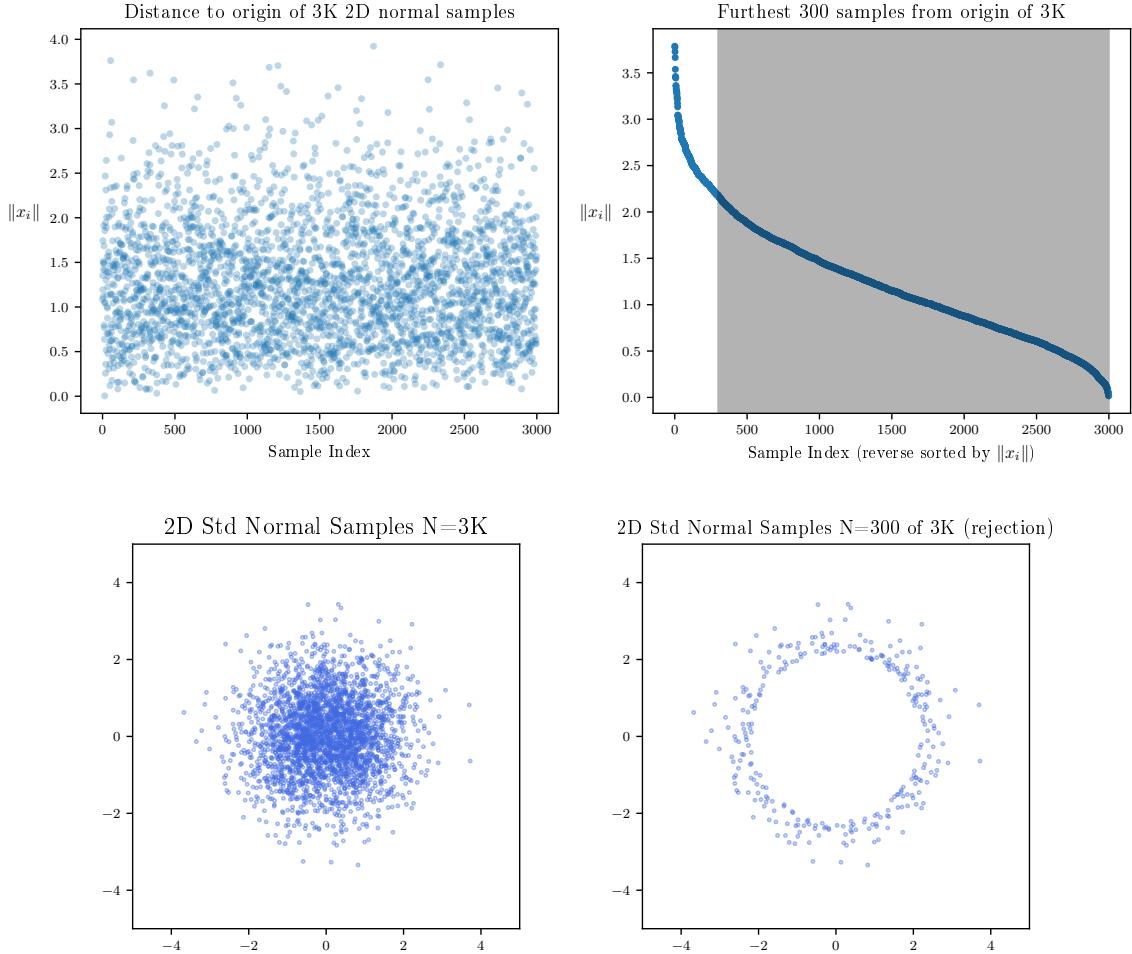


Figure 3.11: Distance to Origin of 2D Normal Samples Unordered and Ordered by Distance to Origin

In such a scenario, rejection sampling does not scale well with sample size since the acceptance probability is inversely proportion to the sample size. For example, suppose 10^{14} samples (roughly 7 sigma) are to be simulated with a simulation budget of 10^6 simulations. Take a simple 6 transistor circuit where each transistor is

described by 7 random numbers (42 total). Assume, on average, the generation of a normal random number takes 5ns (a very competitive generation rate). In total, 5.8 hours would be required on 1000 cores assuming perfect scaling just to generate 10^{14} random numbers to perform Rejection Sampling. In Section 3.9, we propose a new method called In-Order Multivariate Sampling which can directly generate tail samples efficiently. We will first lay the groundwork in the next two sections.

3.7 Generating List of Random Numbers in Sorted Order

The most direct way to create a list of n sorted uniform random numbers is by generating n uniform random numbers and sorting the list, as shown in Algorithm 3. The function `random()` used here and in subsequent algorithms returns an independent random number from the uniform distribution on $[0,1]$.

Algorithm 3 Generate Sorted Uniform Random Numbers by Sorting

```

1: procedure SORTEDUNIFORMBYSORTING
2:   for  $k \leftarrow 1, n$  do
3:      $x[k] \leftarrow \text{random}()$ 
4:    $x \leftarrow \text{sort}(x)$ 
```

The approach of Algorithm 3 requires a sort, which (see Knuth[28]) in linear space is of complexity $\mathcal{O}(n \log n)$. Direct sorting may be avoided by employing more sophisticated approaches which leverage probability theory[27]. Such approaches are possible due to the fact that the gaps between samples drawn from a uniform probability distribution are themselves distributed exponentially. More specifically,

the order statistics of a random sample of size n from the uniform distribution on $[0,1]$ can be computed by

$$y_k = \sum_{j=1}^k x_j \Bigg/ \sum_{j=1}^{n+1} x_j \quad (3.6)$$

for $k \in 1, \dots, n$ and x_j independent exponentially distributed random variables of any fixed mean[26]. This property is utilized in Algorithm 4 whose output is a list of n sorted random numbers independently drawn from $U[0,1]$.

Algorithm 4 Generate Sorted Uniform Random Numbers in Two Passes

```

1: procedure SORTEDUNIFORMTWO PASS
2:    $s \leftarrow 0$ 
3:   for  $k \leftarrow 1, n$  do
4:      $s \leftarrow s - \log(\text{random}())$ 
5:      $x[k] \leftarrow s$ 
6:    $s \leftarrow s - \log(\text{random}())$ 
7:   for  $k \leftarrow 1, n$  do
8:      $x[k] \leftarrow x[k]/s$ 
```

While Algorithm 4 is a great improvement in efficiency over Algorithm 3, it does require two passes to produce the sorted random list. Friedman[11] showed that this can be further improved into a one-pass approach as shown in Algorithm 5.

Algorithm 5 Generate Sorted Uniform Random Numbers in One Pass, Version 1

```

1: procedure SORTEDUNIFORMONEPASSV1
2:    $c \leftarrow 1$ 
3:   for  $k \leftarrow n, 1$  do
4:      $c \leftarrow c * \text{pow}(\text{random}(), 1/k)$ 
5:      $x[k] \leftarrow c$ 
```

Bentley and Saxe[4] of Carnegie-Mellon University proposed an alternative more

3.8. INVERSE TRANSFORM SAMPLING

numerically robust computation given by Algorithm 6, which replaces the call to `pow()` with calls to `log()` and `exp()`.

Algorithm 6 Generate Sorted Uniform Random Numbers in One Pass, Version 2

```

1: procedure SORTEDUNIFORMONEPASSV2
2:    $c \leftarrow 0$ 
3:   for  $k \leftarrow n, 1$  do
4:      $c \leftarrow c * \log(\text{random})/k$ 
5:      $x[k] \leftarrow \exp(c)$             $\triangleright$  for streaming computation, omit store to  $x[k]$ 
```

When the value of n is large (for example, in the billions), it may not be desirable to store all n numbers in array x as shown. Algorithm 6 can easily be converted into a streaming sorted random number generator by omitting the storage step in Line 5.

3.8 Inverse Transform Sampling

In Section 3.7, a method of producing lists of random numbers from the uniform distribution on $[0,1]$ was presented. Not specific to sorted lists, a value x for a random variable X with non-uniform distribution may also be generated using inverse transform sampling.

Lemma 1 (Inverse Transform Sampling). *If U is a uniform random variable on $[0,1]$ then X as defined below has a distribution function F_X .*

$$X = F_X^{-1}(U) \tag{3.7}$$

$$\begin{aligned} \textit{Proof. } \Pr(F^{-1}(U) \leq x) &= \Pr(\inf\{x : F(x) = U\} \leq x) && (\text{definition of } F^{-1}) \\ &= \Pr(U \leq F(x)) && (\text{application of } F \text{ to each side}) \end{aligned}$$

$$= F(x) \quad (\text{definition of CDF of } U)$$

□

Using Algorithm 6 together with Lemma 1, lists of sorted random numbers can be obtained for a wide range of continuous probability distributions. In Section 3.9 this idea will be extended from univariate to multivariate random numbers.

3.9 In-order Multivariate Sampling

In Section 3.7, it was shown that sorted lists of uniform random numbers can be efficiently generated. Then in Section 3.8, this idea was extended to the sorted lists of arbitrarily distributed random numbers via the inverse transform sampling method. In this section, this idea is extended further to multivariate standard normal random vectors.

It is well known that the sum of squares of k standard normal distributed random variables follows a chi-square distribution with k degrees of freedom[38]. In other words, the square of the distance from the origin of a k -dimensional standard normal random vector is distributed as the k degrees of freedom chi-square. Given a random variable $Y \sim U(0, 1)$, inverse transform sampling can be used to find $X = F_{\chi^2}^{-1}(k, Y) \sim \chi^2(k, X)$. Now the distance to the origin of the k -dimensional standard normal random vector is distributed as \sqrt{X} . The values in each of the k dimensions need only satisfy the requirement that the sum of their squares equal X . Utilizing the method of Lemma 1, a point on the surface of the hypersphere of radius \sqrt{X} can be sampled to yield the desired k -dimensional standard normal random

3.9. IN-ORDER MULTIVARIATE SAMPLING

vector[36]. In Algorithm 7, these ideas are synthesized into a program which gives lists of multivariate standard normal random vectors which are sorted according to distance from origin.

Algorithm 7 In-order Multivariate Normal Sampling Steps (Original, $d+1$ RVs per sample)

1. Draw N joint standard normal random vectors each of length d , call each vector \mathbf{x}_i
 2. For each vector \mathbf{x}_i , let s_i be the square of its distance from the origin
 3. Draw N uniform random variables on $(0, 1)$, call values u_i
 4. Update $t = t - \log(u_i)/(N - i)$; with $t = 0$ initially
 5. In-order Multivariate Normal samples \mathbf{Z} are given by $\mathbf{z}_i = \mathbf{x}_i \sqrt{F_{\chi^2}^{-1}(e^t, d)/s_i}$
-

Algorithm 8 (an alternative form of Algorithm 7) requires one fewer random variables per sample. This is possible because the required uniformly distributed variables can be calculated with the CDF of the chi-square distribution with s_i in d degrees of freedom instead of drawn independently.

Algorithm 8 In-order Multivariate Normal Sampling Steps (Revised, d RVs per sample)

1. Draw N joint standard normal random vectors each of length d , call each vector \mathbf{x}_i
 2. For each vector \mathbf{x}_i , let s_i be the square of its distance from the origin
 3. Calculate \mathbf{u} by $u_i = \gamma(\frac{d}{2}, \frac{s_i}{2})/\Gamma(\frac{d}{2}) = F_{\chi^2}(s_i, d)$
 4. Update $t = t - \log(u_i)/(N - i)$; with $t = 0$ initially
 5. In-order Multivariate Normal samples \mathbf{Z} are given by $\mathbf{z}_i = \mathbf{x}_i \sqrt{F_{\chi^2}^{-1}(e^t, d)/s_i}$
-

Figure 3.12 shows the furthest 1K samples out of 1K, 100K, 100M, and 100B samples, respectively. Note that in the first instance of drawing 1K samples of 1K

samples, this sample distribution is the same as standard Monte Carlo. In the latter three instances, the drawn distribution is distributed as if by standard Monte Carlo with rejection sampling, however the samples are drawn sequentially. As the total virtual sample size increases, the first 1K samples drawn become located further and further from the origin.

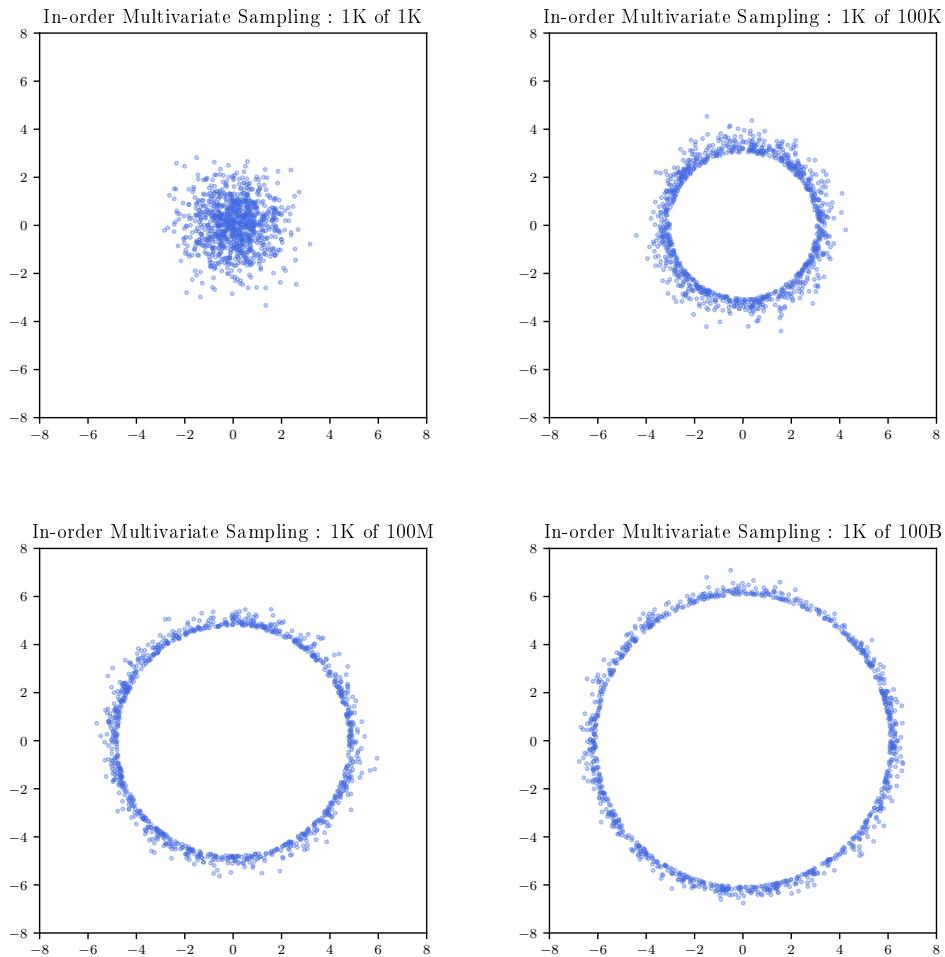


Figure 3.12: In-order Multivariate Sampling

3.9. IN-ORDER MULTIVARIATE SAMPLING

Such sampling can be done in higher dimensions, as shown in Figure 3.13. Unlike Figure 3.12 which is distributed in just two dimensions, Figure 3.13 shows a two dimensional projection of the 5D In-Order Multivariate Sampling distribution.

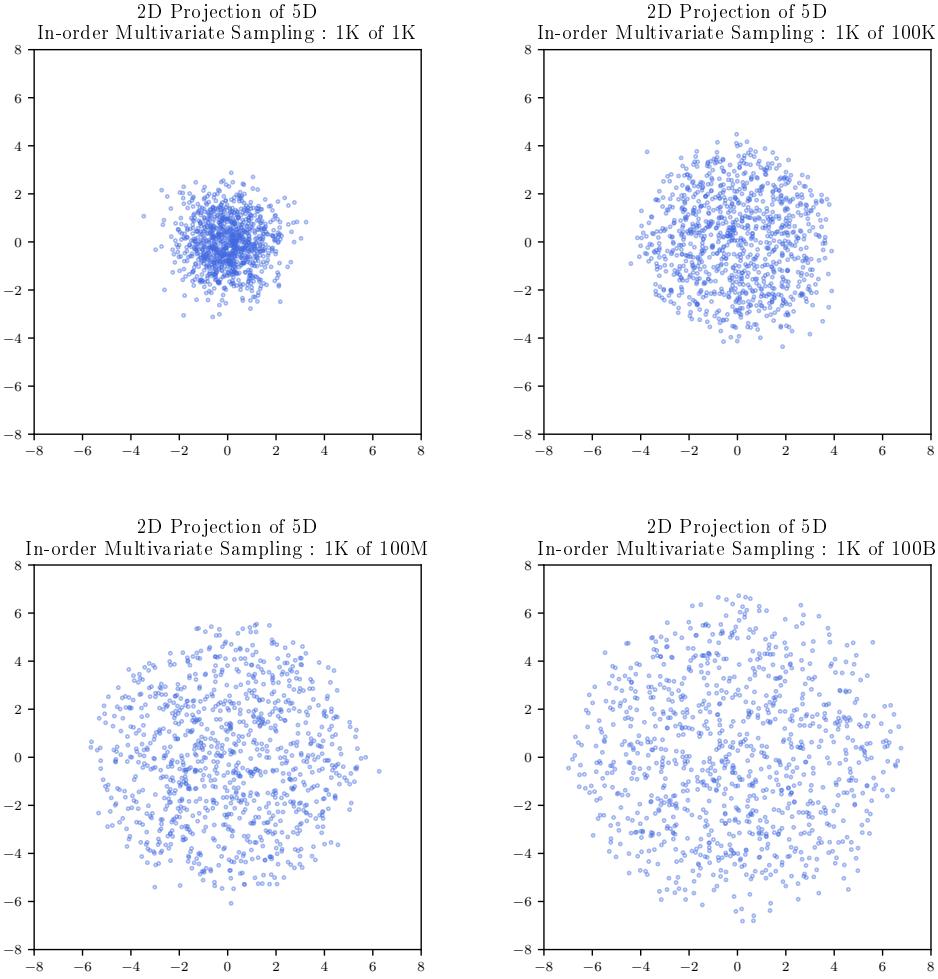


Figure 3.13: In-order Multivariate Sampling

The radius of the hypersphere is related to the dimension and probability in terms of σ . The hyperspherical iso-density contours of the d -dimensional standard

multivariate normal have radius r given by

$$r = \sqrt{F_{\chi^2}^{-1}(P; d)} \quad (3.8)$$

where P is the probability that a random sample is drawn from the interior of that hypersphere. In terms of In-order Multivariate Sampling, this can be restated as follows

$$r = \sqrt{F_{\chi^2}^{-1}(1 - (n - 0.5)/N; d)} \quad (3.9)$$

where N is the total sample size of the distribution from which n samples furthest from the origin are collected.

Example : Radius Calculation

For a standard normal random vector in 42 dimensions, at 5σ , 99.99997% of samples are expected to be contained within a hypersphere centered at the origin with a radius given by

$$r = \sqrt{F_{\chi^2}^{-1}(\Phi(5); 42)} = 10.23268 \quad (3.10)$$

which can be computed in the R language by : `sqrt(qchisq(pnorm(5), 42))` .

Example : σ Calculation

If the failure point closest to the origin lies at a distance 10.23268 from the origin in 42 dimensions, then the σ value associated with the joint probability contained

3.9. IN-ORDER MULTIVARIATE SAMPLING

within that hyperspheric surface is given by

$$\sigma = \Phi^{-1}(F_{\chi^2}(r^2, d)) = \Phi^{-1}(F_{\chi^2}(10.23268^2, 42)) = 5 \quad (3.11)$$

which can be computed in the R language by : `qnorm(pchisq(10.23268^2, 42))` .

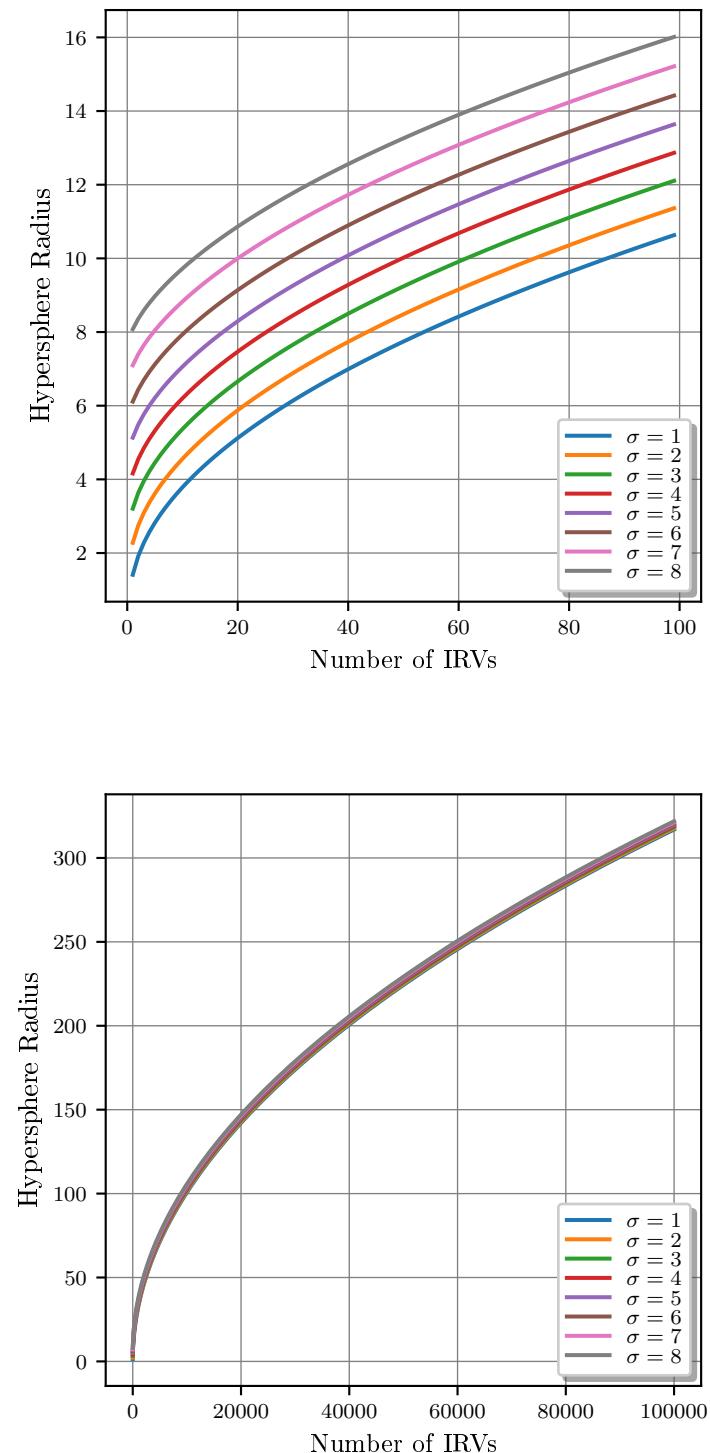


Figure 3.14: Hypersphere radius as a function of the number of IRVs and probability in terms of σ

3.10 Extensions

It is possible to extend the Hyperspherical Uniform Random Sampling (HURS) and In-order Multivariate Sampling (IMS) methods for sampling outlined in this chapter in a variety of ways. Examples of three such extensions are outlined in this section : correlated sampling, Hyperspherical Non-uniform Random Sampling (HNRS), and directional sampling. The HNRS method is an extension of the HURS method only, whereas correlated sampling and directional sampling can be applied both to HURS and IMS. It should be noted that these extensions are not revisited elsewhere in the thesis. They are provided to give evidence for the flexibility of the established methods to cater to differing situational requirements.

