

**UCC Library and UCC researchers have made this item openly available.
Please [let us know](#) how this has helped you. Thanks!**

Title	Option pricing and CVA calculations using the Monte Carlo-Tree (MC-Tree) method
Author(s)	Trinh, Yen Thuan
Publication date	2022-07-18
Original citation	Trinh, Y. T. 2022. Option pricing and CVA calculations using the Monte Carlo-Tree (MC-Tree) method. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2022, Yen Thuan Trinh. https://creativecommons.org/licenses/by-nc-nd/4.0/
Item downloaded from	http://hdl.handle.net/10468/13565

Downloaded on 2022-12-08T09:06:09Z



Option Pricing and CVA Calculations Using The Monte Carlo-Tree (MC-Tree) Method

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Mathematics
National University of Ireland, Cork
July, 2022

Thesis by Yen Thuan Trinh
Under the Supervision of
Prof. Bernard Hanzon,
Chair of Mathematics

School of Mathematical Sciences
University College Cork
Western Rd, Cork, Ireland.

Contents

Declaration of Work	ix
Thesis Structure	xi
Publications and Working Papers	xiii
Abstract	xv
Acknowledgements	xvii
List of Figures	xix
List of Tables	xxiii
Glossary	xxvii
1 Introduction	1
1.1 Motivation	1
1.1.1 Benefits and Drawbacks of Monte Carlo Simulation and Tree Methods	1
1.1.2 An MC-Tree Method	2

1.2	Literature Review	4
2	Pricing Single-Asset Options Using MC-Tree Method	9
2.1	Introduction	9
2.2	Parameters	10
2.3	Preliminaries to the Derivation of the General Compound Density Formula	13
2.4	General Compound Density Formula	16
2.5	Mixing Density	18
2.6	Monte Carlo Drawing	23
2.7	A Particular Compound Density	25
2.8	Option Pricing	32
2.8.1	Usage of a Bias-Correction	32
2.8.2	Usage of a Distribution Correction Factor	33
2.9	Convex Combination and Convex Optimization	34
2.9.1	Convex Combination	34
2.9.2	Convex Optimization	36
2.10	Numerical Results	37
2.10.1	Numerical Results of Entropy, KL Divergence, and L_1 Distance	37
2.10.2	Distribution Correction Factor	40
2.10.3	Convex Combination and Convex Optimization	42
2.10.4	Pricing European Options	43
2.10.4.1	Comparison to Plain MC Method	45
2.10.4.2	Comparison to Binomial Models	46
2.10.5	Pricing American Put Option	47

2.10.5.1	Comparison to LSM Method	52
2.10.5.2	Comparison to Binomial Models	54
2.10.5.3	The optimal N and M of MC-Tree	56
3	Pricing Multi-Asset Options Using An MC-Tree Method	57
3.1	Introduction	57
3.2	Mixing Density and MC Drawings	58
3.3	The Radius	60
3.4	RI Entropy	62
3.4.1	RI Entropy in the Two Assets Case	63
3.4.2	RI Entropy