3.10.1 Correlated Sampling

In this section, we will show how Hyperspherical Uniform Random Sampling (HURS) and In-order Multivariate Sampling (IMS) methods can be extended to incorporate correlated sampling. The multivariate standard normal distribution can be thought of as a hyperspherical distribution as the boundary of equally probable density contours form hyperspheres in d dimensions. Given non-equal variance among these elements, the hypersphere becomes a hyperellipse. Further, if there exists a covariance among the elements that is non-zero, then the hyperellipse becomes a rotated hyperellipse. A covariance matrix Σ represent such conditions, with variances on the diagonal and covariances off the diagonal. The covariance matrix can be factored uniquely as the product of two upper triangular matrices as follows

$$\boldsymbol{\Sigma} = \mathbf{U}^T \mathbf{U} . \quad (3.12)$$

The matrix \mathbf{U} is the Cholesky matrix, though it is often written[14] instead as a lower diagonal matrix \mathbf{L} defined by

$$\boldsymbol{\Sigma} = \mathbf{L} \mathbf{L}^T . \quad (3.13)$$

The Cholesky matrix transforms uncorrelated variables into variables whose variances and covariances are given by $\boldsymbol{\Sigma}$. In particular, the Cholesky transformation in Equation 3.14 maps the multivariate standard normal distribution \mathbf{X} into the multivariate normal distribution with covariance matrix $\boldsymbol{\Sigma}$ centered at the origin, denoted $\mathbf{Y} = N_d(\mathbf{0}, \boldsymbol{\Sigma})$.

$$\mathbf{Y} = \mathbf{L} \mathbf{X} . \quad (3.14)$$

In the following example, let $\boldsymbol{\Sigma}$ take the following value

$$\boldsymbol{\Sigma} = \begin{bmatrix} 3.25 & 3.5 \\ 3.5 & 5 \end{bmatrix} . \quad (3.15)$$

The lower Cholesky matrix \mathbf{L} is then computed as

$$\mathbf{L} = \begin{bmatrix} 1.80277564 & 0 \\ 1.94145069 & 1.10940039 \end{bmatrix} . \quad (3.16)$$

Using the \mathbf{L} in Equation 3.16 as a mapping of the input distribution by Equation 3.14 yields a correlated distributions as shown in Figure 3.15

3.10. EXTENSIONS

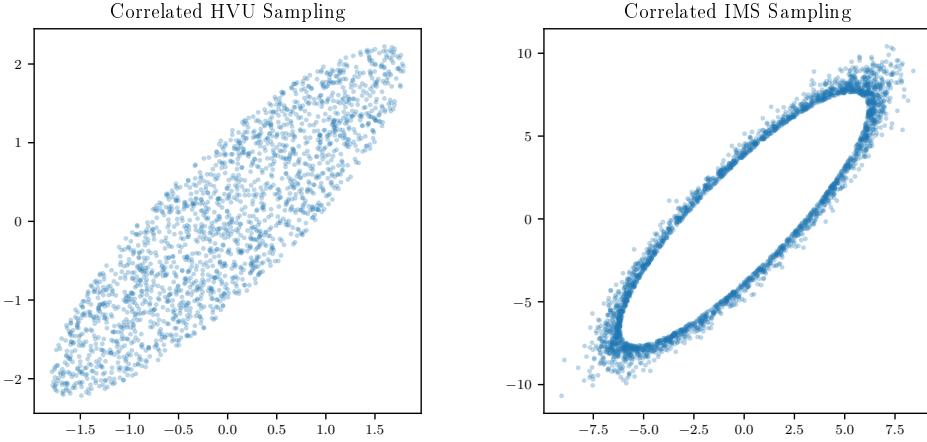


Figure 3.15: Correlated sampling with Hyperspherical Uniform Random Sampling and In-order Multivariate Sampling

3.10.2 Hyperspherical Non-Uniform Random Sampling

Algorithm 2 provided a method to generate a volumetric uniform samples within the surface of hyperspheres. This sample distribution can be generalized to include a radient density parameter r as in Algorithm 9. An outline of the behavior, in relation to the value of r , of samples drawn from the non-uniform distribution is shown in Table 3.1.

Algorithm 9 Volumetric Non-uniform Sampling Steps with radial density parameter r

1. Draw N joint standard normal random vectors each of length d , call each vector \mathbf{x}_i
 2. For each vector \mathbf{x}_i , let s_i be the square of its distance from the origin
 3. Calculate \mathbf{u} by $u_i = \gamma(\frac{d}{2}, \frac{s_i}{2})/\Gamma(\frac{d}{2}) = F_{\chi^2}(s_i, d)$
 4. Volumetric uniform samples \mathbf{Z} are given by $\mathbf{z}_i = \mathbf{x}_i u_i^{\frac{r}{d}} / \sqrt{s_i}$
-

Table 3.1: The effect of non-uniform distribution radial density parameter r on resulting sample distribution

Condition	Description
$r < 0$	samples distributed beyond hyperspheric surface
$r = 0$	samples distributed on hyperspheric surface only
$0 < r < 1$	samples distributed with lower probability density near origin
$r = 1$	samples distributed with uniform density within hypersphere
$r > 1$	samples distributed with higher probability density near origin

3.10. EXTENSIONS

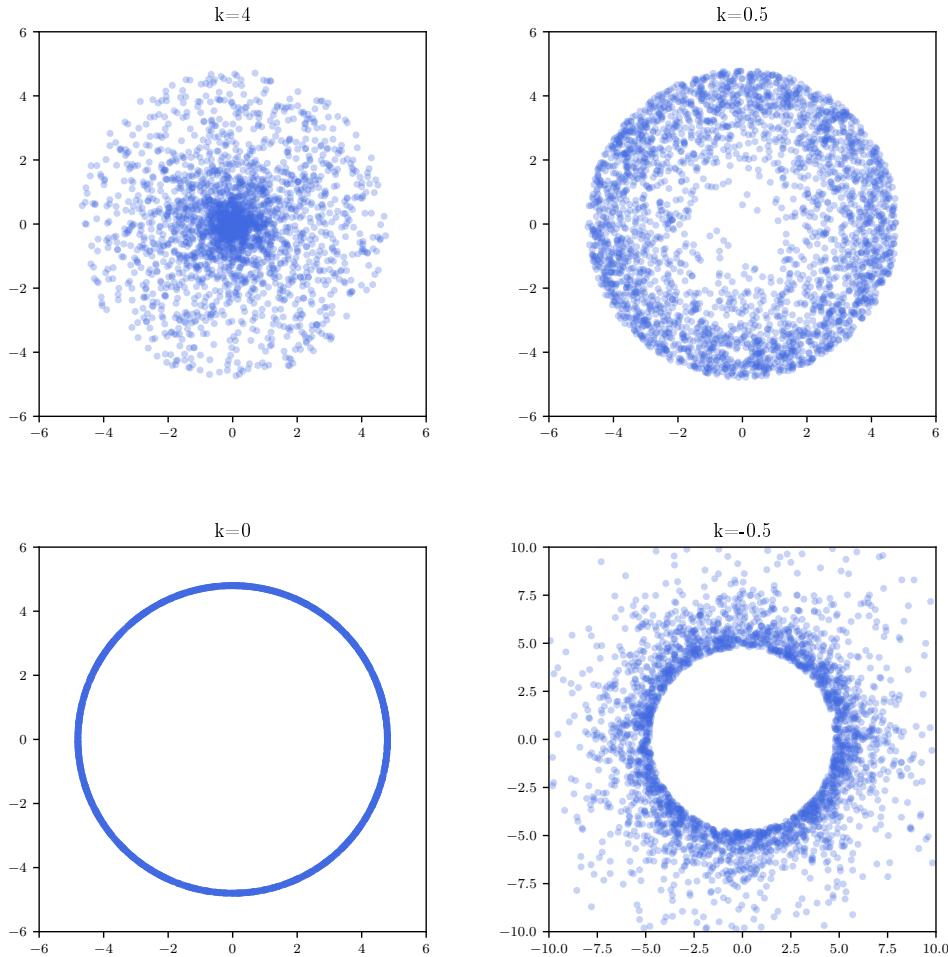


Figure 3.16: Hyperspherical Non-uniform Volumetric Sampling distributions for $k=4, 0.5, 0, -0.5$

3.10.3 Directional Sampling

Named after Richard von Mises, the von Mises distribution[10] (also known as the circular normal distribution or Tikhonov distribution) is a probability distribution

used in directional statistics on the circle and has a probability density function given by

$$f(x \mid \mu, \kappa) = \frac{e^{\kappa \cos(x-\mu)}}{2\pi I_0(\kappa)} \quad (3.17)$$

where $I_0(\kappa)$ is the order 0 modified Bessel function, μ is the location parameter which indicated the angular direction, and κ is the concentration parameter. When κ is zero, the distribution is uniform on the circle whereas when κ is large the samples are highly clustered on the circle about the angle μ .

Moving to higher dimensions, the Fisher distribution (named for Ronald Fisher) extends the von Mises distribution to the sphere and the von Mises-Fisher distribution generalizes this further to the $(d-1)$ -dimensional sphere in \mathbb{R}^d . The probability density function of the von Mises-Fisher distribution for the random d -dimensional unit vector \mathbf{x} is given by:

$$f_p(\mathbf{x}; \boldsymbol{\mu}, \kappa) = c_d(\kappa) \exp(\kappa \boldsymbol{\mu}^T \mathbf{x}) \quad (3.18)$$

where $\kappa \geq 0, d \geq 2, \|\boldsymbol{\mu}\| = 1$ and the normalization constant $c_d(\kappa)$ is given by

$$c_d(\kappa) = \frac{\kappa^{d/2-1}}{(2\pi)^{d/2} I_{d/2-1}(\kappa)}. \quad (3.19)$$

where I_v denotes the modified Bessel function of the first kind of order v .

For visualization purposes, the samples drawn from the von-Mises Fisher distribution in two dimensions (on the unit circle) with constant $\boldsymbol{\mu} = (\sqrt{2}/2, \sqrt{2}/2)$ and $\kappa = 1, 2, 5, 15$ are shown in Figure 3.17. The angular densities of these samples are

3.10. EXTENSIONS

shown in Figure 3.18, which clearly illustrates the high density of sample angles at $\pi/4$, which corresponds to the specified location parameter of $\mu = (\sqrt{2}/2, \sqrt{2}/2)$.

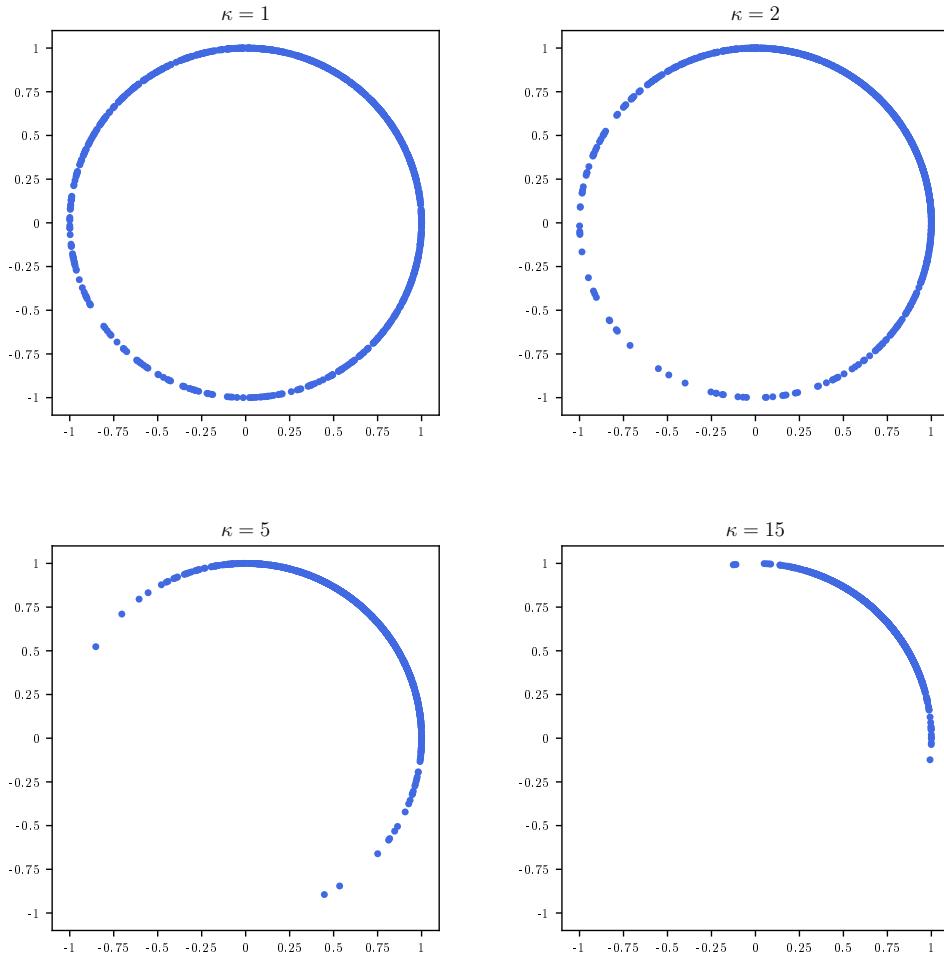


Figure 3.17: Samples from the von Mises-Fisher distribution on the unit circle for $\kappa = 1, 2, 5, 15$ in direction $\mu = (\sqrt{2}/2, \sqrt{2}/2)$

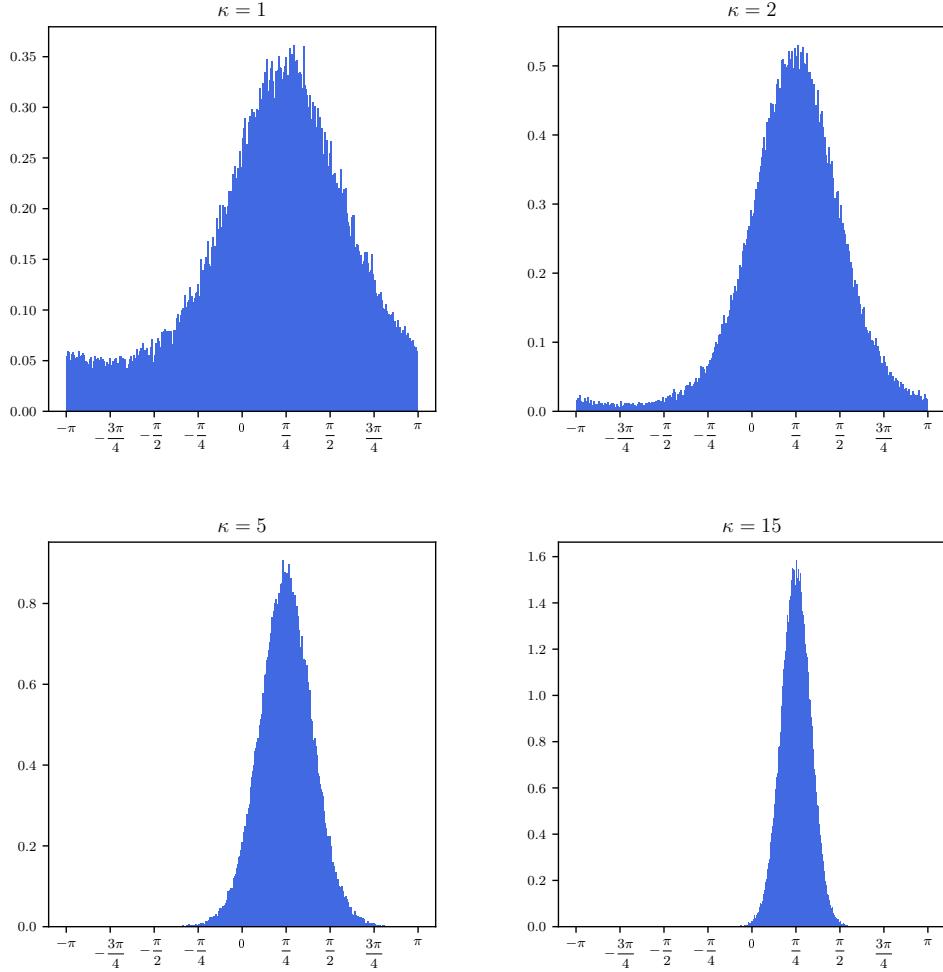


Figure 3.18: Angular densities from the von Mises-Fisher Distribution on the unit circle for $\kappa = 1, 2, 5, 15$ in direction $\mu = (\sqrt{2}/2, \sqrt{2}/2)$

It has been shown that the von Mises-Fisher distribution may be applied to Hyperspherical Uniform Random Sampling in the case where the radius density parameter is set to zero to achieve samples on the hyperspheric surface only. This same technique may be applied to arbitrary Hyperspherical Uniform Random Sampling

3.11. DISCUSSION

and In-order Multivariate Sampling methods as well to achieve directional versions of these sampling methods.

3.11 Discussion

In this chapter, many concepts related to Monte Carlo sampling have been presented. Simple Random Sampling from the normal, i.e. drawing samples from the standard normal distribution in our case, was shown to be an inefficient method to learn about the output distribution tails. Next, sampling from a normal distribution with standard deviation greater than unity instead allowed for faster exploration of tail samples, though a majority of samples were still located in close proximity to the origin. Sampling from the uniform distribution was then presented and details about statistical practicality of drawn samples in higher dimensions was discussed. The inefficiency of rejection sampling to learn about distribution tails was explored and motivated the development a more advanced sampling method.

A novel method, In-order Multivariate Sampling (IMS), was proposed to efficiently explore tail samples from the multivariate normal distribution. An alternative IMS algortihm was given as well as several examples to illustrate the power and theoretical foundation of the method. In-Order Multivariate Sampling is revisited again in Chapter 6 with circuit examples.

Finally, these methods were extended in three ways : correlated sampling, Hyperspherical Non-uniform Random Sampling, and directional sampling. The extensions provide several ideas for leveraging the established sampling methods to apply to a

wider domain of practical problems.

Chapter 4

Quadratic Forms of the Multivariate Standard Normal

In the last chapter, a first order Taylor series expansion (the linear form) of the true MC mapping of the multivariate standard normal was studied. This effort is taken one step further to consider the second order Taylor series expansion, the quadratic form.

Given a set of data, a procedure is established to model the data by a sparsified quadratic polynomial leveraging Legendre polynomials. Once the quadratic form is available, it is transformed into a linear form of normal and/or chi-square random variables. Subsequently, the raw and central moments as well as the cumulants are computed. Finally, using the Cornish-Fisher expansion, the distribution quantiles are estimated.

4.1 Quadratic Forms

As was seen in the previous chapter, linear forms of multivariate standard normals have moments and quantiles which are easily computable. Although the computations are more complex, assuming the vector of independent random variables are joint normal, the moments and quantiles of the quadratic form can be directly computed from the form itself[31]. In particular, a quadratic form of a multivariate standard normal vector has the following properties:

1. Can be expressed as a linear form of independent chi-square and normal random variables
2. Has moments which can be calculated in closed form.
3. Has quantile estimates which can be estimated in closed form.

4.2 Quadratic Modeling with Monomial Polynomials

The problem of finding a polynomial $P_n(x)$ of degree at most n which models a function $g(x)$ which is continuous on $[a, b]$ such that the integral of the square of the error is minimized can be written as

$$\min_{\mathbf{w}} \int_a^b [g(x) - P_n(x)]^2 dx \quad (4.1)$$

where $P_n(x)$ is given by

4.3. QUADRATIC MODELING WITH LEGENDRE POLYNOMIALS

$$P_n(x) = w_0 + w_1x + w_2x^2 + \cdots + w_nx^n . \quad (4.2)$$

To perform the same minimization using a set of m discrete data points $(x_i, y_i), i = 1, 2, \dots, m$, the $P_n(x)$ of Equation 4.2 the new objective becomes

$$\min_{\mathbf{w}} \sum_1^m (y_i - w_0 - w_1x_i - w_2x_i^2 - \cdots - w_nx_i^n)^2 dx \quad (4.3)$$

provided that $n < m$.

While solutions to the problems outlined in this section, they are seldom used in practice due to the problem of ill-conditioning. This computation can be made computationally effective by using a special type of polynomials, called orthogonal polynomials, of which the Legendre polynomials are an important class.

4.3 Quadratic Modeling with Legendre Polynomials

Legendre polynomials form an orthogonal polynomial basis over the interval $[-1,1]$, thus the n th-order Legendre polynomial has the property

$$\int_{-1}^1 P_n(x)P_m(x) dx = 0 \quad \text{if } n \neq m \quad (4.4)$$

where $P_n(x)$ denote the Legendre polynomials and are given by Equations 4.6-4.11 and plotted in Figure 4.1 for $x \in [-1, 1]$ and $n = 0, 1, 2, \dots, 5$.

$$P_0(x) = 1 \quad (4.5)$$

$$P_1(x) = x \quad (4.6)$$

$$P_2(x) = \frac{1}{2}(3x^2 - 1) \quad (4.7)$$

$$P_3(x) = \frac{1}{2}(5x^2 - 3x) \quad (4.8)$$

$$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3) \quad (4.9)$$

$$P_5(x) = \frac{1}{8}(63x^5 - 30x^2 + 3) \quad (4.10)$$

$$P_n(x) = \frac{1}{2^n} \sum_{k=0}^n \binom{n}{k}^2 (x-1)^{n-k} (x+1)^k. \quad (4.11)$$

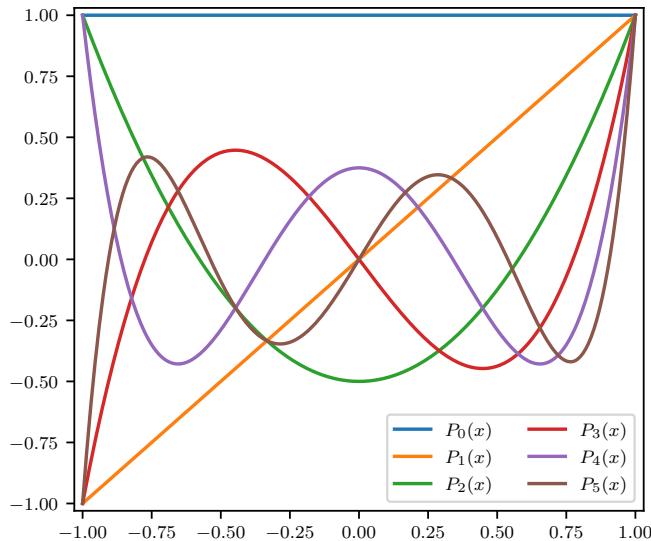


Figure 4.1: Plots of the first six Legendre polynomials

Using Legendre polynomials to model unknown functions is more computation-

4.3. QUADRATIC MODELING WITH LEGENDRE POLYNOMIALS

ally effective as mentioned in the previous section and useful in that small perturbations in the input sample data result in small coefficient updates. Partial fits are also possible with series truncation. In 1 dimension, Legendre polynomials of order $0 \dots n$ form a 1D orthogonal basis on $-1 \leq x \leq 1$ in the degree n univariate polynomial space. To approximate sample data from an unknown function by a n th order Legendre polynomial, the $n + 1$ coefficients a_k are used as follows

$$L_n(x) = \sum_{k=0}^n a_k P_k(x) . \quad (4.12)$$

In higher dimensions, the tensor product of the span of the Legendre polynomial basis functions are considered

$$\mathbf{L}_n^d = L_{n,1} \otimes L_{n,2} \otimes L_{n,3} \cdots \otimes L_{n,d} . \quad (4.13)$$

This may be solved as in the 1D case, but in d dimensions the number of Legendre coefficients becomes $(n_1 + 1) \times (n_2 + 1) \times (n_3 + 1) \times \cdots \times (n_d + 1)$ in d dimensions, when $n_k = n \forall k$ then the number of fitting coefficients number $(n + 1)^d$.

Since the quadratic form is being considered, a d -order tensor space of order up to 2 can be modeled by Legendre polynomials with $k = 3^d$ fitting terms, which can be expressed as c and $w_1 \dots w_{k-1}$. Utilizing LASSO, a sparse quadratic fit is produced by finding c and \mathbf{w} .

$$\arg \min_{c, \mathbf{w}} \left\{ \frac{1}{N} \|y - c\mathbf{1}_N - X\mathbf{w}\|_2^2 \right\} \text{ subject to } \|\mathbf{w}\|_1 \leq \lambda \quad (4.14)$$

Finally, the solved coefficients of c and \mathbf{w} can be remapped to a , \mathbf{b} , and \mathbf{C}

which represent the constant, linear and quadratic terms of the quadratic form, respectively. The dependent random variable y may now be expressed in terms of the vector of independent random variables $\mathbf{x} \sim N_d(\mathbf{0}, \mathbf{I})$ as follows

$$y = \mathbf{x}^\top \mathbf{C} \mathbf{x} + \mathbf{b}^\top \mathbf{x} + a . \quad (4.15)$$

4.4 Linearization via Eigenvector Decomposition

The quadratic form modeled in the previous section can be expressed as a linear combination of normal and chi-square variables. The linear function has a very similar structure given by

$$y = \tilde{\mathbf{x}}^\top \tilde{\mathbf{C}} \tilde{\mathbf{x}} + \tilde{\mathbf{b}}^\top \tilde{\mathbf{x}} + \tilde{a} \quad (4.16)$$

where a matrix \mathbf{U} , comprised of the orthonormal eigenvectors of \mathbf{C} , is used to compute $\tilde{\mathbf{x}}$, $\tilde{\mathbf{C}}$, $\tilde{\mathbf{b}}$ and \tilde{a} as follows

$$\tilde{\mathbf{x}} = \mathbf{U} \mathbf{x} \quad (4.17)$$

$$\tilde{\mathbf{C}} = \mathbf{U} \mathbf{C} \mathbf{U}^\top = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix} \quad (4.18)$$

$$\tilde{\mathbf{b}} = \mathbf{b} \mathbf{U}^\top \quad (4.19)$$

$$\tilde{a} = a . \quad (4.20)$$

Since $\tilde{\mathbf{C}}$ has no off-diagonal elements, the matrix algebra can be rewritten in summation form as

$$y = \sum_{i=1}^d \tilde{\mathbf{C}}_{i,i} \tilde{\mathbf{x}}_i^2 + \tilde{\mathbf{b}}_i \tilde{\mathbf{x}}_i + \tilde{a} . \quad (4.21)$$

In these terms, the dependence of y upon $\tilde{\mathbf{x}}$ can be express in four ways :

1. $\tilde{\mathbf{C}}_{i,i} = 0$ and $\tilde{\mathbf{b}}_i = 0$: No dependence
2. $\tilde{\mathbf{C}}_{i,i} = 0$ and $\tilde{\mathbf{b}}_i \neq 0$: Linear dependence
3. $\tilde{\mathbf{C}}_{i,i} \neq 0$ and $\tilde{\mathbf{b}}_i = 0$: Central quadratic dependence
4. $\tilde{\mathbf{C}}_{i,i} \neq 0$ and $\tilde{\mathbf{b}}_i \neq 0$: Non-central quadratic dependence

4.5 Raw Moments

Derivation details of the calculation of raw moments by the quadratic form and its linearization is covered by Mathai and Provost[31]. The raw moments are given by

$$\mathbb{E}[\mathbf{y}^r] = \sum_{r_1=0}^{r-1} \binom{r-1}{r_1} g^{(r-1-r_1)} \sum_{r_2=0}^{r_1-1} \binom{r_1-1}{r_2} g^{(r_1-1-r_2)} \dots \quad (4.22)$$

$$(4.23)$$

where $g^{(k)}$ is defined by

$$g^{(k)} = \begin{cases} \tilde{a} + \frac{1}{2} \sum_{j=1}^p (2\lambda_j) & k = 0 \\ \frac{k!}{2} \sum_{j=1}^p (2\lambda_j)^{k+1} + \frac{(k+1)!}{2} \sum_{j=1}^p \tilde{\mathbf{b}_j}^2 (2\lambda_j)^{k-1} & k \geq 1 \end{cases}. \quad (4.24)$$

4.6 Central Moments

From the raw moments, the central moments can be calculated as follows:

$$\mu_n = \mathbb{E}[(\mathbf{y} - \mathbb{E}[\mathbf{y}])^n] = \sum_{j=0}^m \binom{n}{j} (-1)^{n-j} \mu'_j \mu^{m-j} \quad (4.25)$$

$$\mu'_m = \mathbb{E}[\mathbf{y}^m] \quad (4.26)$$

$$\mu_1 = 0 \quad (4.27)$$

$$\mu_2 = \mu'_2 - \mu^2 \quad (4.28)$$

$$\mu_3 = \mu'_3 - 3\mu\mu'_2 + 2\mu^3 \quad (4.29)$$

$$\mu_4 = \mu'_4 - 4\mu\mu'_3 + 6\mu^2\mu'_2 - 3\mu^4 \quad (4.30)$$

$$\mu_5 = \mu'_5 - 5\mu\mu'_4 + 10\mu^2\mu'_3 - 10\mu^3\mu'_2 + 4\mu^5. \quad (4.31)$$

4.7 Cumulants

The cumulants, κ_j , can be expressed in terms of the central moments μ_j .

$$\kappa_1 = \mu \quad (4.32)$$

$$\kappa_2 = \mu_2 \quad (4.33)$$

$$\kappa_3 = \mu_3 \quad (4.34)$$

$$\kappa_4 = \mu_4 - 3\mu_2^2 \quad (4.35)$$

$$\kappa_5 = \mu_5 - 10\mu_2\mu_3 . \quad (4.36)$$

4.8 Quantiles

By the Cornish-Fisher expansion[8], the probability distribution quantiles can calculated based on the cumulants.

$$\begin{aligned} F_y^{-1}(q) \approx & \Phi^{-1}(q) + \frac{\Phi^{-1}(q)^2 - 1}{6}\kappa_3 + \frac{\Phi^{-1}(q)^3 - 3\Phi^{-1}(q)}{24}\kappa_4 \\ & - \frac{2\Phi^{-1}(q)^3 - 5\Phi^{-1}(q)}{36}\kappa_3^2 + \frac{\Phi^{-1}(q)^4 - 6\Phi^{-1}(q)^2 + 3}{120}\kappa_5 \\ & - \frac{\Phi^{-1}(q)^4 - 5\Phi^{-1}(q)^2 + 2}{24}\kappa_3\kappa_4 + \frac{12\Phi^{-1}(q)^4 - 53\Phi^{-1}(q)^2 + 17}{324}\kappa_3^3 \end{aligned} \quad (4.37)$$

where $\Phi^{-1}(q)$ is the quantile function of the standard normal distribution.

The expansion is performed using the normalized κ^* rather than κ as the expansion requires standard normality. This yields the quantile of y^* which is subsequently un-normalized. The relationships between these terms is given by

$$y^* = \frac{y - \mu}{\sigma} \quad (4.38)$$

$$\mu_j^* = \frac{\mu_j}{\sigma^j} \quad (4.39)$$

$$\sigma = \sqrt{\mu_2} . \quad (4.40)$$

4.9 Toy Example

In this section, the Booth function is used to illustrate the quadratic modeling and the extraction of distribution properties directly from the quadratic model as described earlier in this chapter. The Booth function is given in both standard quadratic form as well as in matrix form as follows:

$$y = 5x_0^2 + 5x_1^2 + 8x_0x_1 - 34x_0 - 38x_1 + 74$$

$$\begin{aligned} y &= \mathbf{x}' \mathbf{C} \mathbf{x} + \mathbf{b}' \mathbf{x} + a \\ \mathbf{C} &= \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \\ \mathbf{b} &= \begin{bmatrix} -34 & -38 \end{bmatrix} \\ a &= 74 . \end{aligned}$$

Now define the rows of square matrix \mathbf{U} , comprised of the orthonormal eigenvectors of \mathbf{C} , as follows

4.9. TOY EXAMPLE

$$\mathbf{U} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} .$$

Completing the transformation, it can be written as

$$\begin{aligned} y &= \tilde{\mathbf{x}}' \tilde{\mathbf{C}} \tilde{\mathbf{x}} + \tilde{\mathbf{b}} \tilde{\mathbf{x}} + \tilde{a} \\ \tilde{\mathbf{x}} &= \mathbf{U} \mathbf{x} \\ \tilde{\mathbf{C}} &= \mathbf{U} \mathbf{C} \mathbf{U}' = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix} \\ \tilde{\mathbf{b}} &= \mathbf{b} \mathbf{U}' = \begin{bmatrix} -36\sqrt{2} & -2\sqrt{2} \end{bmatrix} \\ \tilde{a} &= a = 74 \end{aligned}$$

$$\begin{aligned} y &= 9\tilde{x}_0^2 + \tilde{x}_1^2 - 36\sqrt{2}\tilde{x}_0 - 2\sqrt{2}\tilde{x}_1 + 74 \\ &= 9(\tilde{x}_0 - 2\sqrt{2})^2 + (\tilde{x}_1 - \sqrt{2})^2 . \end{aligned}$$

Now y can be expressed as a linear combination of two independent chi-squared random variables

$$(\tilde{x}_0 - 2\sqrt{2})^2 \sim \chi^2(1, 8)$$

$$(\tilde{x}_1 - \sqrt{2})^2 \sim \chi^2(1, 2) .$$

The analytic raw moments, calculated by Equation 4.22, are compared with the sample raw moments in Table 4.1. The correspondence between the analytic and

*CHAPTER 4. QUADRATIC FORMS OF THE MULTIVARIATE STANDARD
NORMAL*

sample raw moments is good, with the error of the analytic moments versus the sample moments within 0.3%.

Table 4.1: Sample versus analytic raw moments : Booth function

r	Analytic	Sample	Rel. Error
1	84.000	84.002	0.003%
2	9810	9821	0.111%
3	1432512	1435297	0.194%
4	248515452	249157550	0.258%
5	49659227280	49807988025	0.300%

Using the raw moments, the central moments are computed by Equations 4.25-4.31. A comparison between the analytic central moments and sample central moments is given in Table 4.2. The agreement between the analytic and sample central moments is also good, with a maximum relative error of 0.2%.

Table 4.2: Sample versus analytic central moments : Booth function

r	Analytic	Sample	Rel. Error
1	0	0	0.000%
2	2765	2764	-0.026%
3	145901	145855	-0.031%
4	33329811	33302404	-0.082%
5	4965361922	4955431600	-0.200%

4.9. TOY EXAMPLE

The normalized central moments and cumulants, computed by Equations 4.40 and 4.33-4.36 respectively, are reported in Table 4.3. These quantities are subsequently used in order to calculate analytic quantile information.

Table 4.3: Normalized central moments and cumulants : Booth function

j	μ_j^*	κ_j^*
1	0	0
2	1	1
3	1.0037	1.0037
4	4.3605	1.3605
5	12.3548	2.3181

Lastly, the quantiles are computed analytically by the Cornish-Fisher expansion given by Equation 4.38. A comparison between the analytic and sample values is given in Table 4.4. The correspondence between the two quantile calculation methods is very close, with the largest relative error being less than 0.1% up to a quantile of 0.999999.

Table 4.4: Sample versus analytic quantiles : Booth function

Quantile	Analytic	Sample	Rel. Error
0.5	75.06	75.02	-0.047%
0.75	113.54	113.52	-0.023%
0.95	183.31	183.23	-0.039%
0.99	242.47	242.29	-0.075%
0.999	318.73	318.51	-0.068%
0.9999	389.30	389.19	-0.028%
0.99999	456.22	456.49	0.059%
0.999999	520.50	520.16	-0.065%
0.9999999	582.74	596.61	2.380%

4.10 Quadratic Modeling Examples

In the previous section, the modeling of the Booth function was used as a vehicle for demonstrating the extraction of probability distribution statistics such as moments, cumulants, and quantiles from the quadratic form directly. In the last section, in order to outline the ideas clearly, the quadratic form of the model for the Booth function was given as an input. In this section, the more general case is considered where only sample data from an unknown function is provided. Several different types of functions are considered, both univariate and multivariate. In each case, the analytic quantiles of the linear and quadratic models are compared to the quantiles

4.10. QUADRATIC MODELING EXAMPLES

of the sample data. For reference, the linear and quadratic model equations are also provided.

4.10.1 Univariate Linear

Table 4.5: Univariate Linear Example True and Modeled Forms

True Form	$2X + 1$
Modeled Linear Form	$2X + 1$
Modeled Quadratic Form	$1.9956X + 0.999$

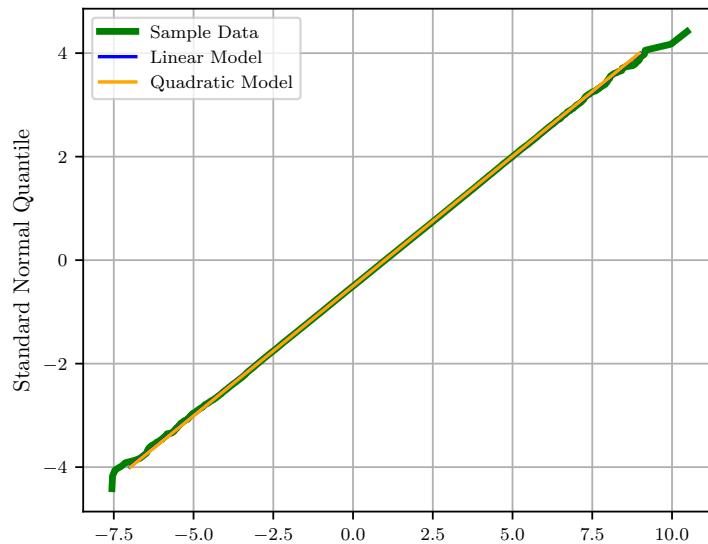


Figure 4.2: Sample data of univariate linear form compared with analytic quantiles of modeled linear and quadratic forms

4.10.2 Univariate Quadratic

Table 4.6: Univariate Quadratic Example True and Modeled Forms

True Form	$2X^2 - 3X + 7$
Modeled Linear Form	$-2.9937X + 8.9999$
Modeled Quadratic Form	$1.9156X^2 - 2.9599X + 7.075$

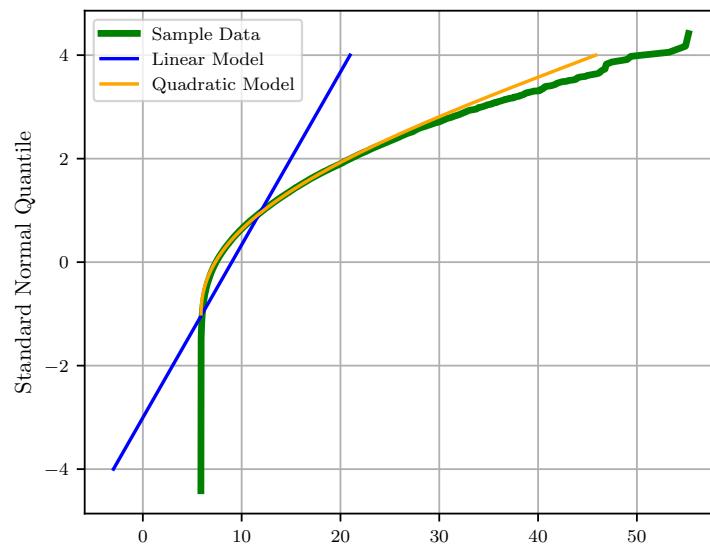


Figure 4.3: Sample data of univariate quadratic form compared with analytic quantiles of modeled linear and quadratic forms

4.10.3 Multivariate Quartic

Table 4.7: Multivariate Quartic Example True and Modeled Forms

True Form	Himmelblau's function : $(X_0^2 + X_1 - 11)^2 + (X_0 + X_1^2 - 7)^2$ $= X_0^4 + 2X_1X_0^2 - 21X_0^2 + 2X_1^2X_0 - 14X_0 + X_1^4$ $- 13X_1^2 - 22X_1 + 170$
Modeled Linear Form	$-12.083X_0 - 19.966X_1 + 142.05$
Modeled Quadratic Form	$-14.305X_0 - 6.4260X_1^2 - 11.379X_0 - 19.398X_1 + 162.62$

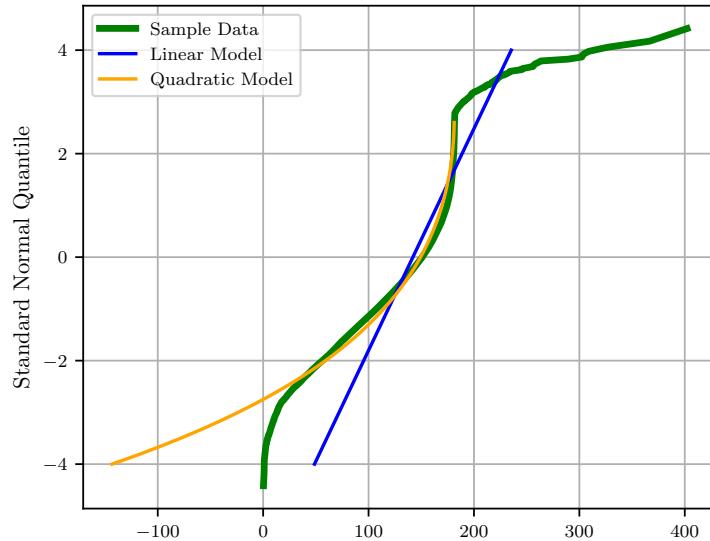


Figure 4.4: Sample data of multivariate quartic form compared with analytic quantiles of modeled linear and quadratic forms

4.10.4 Multivariate Exponential and Quadratic

Table 4.8: Multivariate Exponential and Quadratic Example True and Modeled Forms

True Form	$e^{X_0} + 8X_1^2 - 17X_0$
Modeled Linear Form	$-15.339X_0 - 0.022676X_1 + 9.6711$
Modeled Quadratic Form	$0.70111X_0^2 + 7.8921X_1^2 - 5.3091X_0 + 1.0426$

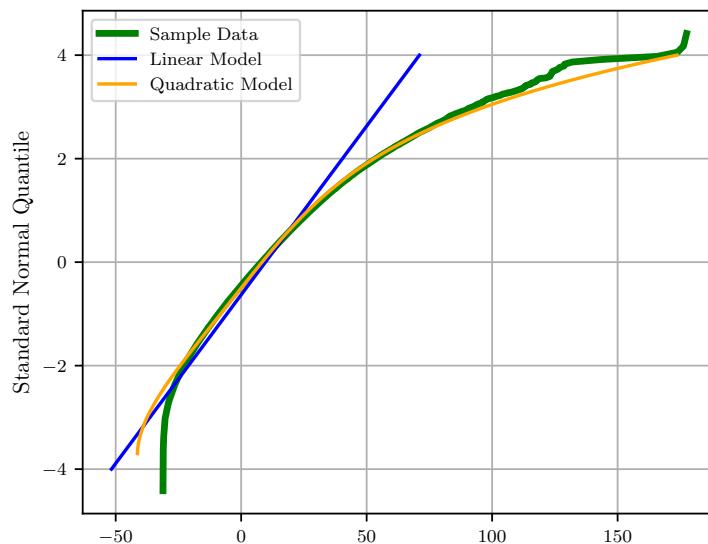


Figure 4.5: Sample data of multivariate exponential and quadratic form compared with analytic quantiles of modeled linear and quadratic forms

4.10.5 Quadratic Modeling Example Observations

Linear and quadratic forms of multivariate standard normal vectors are useful to capture the first and second order terms of a Taylor series expansion. From the examples given in this section, it is clear that the quadratic form is not well suited to model the distribution of black box forms for sampling far from the origin which vary significantly from linear or quadratic. The consistency of the quantiles of the quadratic form near the origin could be useful for modeling high dimensional systems for lower sigma as the Taylor expansion is performed about the origin. As shown in this chapter, the quantiles (and other statistics such as moments and cumulants) of the quadratic form may be determined analytically, making the quadratic form a useful modeling tool.

4.11 Discussion

In this chapter, a complete flow for modeling an unknown function by a quadratic form was outlined. Distribution parameters were then computed not by sampling, but analytically from the quadratic form itself.

A multidimensional Legendre polynomial of order two is used to model the unknown form of the standard multivariate normals as a quadratic form. Next, the quadratic form is broken down by eigenvalue decomposition to reveal an underlying linear equation which is a combination of normal and chi-squared distributed vectors of random variables. Given this new functional formulation, raw moments are calculated whereupon central moments are derived. Once the raw and central

*CHAPTER 4. QUADRATIC FORMS OF THE MULTIVARIATE STANDARD
NORMAL*

moments are available, the cumulants may then be subsequently computed. Normalized quantiles are calculated by the Cornish-Fisher expansion given the normalized cumulants. A final reverse normalization of the normalized quantiles then provides the quantiles of the quadratic model fit to the unknown form.

The procedure outlined above can be useful to provide rough distribution properties where extensive sampling is expensive or impractical. Abstractly, the quadratic form represents the second order Taylor series expansion of the unknown form.

The main contribution in this chapter is the outline of a coherent flow which starts from raw data and ends with extracted statistical information from a sparse quadratic form including moments, cumulants, and quantiles.

Chapter 5

Quantile-Based Moments

5.1 Computing Moments from Quantile Information

The r th raw moment of probability distribution of random variable \mathbf{X} is given by

$$\mathbb{E}[\mathbf{X}^r] = \int_{-\infty}^{\infty} x^r f(x) dx \quad (5.1)$$

where f is the probability density function (PDF) of \mathbf{X} .

Analogously, moments can also be written in terms of the cumulative density function (CDF) as follows

$$\mathbb{E}[\mathbf{X}^r] = -r \int_{-\infty}^0 x^{r-1} F(x) dx + r \int_0^{\infty} x^{r-1} (1 - F(x)) dx . \quad (5.2)$$

Now, assuming \mathbf{X} is piece-wise normally distributed in n samples, $\mathbb{E}[\mathbf{X}]$ can be approximated by

$$\mathbb{E}[\mathbf{X}^r] \approx r \sum_{i=1}^{n-1} g_r(x_{(i)}, x_{(i+1)} | \mu_i, \sigma_i) + c \quad (5.3)$$

$$g_r(x_{(i)}, x_{(i+1)} | \mu_i, \sigma_i) = \begin{cases} - \int_{x_{(i)}}^{x_{(i+1)}} x^{r-1} \Phi(x | \mu_i, \sigma_i^2) dx & \text{if } 0 \leq x_{(i)} \leq x_{(i+1)} \\ - \int_0^{x_{(i)}} x^{r-1} \Phi(x | \mu_i, \sigma_i^2) dx & \text{if } x_{(i)} \leq 0 \leq x_{(i+1)} \\ + \int_0^{x_{(i+1)}} x^{r-1} (1 - \Phi(x | \mu_i, \sigma_i^2)) dx \\ \int_{x_{(i)}}^{x_{(i+1)}} x^{r-1} (1 - \Phi(x | \mu_i, \sigma_i^2)) dx & \text{if } x_{(i)} \leq x_{(i+1)} \leq 0 \end{cases} \quad (5.4)$$

where Φ is the standard normal CDF and $x_{(i)}$ is the i th order statistic. By subtracting out the minimum value $x_{(1)}$, the discrete distribution becomes non-negative allowing for the significant simplification as follows

$$\mathbb{E}[\mathbf{X}^r] = r \sum_{i=1}^{n-1} \int_{x_{(i)} - x_{(1)}}^{x_{(i+1)} - x_{(1)}} x^{r-1} \left[1 - \Phi \left(\frac{x - \mu_i}{\sigma_i} \right) \right] dx + x_{(1)} \delta_{i-1} \quad (5.5)$$

The moments can now be computed in closed form for all r . Results for $r \in \{1, 2, 3, 4\}$ are given as follows.

$$\mathbb{E}[\mathbf{X}] = \frac{1}{2} \sum_{i=1}^{n-1} \left[x - \sigma_i \sqrt{\frac{2}{\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} + (\mu_i - x) \operatorname{erf}\left(\frac{x-\mu_i}{\sigma_i \sqrt{2}}\right) \right] \Big|_{x_{(i)} - x_{(1)}}^{x_{(i+1)} - x_{(1)}} + x_{(1)} \quad (5.6)$$

$$\begin{aligned} \mathbb{E}[\mathbf{X}^2] &= \frac{1}{2} \sum_{i=1}^{n-1} \left[x^2 - \sigma_i (\mu_i + x) \sqrt{\frac{2}{\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \right. \\ &\quad \left. + (\mu_i^2 + \sigma_i^2 - x^2) \operatorname{erf}\left(\frac{x-\mu_i}{\sigma_i \sqrt{2}}\right) \right] \Big|_{x_{(i)} - x_{(1)}}^{x_{(i+1)} - x_{(1)}} \end{aligned} \quad (5.7)$$

$$\begin{aligned} \mathbb{E}[\mathbf{X}^3] &= \frac{1}{2} \sum_{i=1}^{n-1} \left[x^3 - \sigma_i (\mu_i^2 + 2\sigma_i^2 + \mu_i x + x^2) \sqrt{\frac{2}{\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \right. \\ &\quad \left. + (\mu_i^3 + 3\mu_i \sigma_i^2 - x^3) \operatorname{erf}\left(\frac{x-\mu_i}{\sigma_i \sqrt{2}}\right) \right] \Big|_{x_{(i)} - x_{(1)}}^{x_{(i+1)} - x_{(1)}} \end{aligned} \quad (5.8)$$

$$\begin{aligned} \mathbb{E}[\mathbf{X}^4] &= \frac{1}{2} \sum_{i=1}^{n-1} \left[x^4 - \sigma_i (\mu_i^3 + 5\mu_i \sigma_i^2 + \mu_i^2 x + 3\sigma_i^2 x + \mu_i x^2 + x^3) \sqrt{\frac{2}{\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \right. \\ &\quad \left. + (\mu_i^4 + 6\mu_i^2 \sigma_i^2 + 3\sigma_i^4 - x^4) \operatorname{erf}\left(\frac{x-\mu_i}{\sigma_i \sqrt{2}}\right) \right] \Big|_{x_{(i)} - x_{(1)}}^{x_{(i+1)} - x_{(1)}} \end{aligned} \quad (5.9)$$

The mean, standard deviation, skewness and kurtosis can then be computed from these expectations with

$$\mu = \mathbb{E}[\mathbf{X}] \quad (\text{mean}) \quad (5.10)$$

$$\sigma = \sqrt{\mathbb{E}[\mathbf{X}^2] - \mathbb{E}[\mathbf{X}]^2} \quad (\text{standard deviation}) \quad (5.11)$$

$$\gamma = \frac{2\mathbb{E}[\mathbf{X}]^3 - 3\mathbb{E}[\mathbf{X}]\mathbb{E}[\mathbf{X}^2] + \mathbb{E}[\mathbf{X}^3]}{\sigma^3} \quad (\text{skewness}) \quad (5.12)$$

$$\kappa = \frac{-3\mathbb{E}[\mathbf{X}]^4 + 6\mathbb{E}[\mathbf{X}^2]\mathbb{E}[\mathbf{X}]^2 - 4\mathbb{E}[\mathbf{X}]\mathbb{E}[\mathbf{X}^3] + \mathbb{E}[\mathbf{X}^4]}{\sigma^4} \quad (\text{kurtosis}) . \quad (5.13)$$

5.2 Experiments

In order to test the quantile based moment calculation method developed in the previous section, it is compared against the sample based moment computation for a wide range of probability distributions. For each probability distribution setting, 10K trials of 10K samples are simulated and the sample and quantile based moment calculations methods are compared.

Tables 5.1, 5.2, 5.3, and 5.4 provide the mean, standard deviation, skewness, and kurtosis comparisons, respectively. Further details about the statistic distribution for selected probability distributions across trials are given in Figures 5.1-5.8. Where a red line appears on the plot, it indicates the given population statistic.

5.2. EXPERIMENTS

Table 5.1: Sample vs. Quantile Based Moment Computation : Mean

Prob Distribution	(Population)	Sample	Quantile	Abs. Error
Normal ($\mu = 1$)	0.000e+00	-7.185e-05	-6.890e-05	2.952e-06
Normal ($\mu = 2$)	0.000e+00	4.203e-04	4.189e-04	1.403e-06
Normal ($\mu = 4$)	0.000e+00	-1.319e-03	-1.321e-03	1.873e-06
Log Normal ($s = 0.5$)	1.133e+00	1.133e+00	1.133e+00	9.428e-06
Log Normal ($s = 1$)	1.649e+00	1.648e+00	1.648e+00	1.231e-04
Log Normal ($s = 1.5$)	3.080e+00	3.078e+00	3.076e+00	1.222e-03
Skew Normal ($\alpha = -4$)	-7.741e-01	-7.742e-01	-7.742e-01	3.456e-06
Skew Normal ($\alpha = -2$)	-7.136e-01	-7.137e-01	-7.137e-01	3.971e-06
Skew Normal ($\alpha = 0$)	0.000e+00	-5.397e-05	-5.317e-05	8.020e-07
Skew Normal ($\alpha = 2$)	7.136e-01	7.136e-01	7.136e-01	3.616e-06
Skew Normal ($\alpha = 4$)	7.741e-01	7.741e-01	7.741e-01	3.247e-06
Weibull ($\lambda = 1, k = 1$)	1.000e+00	9.998e-01	9.998e-01	1.150e-05
Weibull ($\lambda = 1, k = 2$)	8.862e-01	8.861e-01	8.861e-01	2.446e-06
Weibull ($\lambda = 1, k = 4$)	9.064e-01	9.064e-01	9.064e-01	9.057e-08
Beta ($a = 1.5, b = 1$)	6.000e-01	5.999e-01	5.999e-01	3.571e-08
Beta ($a = 1.5, b = 2$)	4.286e-01	4.286e-01	4.286e-01	2.837e-08
Beta ($a = 1.5, b = 4$)	2.727e-01	2.727e-01	2.727e-01	3.987e-07
Student's t ($\nu = 2$)	0.000e+00	-1.987e-04	-1.214e-04	7.733e-05
Student's t ($\nu = 4$)	0.000e+00	-1.498e-04	-1.737e-04	2.391e-05
Student's t ($\nu = \infty$)	0.000e+00	-1.871e-05	-1.901e-05	3.011e-07
Chi-square ($k = 1$)	1.000e+00	1.000e+00	1.000e+00	2.909e-05
Chi-square ($k = 3$)	3.000e+00	3.000e+00	3.000e+00	2.961e-05
Chi-square ($k = 6$)	1.600e+01	1.600e+01	1.600e+01	1.528e-05

Table 5.2: Sample vs. Quantile Based Moment Computation : Standard Deviation

Prob Distribution	(Population)	Sample	Quantile	Abs. Error
Normal ($\mu = 1$)	1.000e+00	9.999e-01	9.998e-01	7.887e-05
Normal ($\mu = 2$)	2.000e+00	2.000e+00	2.000e+00	1.475e-04
Normal ($\mu = 4$)	4.000e+00	4.000e+00	3.999e+00	3.110e-04
Log Normal ($s = 0.5$)	6.039e-01	6.036e-01	6.035e-01	1.565e-04
Log Normal ($s = 1$)	2.161e+00	2.147e+00	2.144e+00	3.415e-03
Log Normal ($s = 1.5$)	8.974e+00	8.408e+00	8.356e+00	5.197e-02
Skew Normal ($\alpha = -4$)	6.331e-01	6.332e-01	6.331e-01	5.355e-05
Skew Normal ($\alpha = -2$)	7.005e-01	7.005e-01	7.004e-01	6.386e-05
Skew Normal ($\alpha = 0$)	1.000e+00	1.000e+00	9.999e-01	7.908e-05
Skew Normal ($\alpha = 2$)	7.005e-01	7.005e-01	7.005e-01	6.246e-05
Skew Normal ($\alpha = 4$)	6.331e-01	6.331e-01	6.330e-01	5.593e-05
Weibull ($\lambda = 1, k = 1$)	1.000e+00	9.997e-01	9.996e-01	1.648e-04
Weibull ($\lambda = 1, k = 2$)	4.633e-01	4.632e-01	4.632e-01	3.560e-05
Weibull ($\lambda = 1, k = 4$)	2.543e-01	2.542e-01	2.542e-01	1.587e-05
Beta ($a = 1.5, b = 1$)	2.619e-01	2.619e-01	2.619e-01	1.319e-05
Beta ($a = 1.5, b = 2$)	2.333e-01	2.333e-01	2.333e-01	1.187e-05
Beta ($a = 1.5, b = 4$)	1.747e-01	1.747e-01	1.747e-01	1.025e-05
Student's t ($\nu = 2$)	inf	3.111e+00	3.077e+00	3.419e-02
Student's t ($\nu = 4$)	1.414e+00	1.410e+00	1.409e+00	1.273e-03
Student's t ($\nu = \infty$)	1.000e+00	1.000e+00	9.999e-01	9.277e-05
Chi-square ($k = 1$)	1.414e+00	1.414e+00	1.414e+00	3.504e-04
Chi-square ($k = 3$)	2.449e+00	2.449e+00	2.449e+00	3.625e-04
Chi-square ($k = 6$)	5.657e+00	5.656e+00	5.656e+00	5.113e-04

5.2. EXPERIMENTS

Table 5.3: Sample vs. Quantile Based Moment Computation : Skewness

Prob Distribution	(Population)	Sample	Quantile	Abs. Error
Normal ($\mu = 1$)	0.000e+00	1.967e-04	2.062e-04	9.429e-06
Normal ($\mu = 2$)	0.000e+00	2.818e-04	2.962e-04	1.436e-05
Normal ($\mu = 4$)	0.000e+00	-1.267e-04	-1.390e-04	1.229e-05
Log Normal ($s = 0.5$)	1.750e+00	1.731e+00	1.727e+00	4.295e-03
Log Normal ($s = 1$)	6.185e+00	5.332e+00	5.271e+00	6.109e-02
Log Normal ($s = 1.5$)	3.347e+01	1.360e+01	1.322e+01	3.798e-01
Skew Normal ($\alpha = -4$)	-7.844e-01	-7.840e-01	-7.835e-01	4.858e-04
Skew Normal ($\alpha = -2$)	-4.538e-01	-4.536e-01	-4.533e-01	2.781e-04
Skew Normal ($\alpha = 0$)	0.000e+00	4.725e-05	2.779e-05	1.945e-05
Skew Normal ($\alpha = 2$)	4.538e-01	4.536e-01	4.532e-01	3.334e-04
Skew Normal ($\alpha = 4$)	7.844e-01	7.833e-01	7.828e-01	4.895e-04
Weibull ($\lambda = 1, k = 1$)	2.000e+00	1.991e+00	1.988e+00	2.866e-03
Weibull ($\lambda = 1, k = 2$)	6.311e-01	6.306e-01	6.302e-01	3.924e-04
Weibull ($\lambda = 1, k = 4$)	-8.724e-02	-8.730e-02	-8.734e-02	4.415e-05
Beta ($a = 1.5, b = 1$)	-3.395e-01	-3.391e-01	-3.390e-01	5.217e-05
Beta ($a = 1.5, b = 2$)	2.227e-01	2.225e-01	2.224e-01	3.692e-05
Beta ($a = 1.5, b = 4$)	6.939e-01	6.941e-01	6.939e-01	1.579e-04
Student's t ($\nu = 2$)	undefined	-9.000e-02	-1.033e-01	1.329e-02
Student's t ($\nu = 4$)	0.000e+00	2.850e-03	3.402e-03	5.519e-04
Student's t ($\nu = \infty$)	0.000e+00	2.066e-05	2.680e-05	6.143e-06
Chi-square ($k = 1$)	2.828e+00	2.811e+00	2.806e+00	5.238e-03
Chi-square ($k = 3$)	1.633e+00	1.628e+00	1.626e+00	2.253e-03
Chi-square ($k = 6$)	7.071e-01	7.055e-01	7.048e-01	7.431e-04

Table 5.4: Sample vs. Quantile Based Moment Computation : Kurtosis

Prob Distribution	(Population)	Sample	Quantile	Abs. Error
Normal ($\mu = 1$)	3.000e+00	2.999e+00	2.997e+00	2.064e-03
Normal ($\mu = 2$)	3.000e+00	2.999e+00	2.996e+00	2.159e-03
Normal ($\mu = 4$)	3.000e+00	2.999e+00	2.997e+00	2.307e-03
Log Normal ($s = 0.5$)	8.898e+00	8.501e+00	8.441e+00	5.982e-02
Log Normal ($s = 1$)	1.139e+02	5.795e+01	5.610e+01	1.858e+00
Log Normal ($s = 1.5$)	1.008e+04	3.270e+02	3.070e+02	2.003e+01
Skew Normal ($\alpha = -4$)	3.633e+00	3.628e+00	3.624e+00	3.943e-03
Skew Normal ($\alpha = -2$)	3.305e+00	3.302e+00	3.299e+00	2.941e-03
Skew Normal ($\alpha = 0$)	3.000e+00	2.999e+00	2.997e+00	2.178e-03
Skew Normal ($\alpha = 2$)	3.305e+00	3.302e+00	3.299e+00	3.113e-03
Skew Normal ($\alpha = 4$)	3.633e+00	3.627e+00	3.623e+00	3.954e-03
Weibull ($\lambda = 1, k = 1$)	9.000e+00	8.846e+00	8.808e+00	3.835e-02
Weibull ($\lambda = 1, k = 2$)	3.245e+00	3.240e+00	3.237e+00	2.942e-03
Weibull ($\lambda = 1, k = 4$)	2.748e+00	2.749e+00	2.748e+00	1.011e-03
Beta ($a = 1.5, b = 1$)	2.051e+00	2.050e+00	2.050e+00	1.293e-04
Beta ($a = 1.5, b = 2$)	2.140e+00	2.140e+00	2.140e+00	1.849e-04
Beta ($a = 1.5, b = 4$)	2.931e+00	2.930e+00	2.929e+00	9.399e-04
Student's t ($\nu = 2$)	undefined	3.613e+02	3.213e+02	3.998e+01
Student's t ($\nu = 4$)	undefined	1.129e+01	1.093e+01	3.625e-01
Student's t ($\nu = \infty$)	3.000e+00	2.999e+00	2.997e+00	1.966e-03
Chi-square ($k = 1$)	1.500e+01	1.461e+01	1.451e+01	9.566e-02
Chi-square ($k = 3$)	7.000e+00	6.921e+00	6.896e+00	2.450e-02
Chi-square ($k = 6$)	3.750e+00	3.736e+00	3.730e+00	6.080e-03

5.2. EXPERIMENTS

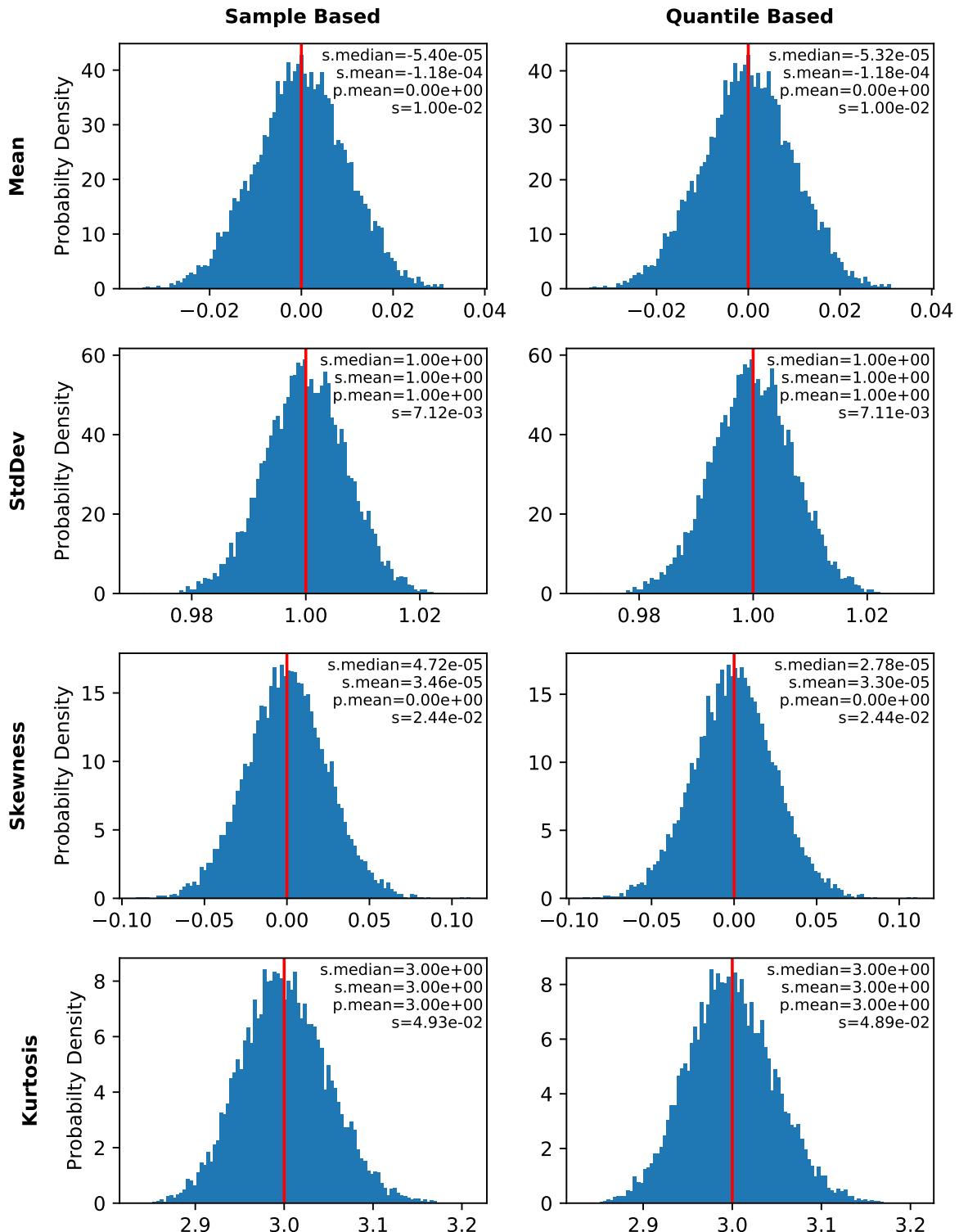
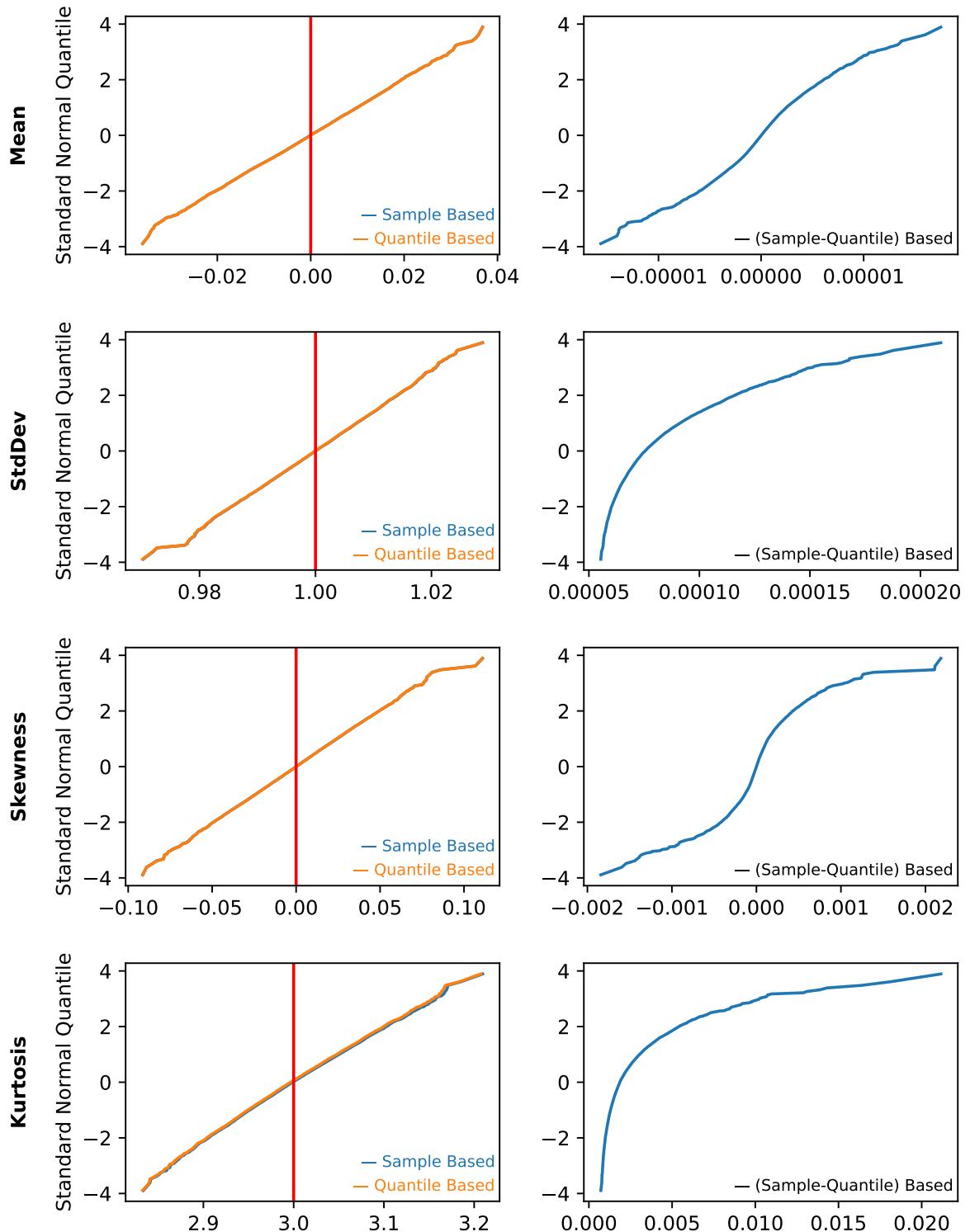


Figure 5.1: Moment Trials Density : Normal ($\mu = 1$)


 Figure 5.2: Moment Trials QQ Plot and Delta QQ Plot : Normal ($\mu = 1$)

5.2. EXPERIMENTS

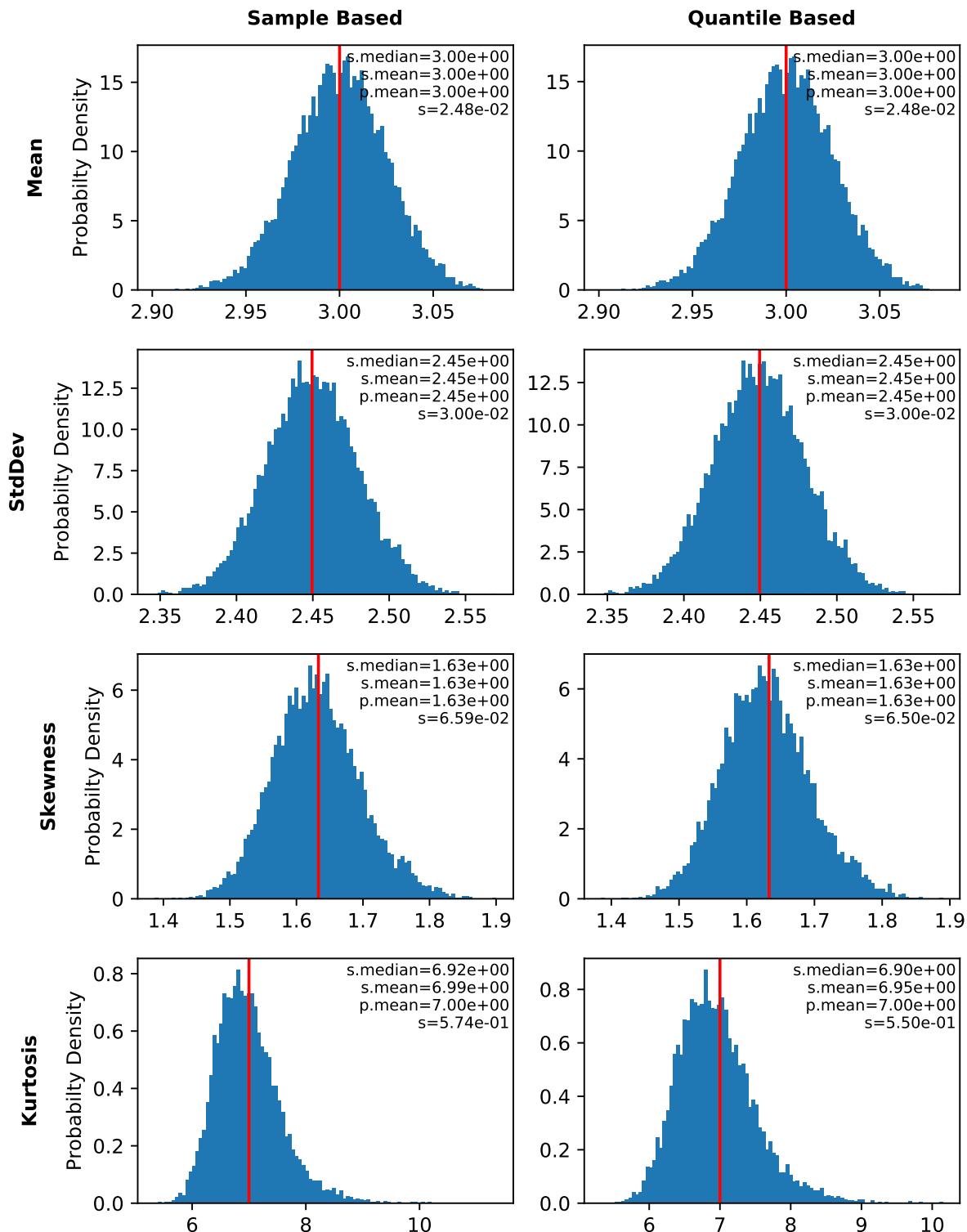
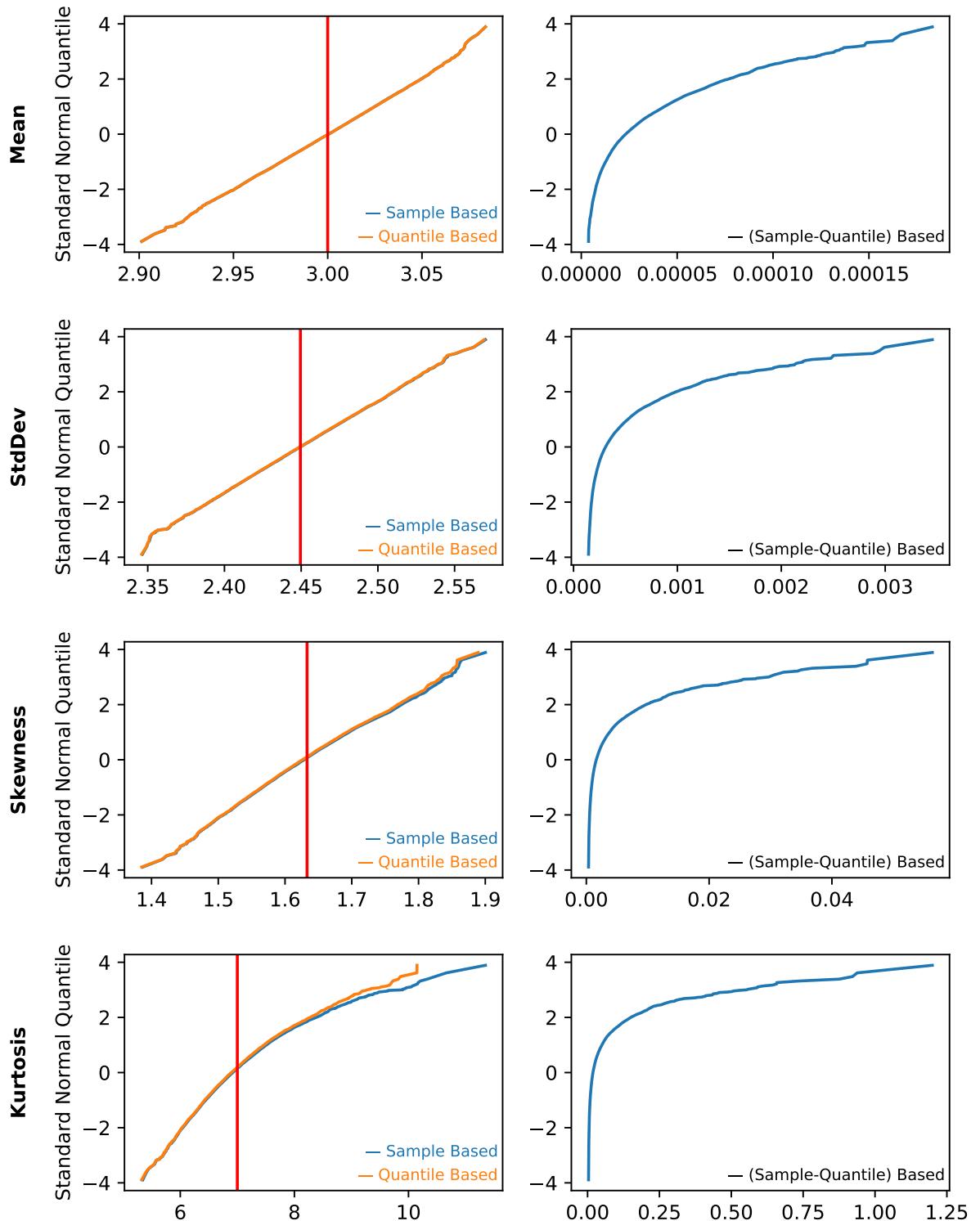


Figure 5.3: Moment Trials Density : Chi-square ($k = 3$)


 Figure 5.4: Moment Trials QQ Plot and Delta QQ Plot : Chi-square ($k = 3$)

5.2. EXPERIMENTS

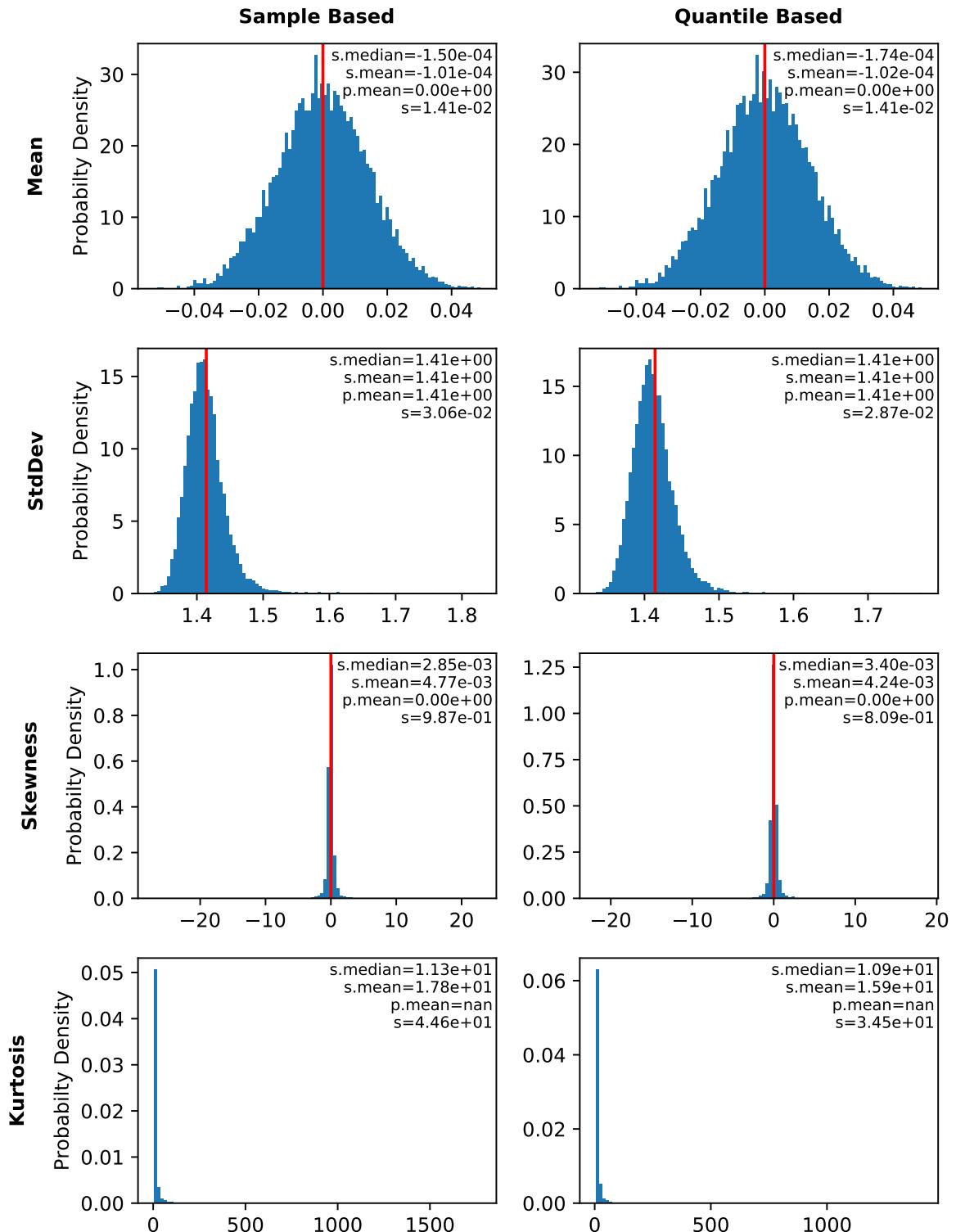
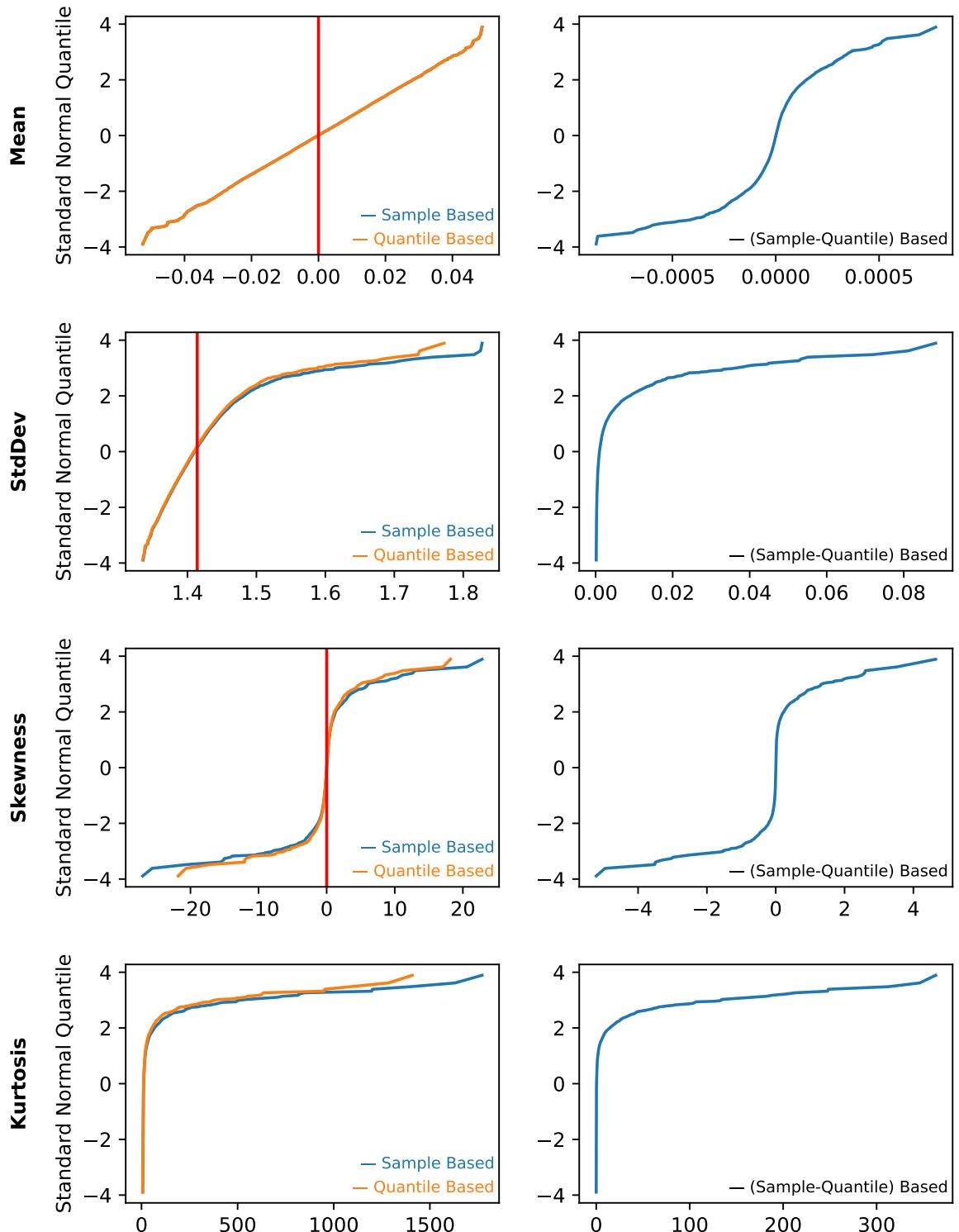


Figure 5.5: Moment Trials Density : Student's t ($\nu = 4$)


 Figure 5.6: Moment Trials QQ Plot and Delta QQ Plot : Student's t ($\nu = 4$)

5.2. EXPERIMENTS

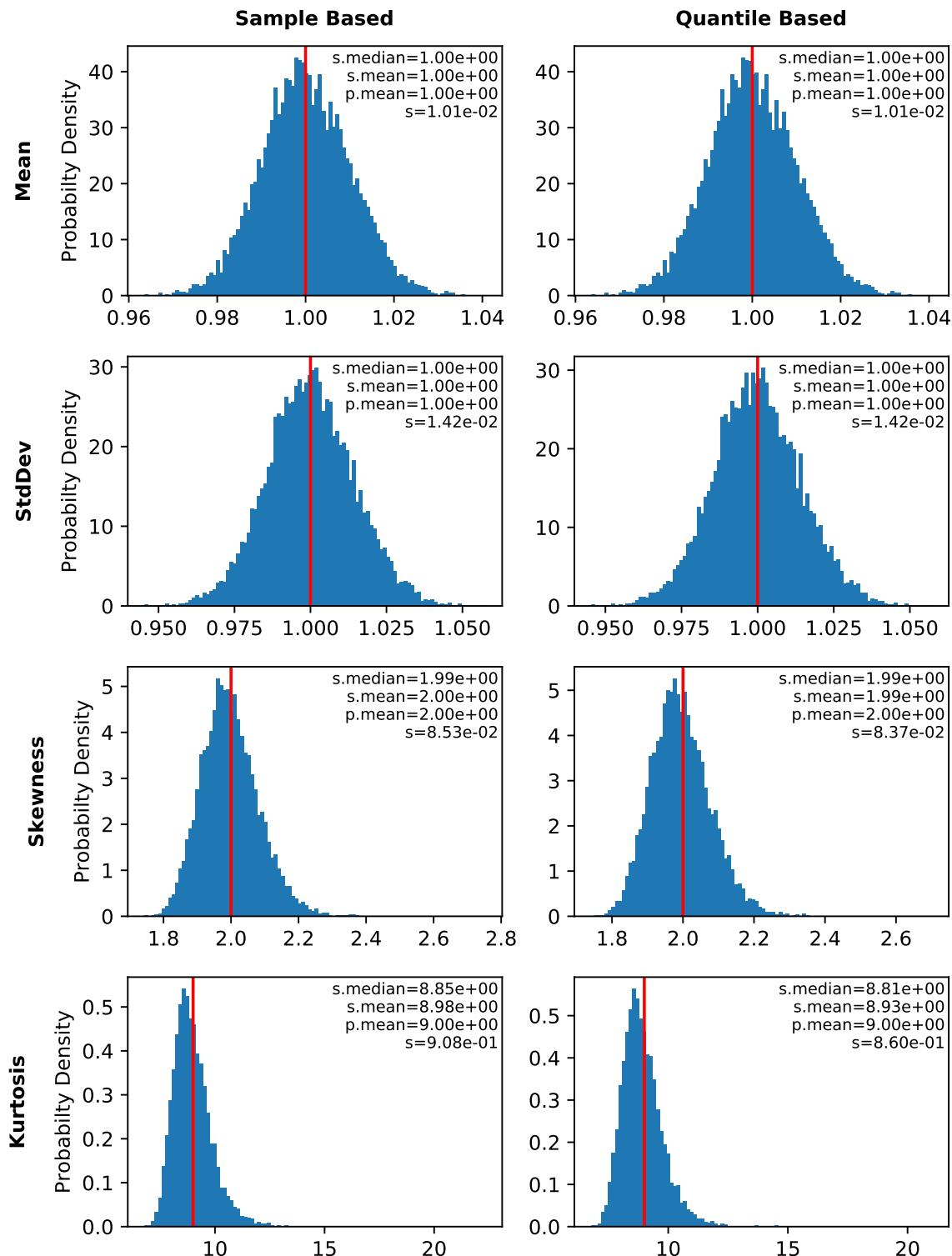
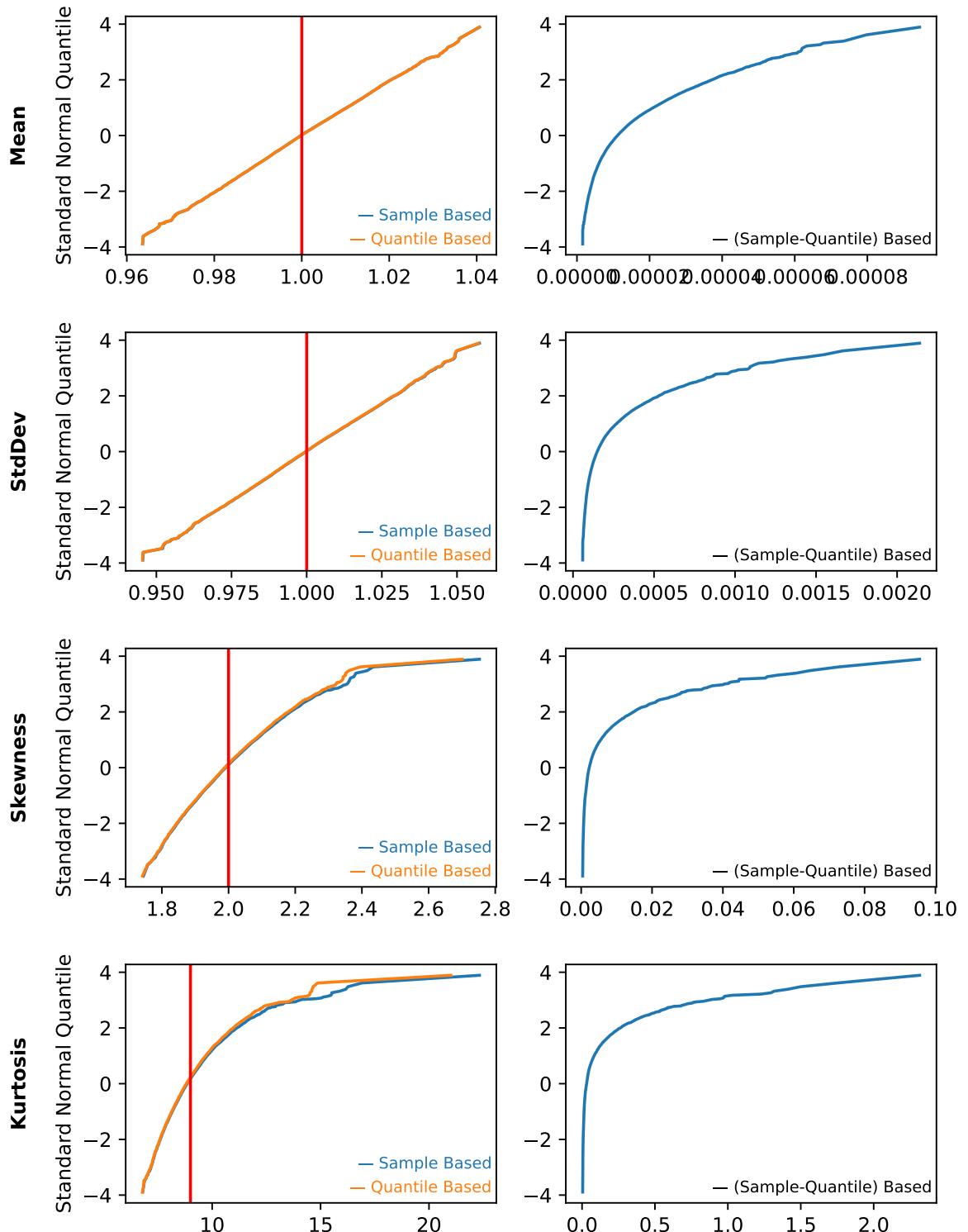


Figure 5.7: Moment Trials Density : Weibull ($\lambda = 1, k = 1$)


 Figure 5.8: Moment Trials QQ Plot and Delta QQ Plot : Weibull ($\lambda = 1, k = 1$)

5.3 Discussion

Quantile based moment calculation, developed in this chapter, is a useful method to approximate sample moments. While the technique is inherently predicated upon the local normality assumption, it is shown to work well for a wide variety of random variable distributions. As is evident from the experimental results, the accuracy of quantile based moment calculations is greater for lower order moments than higher order moments with respect to the sample moment reference. The even order moments appear to have a small positive bias, though this is an artifact of the even order exponents in the quantile based moment calculation. As an approximation to the sample moments, the quantile based moments provide an alternative method of moment calculation when the full sample dataset is not available.

Chapter 6

Applications

This chapter deals with applications of the previously defined methods in the realm of Electronic Design Automation (EDA). Specifically, we consider SPICE circuit simulation. Circuit designers are interested in distribution properties of circuit measurements, as they are directly related to the predicted yield or circuit performance.

For an SRAM bit cell designer, the distribution of read current and noise margin will be very important. For a standard cell logic designer, the distribution of delay and slope will be very important.

In the following sections, the methods of In-order Multivariate Sampling and quantile-based moment calculation are applied to these problems. For In-order Multivariate Sampling, it is shown that the number of Monte Carlo samples required to accurately characterize the tail behavior is greatly reduced compared to simple random sampling. The quantile-based moment calculations are also provided and compared against standard calculation methods.

6.1 Circuits

In this chapter, the following circuits described in Table 6.1 are used for testing the proposed algorithms in Chapters 3-5. These are representative circuits of the larger class of standard cell circuits which are the target circuit type for our investigation. The variation in standard cell circuits, and thus the desire to understand their statistics by Monte Carlo analysis, is driven by their high rate of replication in chip design.

Table 6.1: Circuit Characteristics

Index	Circuit	Node	FET	Res	Cap
1	Inverter	7nm	6	151	259
2	AOI	16nm	80	1064	2600
3	Clock Gate	16nm	122	1879	3578
4	Operational Amplifier	0.2 μ m	8	4	3

6.2 Monotonicity Verification

In Section 3.2, the motivation of the efficacy is built upon the correspondence between rare input and output samples. This correspondence is predicated on the monotonic behavior of the response. In this section, we provide evidence that the monotonicity property holds for many circuits. In addition to the examples given in this section, several dozen more standard cell circuits based on various foundry technologies were tested and no cases of non-monotonic behavior was found.

It should be noted that semiconductor circuits can be non-monotonic in the design space or with respect to external inputs such as frequency or supply voltage for example. When we verify the monotonic property in this section, what is being verified is the monotonicity of the circuit response with respect to the process variation only. In each plot, the value of the independent random variable (IRV) is given on the x-axis while the response is given on the y-axis. In some cases there is a clamp on the output response with respect to a given input and in this condition the circuit response appears flat for IRV values above or below a given value. While such response limits are common in circuit simulation, they do not violate the (weak) monotonic property.

6.2. MONOTONICITY VERIFICATION

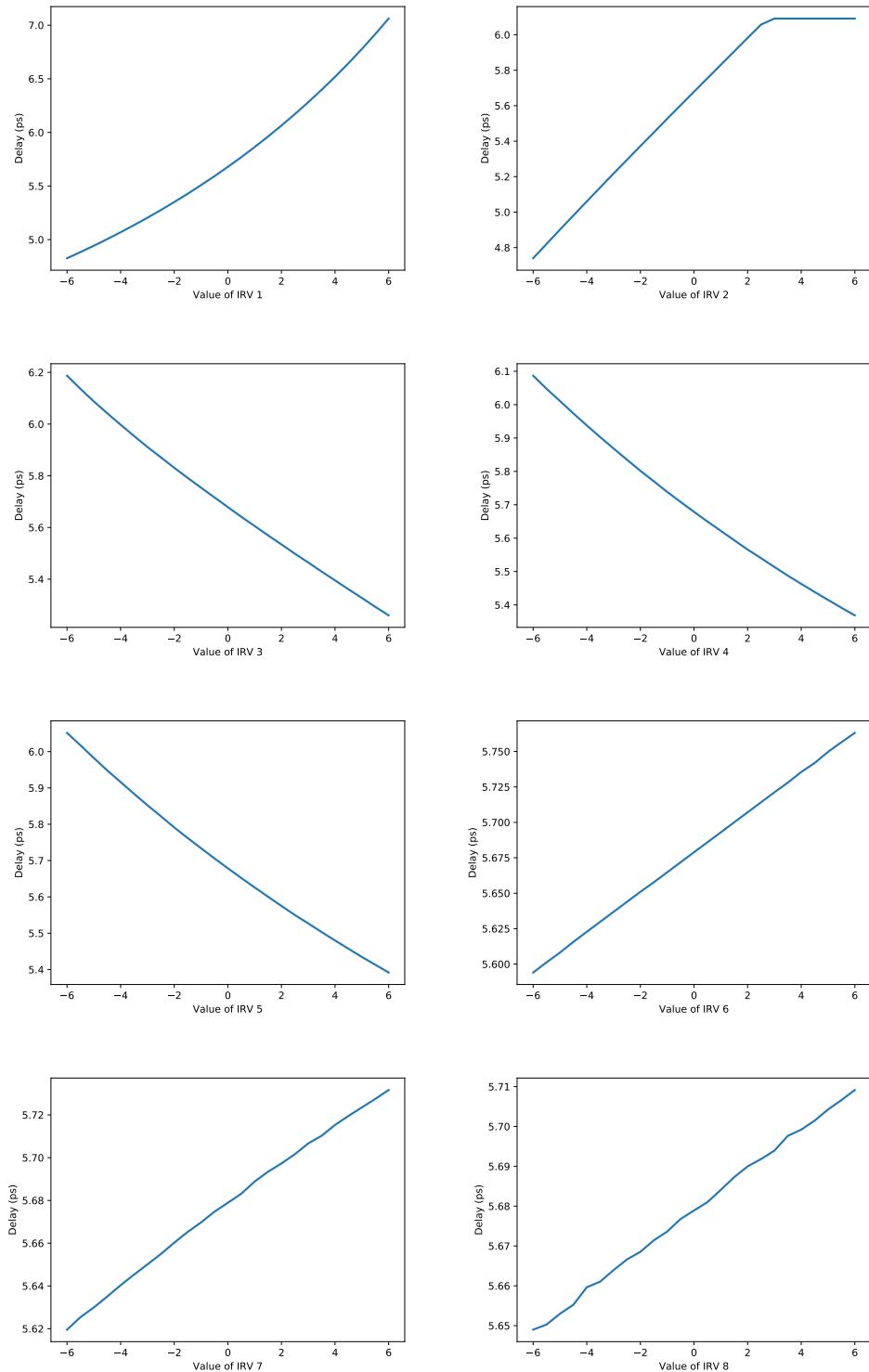


Figure 6.1: Monotonicity plots of delay measurement of an inverter circuit in 16nm technology

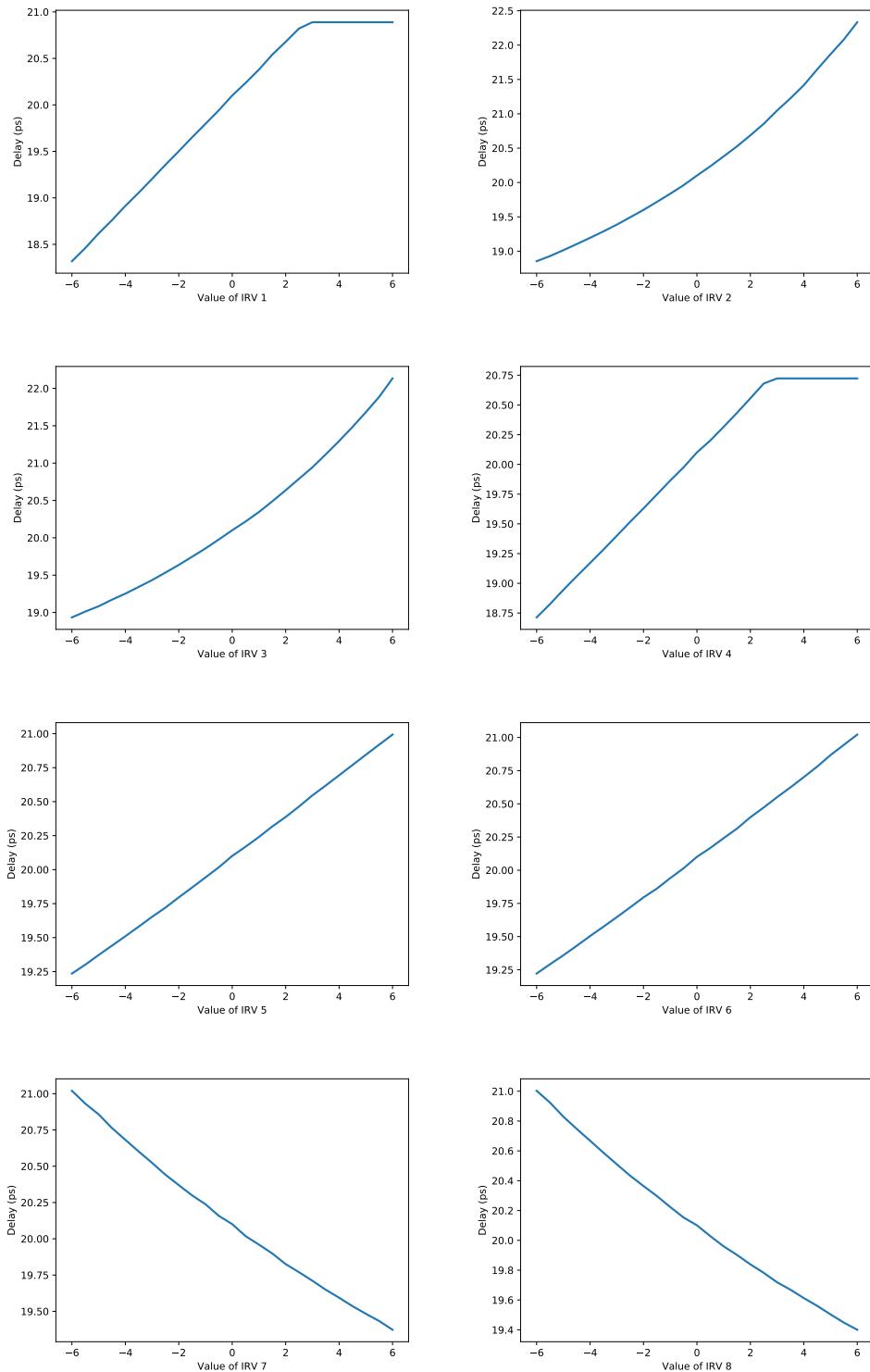


Figure 6.2: Monotonicity plots of delay measurement of a MAJ circuit in 16nm technology

6.2. MONOTONICITY VERIFICATION

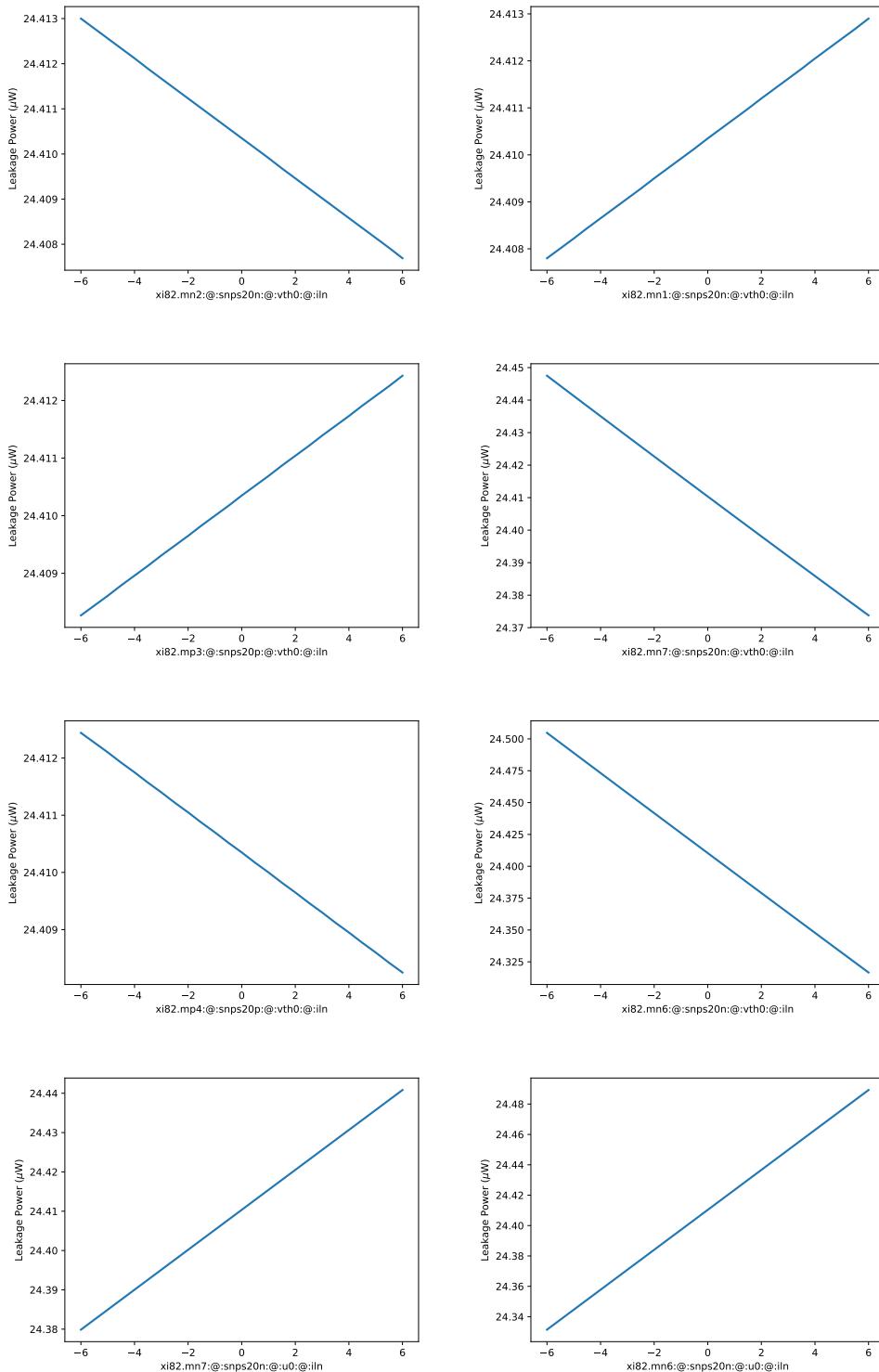


Figure 6.3: Monotonicity plots of delay measurement of a operational amplifier circuit in $0.2 \mu\text{m}$ technology

6.3 In-Order Multivariate Sampling Experiments

In Chapter 3, it was shown that, assuming a linear model for the output response, there is a direct theoretical justification that rare samples in the output space correspond to rare samples in the inputs space. It was further argued that this principle also holds when the output response is monotonic.

In this experiment, we investigate the idea that rare inputs correspond well to rare output responses in the context of circuit simulation. Figures 6.4 - 6.6 show an overlay of the measure value distribution and sample count. The multivariate standard normal input distance from the origin, which is directly related to the input probability density, is recorded on the X axis. The colored lines represent the measure value at a given percentile and is shown on the left-Y axis. The black bell curve corresponds to sample count which is shown on the right-Y axis.

The results confirm that rare samples for the input space do indeed correspond to rare samples in the output space for real circuits. More specifically, when measuring delay for the inverter circuit at in Figure 6.4, delay measurements near the tail of the distribution tend to occur at input samples (drawn from a multivariate normal distribution) which lie further from the origin.

6.3. IN-ORDER MULTIVARIATE SAMPLING EXPERIMENTS

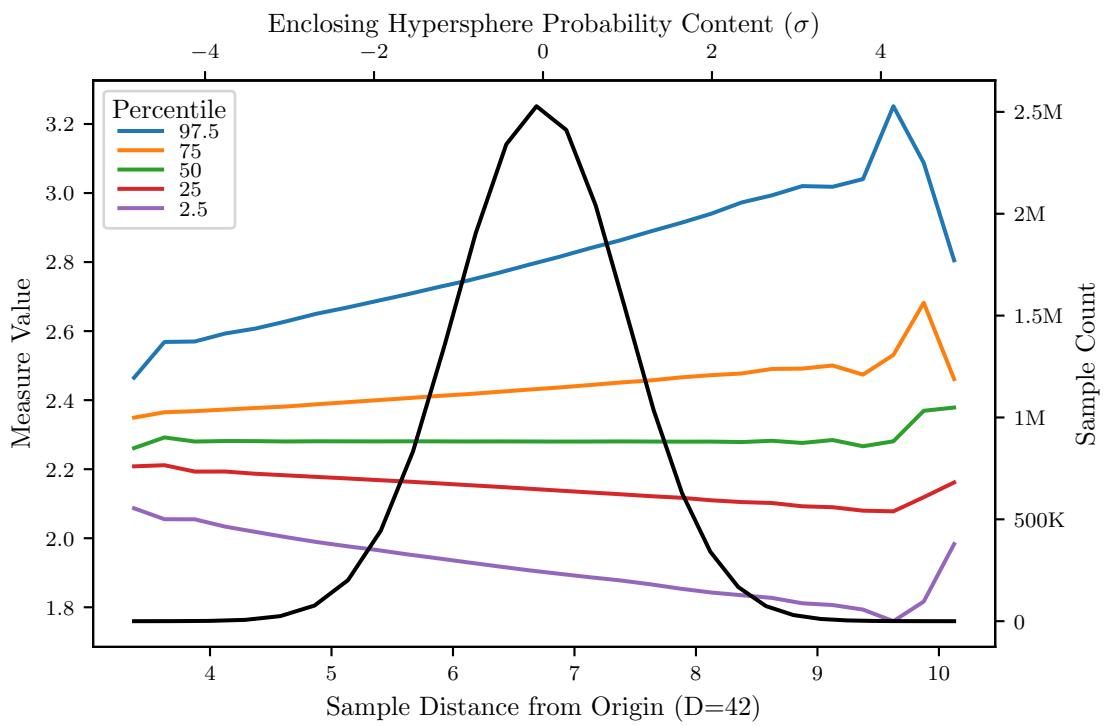


Figure 6.4: Origin Distance versus Measure Percentile : 7nm Inverter Delay

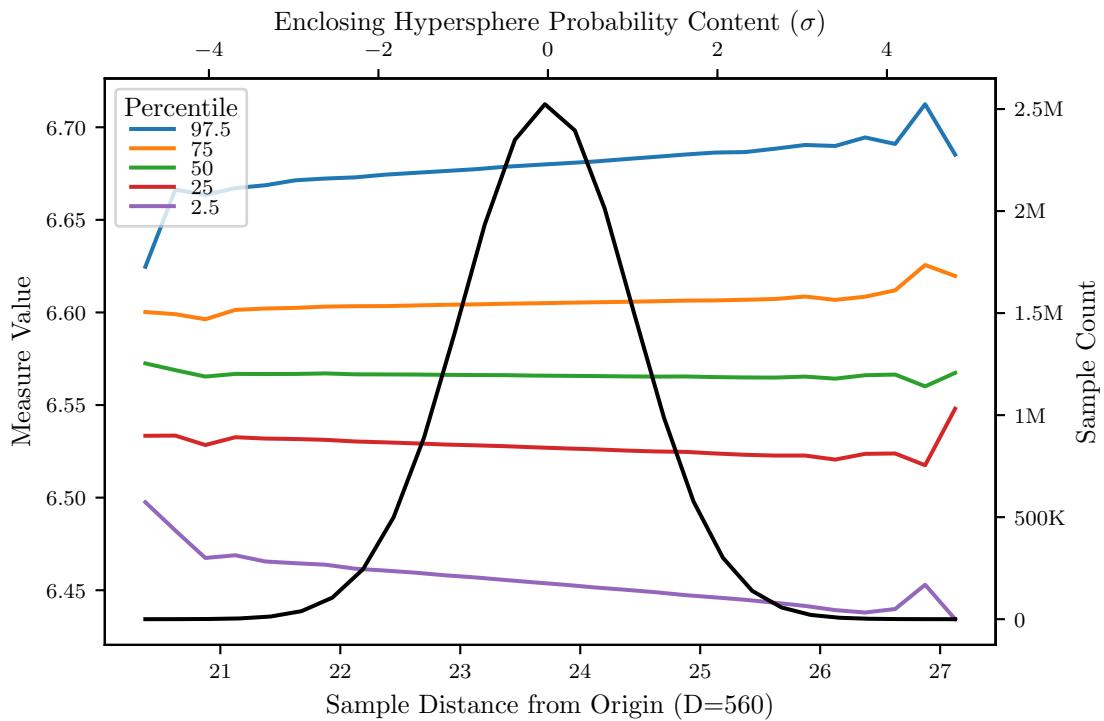


Figure 6.5: Origin Distance versus Measure Percentile : 16nm AOI Delay

6.4. LINEAR AND QUADRATIC MODELING WITH DIRECT MODEL-BASED QUANTILE COMPUTATION

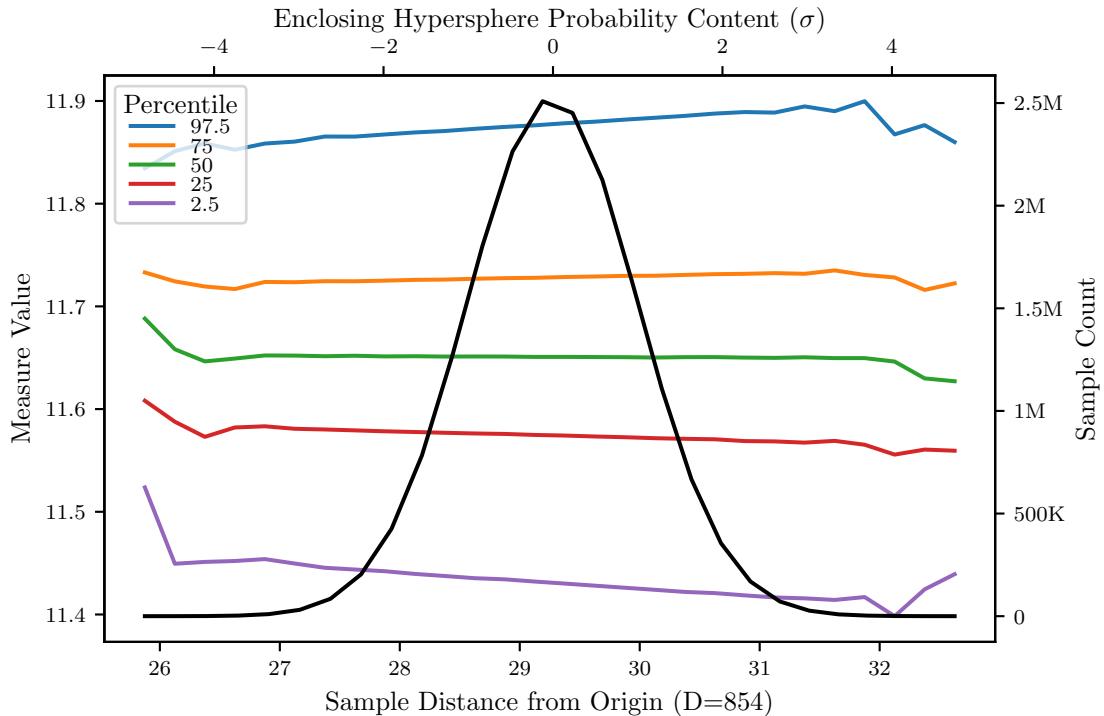


Figure 6.6: Origin Distance versus Measure Percentile : 16nm CKGTPLT Measure

6.4 Linear and Quadratic Modeling with Direct Model-Based Quantile Computation

In this section, we demonstrate the ability to compute quantiles directly from quadratic models for circuit measures. The flow follows that developed in Chapter 4, from quadratic model extraction via sparse Legendre polynomials through quantile estimation via Cornish-Fisher expansion. Figure 6.7 represents the distribution of leakage power of an operational amplifier. The quadratic model is better

able to capture the distribution than the linear. From this example, we find along with previous study of quadratic modeling in Chapter 4, the quadratic form can provide good approximations for certain classes of maps of the multivariate normal distribution, but is not a universal method.

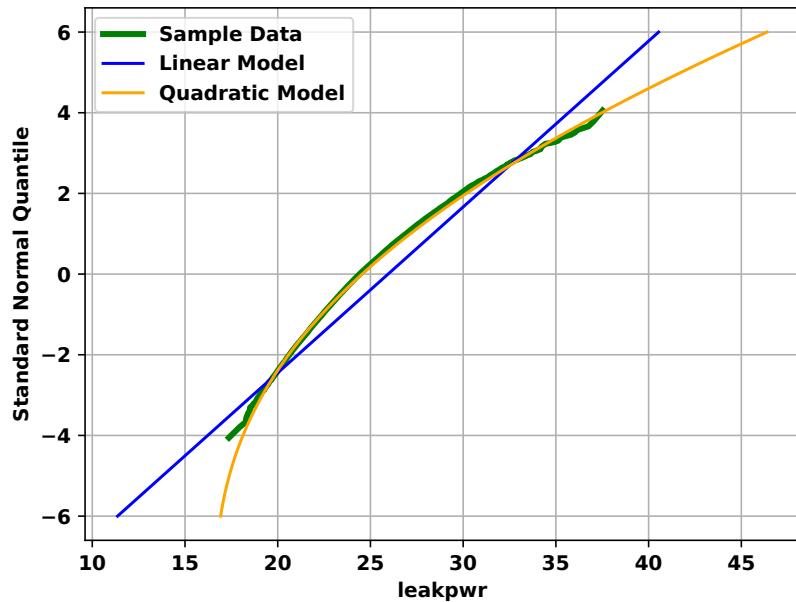


Figure 6.7: QQ plots of MC sample data versus quantiles of linear and quadratic models extracted directly : operational amplifier in 0.2 μm technology

6.5 Quantile-Based Moment Experiments

Note that while the full dataset has 18M samples, the quantile information consist of just 3K samples. This is a significant compression, making computation of moments from quantile information fast despite their rather complicated formulation.

6.5. QUANTILE-BASED MOMENT EXPERIMENTS

The QQ plots of delay measurements for a 7nm inverter circuit is shown in Figure 6.8 and a comparison of the sample and quantile moments is given in Table 6.2

Table 6.2: Comparison of sample and quantile-based moments for a 7nm inverter circuit

	Sample Moments		Quantile Moments		Error	
	meas_1	meas_2	meas_1	meas_2	meas_1	meas_2
Mean	2.2973	6.6810	2.2941	6.6734	0.136%	0.113%
StdDev	0.2213	0.5531	0.2213	0.5532	0.004%	0.004%
Skewness	0.5451	0.4579	0.5460	0.4589	0.165%	0.198%
Kurtosis	3.6750	3.4591	3.6769	3.4608	0.051%	0.049%

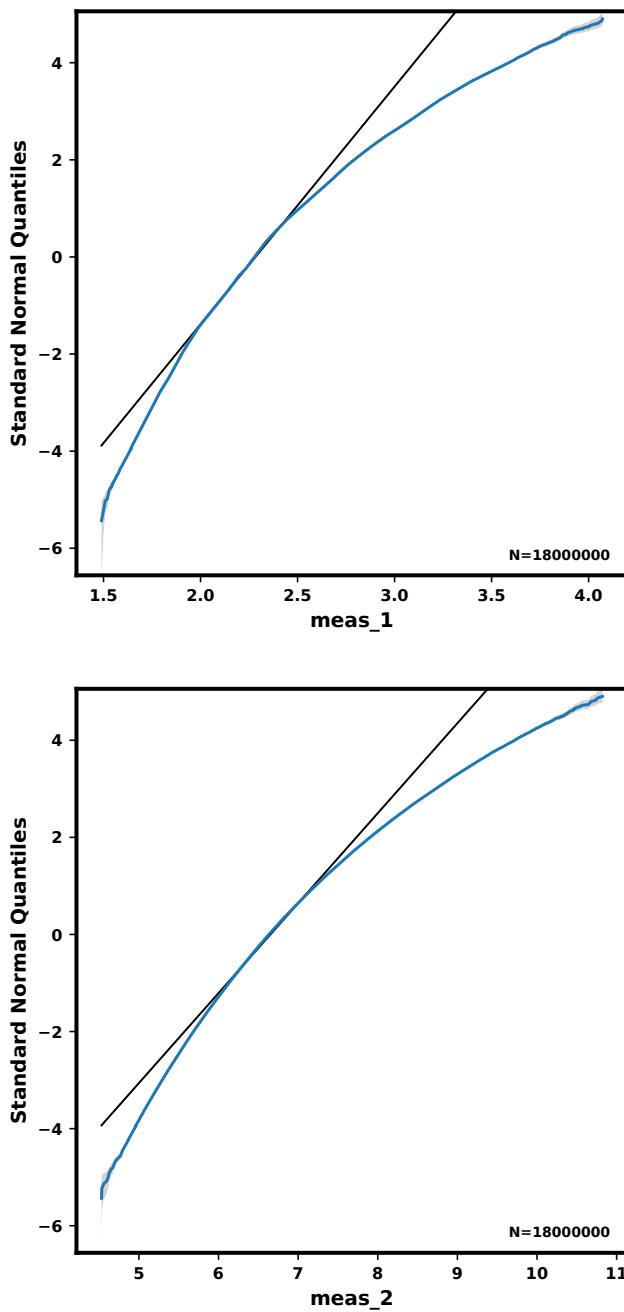


Figure 6.8: QQ plots of two delay measurements of an inverter circuit in 7nm technology

6.6 Discussion

In this chapter, concrete circuit simulation domain examples were provided to reinforce and verify the algorithms proposed in Chapters 3-5. In Section 6.2, it was shown that for many important circuit problems, the black box form presenting the response to process variation is monotone. This monotone property provides the opportunity to perform Monte Carlo sampling in a highly efficient manner using In-order Multivariate Sampling. It is then shown that the probability of finding a sample in the tail region tapers off considerably as distance from the origin of the multivariate standard normal MC sample point diminishes. Because of this, combined with the fact that a majority of simple random MC samples are near the origin, it is directly implied that little sampling is required to find the tail samples of the distribution.

It is then shown that the quantiles of linear and quadratic forms can be computed directly and reliably from the quadratic model itself with no model sampling. While this approach is very convenient, the accuracy of the computed quantiles is naturally limited by the fitting error between the true black box circuit simulation model and the linear or quadratic model. For many circuit measures where the fitting was reasonably good, the agreement between the true model and approximation model was acceptable.

Verification of the quantile-based moment calculation was also carried out for practical circuit examples and the results showed good agreement with that of the sample-based approach. Because the technique does not assume any underlying

distribution, the method can be used to calculate moments from any available CDF or QQ plot in any context.

Chapter 7

Conclusions

7.1 Connections Between Our Proposed Methods

As discussed in Chapter 3, Simple Random Sampling (SRS) are highly concentrated near the origin. If the samples are to be used for machine learning purposes, poor input sample diversity (i.e. low sample count of input samples far from the origin) can lead to poor machine learning models if those sample are to be used for training.

The Uniform Random Sampling (URS) and Hyperspherical Uniform Random Sampling (HURS) methods introduced in Chapter 3 can be used to train machine learning models. The In-order Multivariate Sampling (IMS) method proposed in Chapter 3 can be used to quickly discover rare output distribution samples directly without the need for a surrogate model.

The statistical characterization flow for the quadratic form described in Chapter 4 requires training samples in order to construct the quadratic model. For this

purpose, training samples drawn from URS or HURS may be used rather than samples drawn by SRS, increasing the diversity of the samples in the input space. Once the quadratic model has been extracted, the remaining characterization steps are independent of the novel methods introduced in the thesis.

The computation of distribution moments from quantile information does not depend on the sampling method nor surrogate models such as the quadratic, therefore the quantile-based moment calculation method proposed in Chapter 5 can be considered a self-contained method.

7.2 Connections of Our Proposed Methods with Related Work

7.2.1 Importance Sampling

The importance sampling method, introduced in Section 1.3.1, can be regarded as identical to Simple Random Sampling (SRS) with shifted mean. The methods presented in Chapter 4 (statistical summarization of quadratic forms) and Chapter 5 (moment calculation utilizing quantile information) are not readily applicable to importance sampling. However, because in importance sampling it must be decided by how much and in which direction the mean shift should be performed, a certain number of training samples may be required. In-Order Multivariate Sampling (IMS) may be less useful when used in conjunction with importance sampling as the rare samples drawn from IMS may not have an adequate distribution of samples in the

7.3. THESIS CONTRIBUTIONS

input space to effectively construct a surrogate model. Alternative Monte Carlo sampling methods, such as the Uniform Random Sampling (URS) and Hyperspherical Uniform Random Sampling (HURS) methods introduced in Chapter 3, can be used to produce efficient training samples for such a purpose.

7.2.2 Statistical Blockade

Monte Carlo sampling methods other Simple Random Sampling (SRS) can be used to produce efficient training samples for machine learning models. Statistical Blockade, introduced in Section 1.3.2, can be used together with the Uniform Random Sampling (URS) and Hyperspherical Uniform Random Sampling (HURS) methods introduced in Chapter 3. In general, any improvement in the quality of the surrogate model will lead to more accurate predictions and in the context of Statistical Blockade, this will lead to fewer final simulations for the refined result.

7.3 Thesis Contributions

There are many new concepts, methods, and analyses introduced in the thesis. The In-order Multivariate Sampling (IMS) method introduced in Chapter 3 is one of the most important and novel contributions of our work. It ties together several key ideas to synthesize an innovative sampling method which provides a technique to efficiently study rare events in Monte Carlo. The statistical characterization flow of the quadratic form of the multivariate standard normal defined in Chapter 4 is able to capture the entire quantile range of the given measure distribution, including rare

events at higher sigma. The quantile-based moment computation method introduced in Chapter 5 is not limited to rare events and is a general methods which applies to any continuously distributed dataset, or more specifically the representative quantile data of such a sample dataset. As an extended result, less related to the main body of the thesis, is the consideration of the integrals of the higher order error function which can be found in Appendix A.

The thesis contributions are summarized as follows:

- In-order Multivariate Sampling (IMS)
 - Formulation of a novel method to draw samples from the multivariate standard normal distribution in pre-sorted order
 - Extensions to IMS : directional via von Mises-Fisher distribution, correlated variables
- Flow of Statistical Analysis of Quadratic Forms
 - Fit quadratic models to data using sparse Legendre polynomials
 - Represent the quadratic form as a sum of normal and chi-square distributions
 - Calculation of raw/central moments and cumulants
 - Estimation of quadratic model quantiles via Cornish Fisher expansion
- Moment Calculation from Quantile Information
 - A novel method for calculation of standard moments via quantile information by analytic Cumulative Distribution Function (CDF) integration with the piece-wise normal assumption

7.4. FUTURE WORK

- Higher Order Error Function Integration
 - An analysis of the integrals of higher orders of the error function, proposing a novel closed form solution to those integrals in terms of a newly defined function $S_t(x) = \int e^{-2x^2} \operatorname{erf}^t(x) dx$

7.4 Future Work

The novel method of In-order Multivariate Sampling (IMS) introduced in Chapter 3 is defined in terms of the multivariate normal distribution as it is this distribution which is important for circuit simulation applications. The IMS method could also be used in the context of other multivariate distributions and this extension can be the subject of future work.

The statistical characterization flow of the quadratic form defined in Chapter 4 uses the Cornish-Fisher expansion as an approximation to the distribution quantiles given the cumulants. Alternatively, exact quantiles can be computed by the inversion theorem following the work of Imhof[24] and Davies[9]. In future work, a comparison of these quantiles with the Cornish-Fisher approximation can be considered.

Integration of the higher order error functions is considered in Appendix A. The integration is shown to have a general closed form in terms of a newly defined function $S_t(x)$. The numerical approximation of $S_t(x)$ can be performed in the course of future work.

In summary, possible areas of future work include the following:

- Extension of In-order multivariate sampling to distributions other than the

multivariate standard normal

- Computation of the quadratic form quantiles by the inversion theorem with comparison against the approximation of the Cornish-Fisher expansion
- Numerical approximation to $S_t(x)$ of Appendix A

Bibliography

- [1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, 1964.
- [2] J. Aitchison and J. A. Brown. The lognormal distribution with special reference to its uses in economics. *University of Cambridge Department of Applied Economics Monograph 5*, page 176, 1957.
- [3] A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, pages 171–178, 1985.
- [4] J. L. Bentley and J. B. Saxe. Generating sorted lists of random numbers. *Transactions on Mathematical Software*, 6(3):359–364, Sept. 1979.
- [5] D. C. Brock and G. E. Moore. *Understanding Moore’s law: four decades of innovation*. Chemical Heritage Foundation, 2006.
- [6] P.-L. Chen, C.-T. Tsai, Y.-N. Chen, K.-C. Chou, C.-L. Li, C.-H. Tsai, K.-W. Wu, Y.-C. Chou, C.-Y. Li, W.-S. Lin, S.-H. Yu, R.-B. Chiu, C.-Y. Lin, C.-C. Wang, P.-W. Wang, W.-L. Su, C.-H. Wu, T.-T. Kuo, T. G. McKenzie, Y.-H. Chang, C.-S. Ferng, C.-M. Ni, H.-T. Lin, C.-J. Lin, and S.-D. Lin. A linear ensemble of individual and blended models for music rating prediction. In *Proceedings of KDD Cup 2011*, pages 21–60, 2012.

- [7] C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934.
- [8] E. A. Cornish and R. A. Fisher. Moments and cumulants in the specification of distributions. *Revue de l’Institut International de Statistique*, pages 307–320, 1938.
- [9] R. B. Davies. Numerical inversion of a characteristic function. *Biometrika*, 60(2):415–417, 1973.
- [10] N. I. Fisher. *Statistical Analysis of Circular Data*. Cambridge University Press, 1995.
- [11] J. H. Friedman. Random event generation with preferred frequency distributions. *Journal of Computational Physics*, (7):201, 1970.
- [12] F. Galton. The most suitable proportion between the value of first and second prizes. *Biometrika*, pages 385–399, 1902.
- [13] J. W. L. Glaisher. On a class of definite integrals. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 42(280):294–302, 1871.
- [14] G. H. Golub and C. F. Van Loan. *Matrix Computations*, volume 3. JHU press, 2012.
- [15] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, 2014.
- [16] M. Hane, T. Ikezawa, and T. Ezaki. Atomistic 3d process/device simulation considering gate line-edge roughness and poly-si random crystal orientation effects [mosfets]. In *IEEE International Electron Devices Meeting 2003*, pages 9–5. IEEE, 2003.
- [17] R. Harman and V. Lacko. On decompositional algorithms for uniform sampling from n-spheres and n-balls. *Journal of Multivariate Analysis*, 101(10):2297–2304, 2010.
- [18] F. Hausdorff. Summationsmethoden und momentfolgen. i. *Mathematische Zeitschrift*, 9(1-2):74–109, 1921.
- [19] A. Hazen. The storage to be provided in impounding reservoirs for municipal water supply. *Transactions of the American Society of Civil Engineers*, 77:1539–1669, 1914.

BIBLIOGRAPHY

- [20] G. A. Holton. Value-at-risk: Theory and Practice. 2nd edition. <https://www.value-at-risk.net/>, 2014.
- [21] J. R. Hosking. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(1):105–124, 1990.
- [22] J. R. M. Hosking, J. R. Wallis, and E. F. Wood. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261, 1985.
- [23] R. J. Hyndman and Y. Fan. Sample quantiles in statistical packages. *The American Statistician*, 50(4):361–365, 1996.
- [24] J.-P. Imhof. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3/4):419–426, 1961.
- [25] ISO26262. Road Vehicles – Functional Safety Standard, International Standards Organization, Geneva, Switzerland. <https://www.iso.org/standard/68383.html>, 2018.
- [26] N. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*. Number v. 2 in Wiley series in Probability and Mathematical Statistics: Applied Probability and Statistics. Wiley & Sons, 1995.
- [27] G. Kitagawa. Monte Carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996.
- [28] D. E. Knuth. *The Art of Computer Programming, Volume 3: (2nd Edition) Sorting and Searching*. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 1998.
- [29] C.-Y. Li, W.-L. Su, T. G. McKenzie, F.-C. Hsu, S.-D. Lin, J. Y.-j. Hsu, and P. B. Gib-

- bons. Recommending missing sensor values. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 381–390. IEEE, 2015.
- [30] E. Limpert, W. A. Stahel, and M. Abbt. Log-normal distributions across the sciences: keys and clues: on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: that is the question. *BioScience*, 51(5):341–352, 2001.
- [31] A. M. Mathai and S. B. Provost. *Quadratic Forms in Random Variables: Theory and Applications*. Dekker, 1992.
- [32] T. G. McKenzie, C.-S. Ferng, Y.-N. Chen, C.-L. Li, C.-H. Tsai, K.-W. Wu, Y.-H. Chang, C.-Y. Li, W.-S. Lin, S.-H. Yu, et al. Novel models and ensemble techniques to discriminate favorite items from ones for personalized music recommendation. In *Proceedings of the 2011 International Conference on KDD Cup 2011-Volume 18*, pages 101–135. JMLR.org, 2011.
- [33] S. Mittal. A survey of architectural techniques for managing process variation. *ACM Computing Surveys (CSUR)*, 48(4):54, 2016.
- [34] E. W. Montroll and M. F. Shlesinger. On $1/f$ noise and other distributions with long tails. *Proceedings of the National Academy of Sciences*, 79(10):3380–3383, 1982.
- [35] G. E. Moore et al. Cramming more components onto integrated circuits. *Electronics*, 38(8), 1965.
- [36] M. E. Muller. A note on a method for generating points uniformly on n -dimensional spheres. *Communications of the ACM*, 2(4):19–20, 1959.
- [37] D. B. Owen. Tables for computing bivariate normal probabilities. *The Annals of Mathematical Statistics*, 27(4):1075–1090, 1956.
- [38] K. Pearson. On the criterion that a given system of deviations from the probable

BIBLIOGRAPHY

- in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, 1900.
- [39] K. Pearson. Note on Francis Galton's problem. *Biometrika*, 1(4):390–99, 1902.
 - [40] M. J. Pelgrom and A. C. Duinmaijer. Matching properties of mos transistors. In *ESSCIRC'88: Fourteenth European Solid-State Circuits Conference*, pages 327–330. IEEE, 1988.
 - [41] R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo method*, volume 10. John Wiley & Sons, 2016.
 - [42] J. A. Shohat and J. D. Tamarkin. *The Problem of Moments*. Number 1. American Mathematical Soc., 1943.
 - [43] A. Singhee and R. A. Rutenbar. Statistical blockade: a novel method for very fast monte carlo simulation of rare circuit events, and its application. In *2007 Design, Automation & Test in Europe Conference & Exhibition*, pages 1–6. IEEE, 2007.
 - [44] M. Thulin et al. The cost of using exact confidence intervals for a binomial proportion. *Electronic Journal of Statistics*, 8(1):817–840, 2014.
 - [45] R. M. Vogel and N. M. Fennessey. L moment diagrams should replace product moment diagrams. *Water Resources Research*, 29(6):1745–1752, 1993.
 - [46] H.-F. Yu, H.-Y. Lo, H.-P. Hsieh, J.-K. Lou, T. G. McKenzie, J.-W. Chou, P.-H. Chung, C.-H. Ho, C.-F. Chang, Y.-H. Wei, J.-Y. Weng, E.-S. Yan, C.-W. Chang, T.-T. Kuo, Y.-C. Lo, P. T. Chang, C. Po, C.-Y. Want, Y.-H. Huang, C.-W. Hung, Y.-X. Ruan, Y.-S. Lin, S.-d. Lin, H.-T. Lin, and C.-J. Lin. Feature Engineering and Classifier Ensemble for KDD Cup 2010. *Journal of Machine Learning Research*, 1(16), 2010.

Appendices

Appendix A

Integral of High Order Error Function

Integrals of the error function are well known up to third order, with higher order solutions not available in closed form. In this section, a novel and general representation of the integrals of the high order error function is provided.

The integrals up to third order are given by[1] :

$$\int \operatorname{erf}(x) dx = x \operatorname{erf}(x) + \frac{1}{\sqrt{\pi}} e^{-x^2} \quad (\text{A.1})$$

$$\int \operatorname{erf}^2(x) dx = x \operatorname{erf}^2(x) + \frac{2}{\sqrt{\pi}} e^{-x^2} \operatorname{erf}(x) - \sqrt{\frac{2}{\pi}} \operatorname{erf}(\sqrt{2}x) \quad (\text{A.2})$$

$$\begin{aligned} \int \operatorname{erf}^3(x) dx &= x \operatorname{erf}^3(x) + \frac{3}{\sqrt{\pi}} e^{-x^2} \operatorname{erf}^2(x) - 12 \sqrt{\frac{2}{\pi}} T\left(2x, \frac{1}{\sqrt{2}}\right) \\ &\quad + 3 \sqrt{\frac{2}{\pi}} \end{aligned} \quad (\text{A.3})$$

where T is Owen's T function[37], named after statistician Donald Bruce Owen, defined by

$$T(h, a) = \frac{1}{2\pi} \int_0^a \frac{e^{-\frac{1}{2}h^2(1+x^2)}}{1+x^2} dx, (-\infty < h, a < +\infty). \quad (\text{A.4})$$

To aid in numerical solution, an alternative formulation of $T(h, a)$ is often used in terms of an integral over an area of the standardized bivariate normal distribution with zero correlation defined by

$$T(h, a) = \frac{-1}{2\pi} \int_0^h \int_0^{ax} e^{-\frac{1}{2}(x^2+y^2)} dy dx + \frac{\arctan(a)}{2\pi} \quad (\text{A.5})$$

While integrals of the error function of fourth order and beyond do not have a standard closed form solution, a clear pattern emerges when the error function integrals are instead redefined in the following way.

$$\int \operatorname{erf}(x) dx = x \operatorname{erf}(x) + \frac{1}{\sqrt{\pi}} e^{-x^2} \quad (\text{A.6})$$

$$\int \operatorname{erf}^2(x) dx = x \operatorname{erf}^2(x) + \frac{2}{\sqrt{\pi}} e^{-x^2} \operatorname{erf}(x) - \frac{4}{\pi} \int e^{-2x^2} dx \quad (\text{A.7})$$

$$\int \operatorname{erf}^3(x) dx = x \operatorname{erf}^3(x) + \frac{3}{\sqrt{\pi}} e^{-x^2} \operatorname{erf}^2(x) - \frac{12}{\pi} \int e^{-2x^2} \operatorname{erf}(x) dx \quad (\text{A.8})$$

$$\int \operatorname{erf}^4(x) dx = x \operatorname{erf}^4(x) + \frac{4}{\sqrt{\pi}} e^{-x^2} \operatorname{erf}^3(x) - \frac{24}{\pi} \int e^{-2x^2} \operatorname{erf}^2(x) dx \quad (\text{A.9})$$

$$\begin{aligned} \int \operatorname{erf}^k(x) dx &= x \operatorname{erf}^k(x) + \frac{k}{\sqrt{\pi}} e^{-x^2} \operatorname{erf}^{k-1}(x) \\ &\quad - \frac{2k(k-1)}{\pi} \int e^{-2x^2} \operatorname{erf}^{k-2}(x) dx, k \in \mathbb{N} \end{aligned} \quad (\text{A.10})$$

Define a function $S_t(x)$ as follows:

$$S_t(x) = \int e^{-2x^2} \operatorname{erf}^t(x) dx \quad (\text{A.11})$$

Now integrals of the error function of any positive integer order can be universally defined in closed form in terms of $S_t(x)$ by

$$\int \operatorname{erf}^k(x) dx = x \operatorname{erf}^k(x) + \frac{k}{\sqrt{\pi}} e^{-x^2} \operatorname{erf}^{k-1}(x) - \frac{2k(k-1)}{\pi} S_{k-2}(x), \quad k \in \mathbb{N} \quad (\text{A.12})$$

While more general than Owen's T function, the numerical approximation of $S_t(x)$ for $t > 1$ is left an exercise and is beyond the scope of this thesis.

Appendix B

Operational Amplifier Netlist Listing

```
** Copyright (C) 2005 Synopsys, Inc. All rights reserved
** Test circuit for Monte Carlo analysis
** Classic 7-transistor opamp
** global and local variations
**
.GLOBAL gnda vdda
.PARAM vdd=2.5 vin=vdd/2 k=2

.subckt opamp gnda inn inp out vdda nmosbulk pmosbulk
mn1 net031 inn net044 nmosbulk snps20N L='k*0.5u' W='k*3.5u' M=4
mn2 net18 inp net044 nmosbulk snps20N L='k*0.5u' W='k*3.5u' M=4
mp3 net031 net031 vdda pmosbulk snps20P L='k*0.5u' W='k*4.5u' M=4
mp4 net18 net031 vdda pmosbulk snps20P L='k*0.5u' W='k*4.5u' M=4
mp5 out net18 vdda pmosbulk snps20P L=400e-9 W=10e-6 M=3
mn8 net0148 net0148 gnda nmosbulk snps20N L=3e-6 W=18e-6 M=5
mn7 net044 net0148 gnda nmosbulk snps20N L=3e-6 W=18e-6 M=2
mn6 out net0148 gnda nmosbulk snps20N L=3e-6 W=18e-6 M=12
ccomp out net058 900e-15
rcomp net18 net058 7e3
r0 net0148 vdda 1e6
```

```

.ends opamp

c0 in_neg 0 1m
c1 out 0 5e-12
r1 0 out 1e6
r0 in_neg out 10e6
xi82 gnda in_neg in_pos out vdda gnda vdda opamp
v2 0 gnda DC=0
v1 in_pos 0 DC=vin AC 1
v0 vdda 0 DC=vdd

* model for tsmc 0.2u technology
* from NCSU CDK tsmc20P and tsmc20N
.MODEL snps20p PMOS (
+VERSION = 3.1
+XJ      = 1E-7
+K1      = 0.5955538
+K3B     = 10.7990376
+DVTOW   = 0
+DVTO    = 0.3622323
+U0      = 119.2085214
+UC      = -1E-10
+AGS     = 0.4199318
+KETA    = 0.0126497
+RDSW   = 201.0196067
+WR      = 1
+XL      = -2E-8
+DWB     = 5.977646E-9
+CIT     = 0
+CDSCB   = 0
+DSUB    = 0.940025
+PDIBLC2 = 0.0195957
+PSCBE1  = 2.224265E9
+DELTA   = 0.01
+PRT     = 0
+KT1L    = 0
+UB1     = -7.61E-18
+WL      = 0
+WWN    = 1
+LLN    = 1
+LWL    = 0
TNOM    = 27
NCH     = 4.1589E17
K2       = 0.0265154
W0       = 1E-6
DVT1W   = 0
DVT1    = 0.2560341
UA       = 1.656038E-9
VSAT    = 1.840486E5
B0       = 9.960505E-7
A1       = 0.4537105
PRWG    = 0.5
WINT    = 0
XW       = -1E-8
VOFF    = -0.1035186
CDSC    = 2.4E-4
ETAO    = 0.0079869
PCLM    = 1.9711817
PDIBLCB = -1E-3
PSCBE2  = 6.4242E-10
RSH     = 7.6
UTE     = -1.5
KT2     = 0.022
UC1     = -5.6E-11
WLN     = 1
WWL     = 0
LW      = 0
CAPMOD  = 2
TOX     = 4.1E-9
VTH0    = -0.4215645
K3      = 0
NLX     = 7.393151E-8
DVT2W   = 0
DVT2    = 0.1
UB      = 1E-21
AO      = 1.7659056
B1      = 3.37199E-6
A2      = 0.3
PRWB    = -0.5
LINT    = 2.42201E-8
DWG     = -2.790988E-8
NFACTOR = 1.8044589
CDSCD   = 0
ETAB    = -0.115204
PDIBLC1 = 0
DROUT   = 5.830459E-4
PVAG    = 10.2269693
MOBMOD  = 1
KT1     = -0.11
UA1     = 4.31E-9
AT      = 3.3E4
WW      = 0
LL      = 0
LWN     = 1
XPART   = 0.5
LEVEL   = 49

```

APPENDIX B. OPERATIONAL AMPLIFIER NETLIST LISTING

```

+CGDO      = 6.52E-10      CGSO      = 6.52E-10      CGBO      = 1E-12
+CJ        = 1.156829E-3    PB         = 0.8604313    MJ         = 0.4161985
+CJSW      = 1.800318E-10    PBSW      = 0.6161205    MJSW      = 0.2735145
+CJSWG     = 4.22E-10       PBSWG     = 0.6161205    MJSWG     = 0.2735145
+CF        = 0              PVTH0     = 1.121114E-3   PRDSW     = 12.2644118
+PK2       = 1.671328E-3    WKETA     = 2.478808E-3   LKETA     = -2.85111E-3
+PU0       = -1.8187729    PUA        = -7.23748E-11  PUB        = 1E-21
+PVSAT     = -50            PETAO     = 1E-4          PKETA     = 2.650105E-3   )
                                         )

.MODEL snps20n NMOS (
+VERSION = 3.1
+XJ      = 1E-7
+K1      = 0.5935169
+K3B     = 3.1905105
+DVTOW   = 0
+DVTO    = 1.7203781
+UO      = 269.0634518
+UC      = 2.224818E-11
+AGS     = 0.4169677
+KETA    = -7.704208E-3
+RDSW   = 105
+WR      = 1
+XL      = -2E-8
+DWB     = 1.217904E-8
+CIT     = 0
+CDSCB   = 0
+DSUB   = 0.0110906
+PDIBLC2 = 3.755701E-3
+PSCBE1  = 5.995957E10
+DELTA   = 0.01
+PRT     = 0
+KT1L    = 0
+UB1     = -7.61E-18
+WL      = 0
+WWN    = 1
+LLN    = 1
+LWL     = 0
+CGDO   = 7.45E-10
+CJ      = 9.725136E-4
+CJSW   = 2.269386E-10
+CJSWG  = 3.3E-10

                                         LEVEL   = 49
TNOM      = 27
NCH       = 2.3549E17
K2        = 2.38533E-3
W0        = 1E-7
DVT1W    = 0
DVT1     = 0.4308344
UA        = -1.188565E-9
VSAT     = 9.67502E4
B0        = -1.063955E-8
A1        = 7.99632E-4
PRWG     = 0.5
WINT     = 2.025957E-9
XW        = -1E-8
VOFF     = -0.0901723
CDSC     = 2.4E-4
ETA0     = 1.448044E-3
PCLM     = 1.0622551
PDIBLC2 = -0.1
PSCBE2  = 5.686023E-8
RSH      = 6.7
UTE      = -1.5
KT2      = 0.022
UC1      = -5.6E-11
WLN      = 1
WWL      = 0
LW       = 0
CAPMOD   = 2
CGSO     = 7.45E-10
PB        = 0.7292509
PBSW     = 0.6351005
PBSWG   = 0.6351005

                                         TOX     = 4.1E-9
VTH0     = 0.3796589
K3       = 1E-3
NLX      = 1.786849E-7
DVT2W   = 0
DVT2     = 0.0467521
UB       = 1.930877E-18
A0       = 2
B1       = -1E-7
A2       = 0.999873
PRWB   = -0.2
LINT     = 1.028309E-8
DWG     = -6.4982E-10
NFACTOR = 2.3820479
CDSCD   = 0
ETAB     = -2.754731E-4
PDIBLC1 = 0.3172281
DROUT   = 0.783102
PVAG     = 0.3568363
MOBMOD   = 1
KT1      = -0.11
UA1      = 4.31E-9
AT       = 3.3E4
WW      = 0
LL      = 0
LWN     = 1
XPART   = 0.5
CGBO     = 1E-12
MJ       = 0.3610145
MJSW     = 0.1
MJSWG   = 0.1
)
                                         )

```

```

+CF      = 0          PVTH0   = -2.139932E-3    PRDSW   = -1.2311975
+PK2     = 1.860342E-3 WKETA   = 1.76355E-3     LKETA   = -5.667186E-3
+PU0     = -0.2295277 PUA     = -2.87112E-11   PUB     = 0
+PVSAT   = 1.427606E3 PETAO   = 1E-4           PKETA   = -1.196986E-3 )
```

.variation

.global_variation

 nmos snps20N vth0=0.07 u0=10 %

 pmos snps20P vth0=0.08 u0=8 %

.end_global_variation

.local_variation

 nmos snps20N vth0='1.234e-9/sqrt(get_E(W)*get_E(L)*get_E(M))'

+ u0='2.345e-6/sqrt(get_E(W)*get_E(L)*get_E(M))' %

 pmos snps20P vth0='1.234e-9/sqrt(get_E(W)*get_E(L)*get_E(M))'

+ u0='2.345e-6/sqrt(get_E(W)*get_E(L)*get_E(M))' %

.element_variation

 R r=10 %

.end_element_variation

.end_local_variation

.end_variation

.dc k start=1 stop=4 step=1 monte=1000

.meas DC systoffset1 find V(in_pos,in_neg) at=1

.meas DC systoffset2 find V(in_pos,in_neg) at=2

.meas DC systoffset3 find V(in_pos,in_neg) at=3

.meas DC systoffset4 find V(in_pos,in_neg) at=4

.meas DC leakpwr PARAM='-I(v0)*V(vdda)',

.meas DC logleakpwr PARAM='log(-I(v0)*V(vdda))'

.TEMP 25

.option nomod dcon=1

.option post \$sampling_method=ofat

.END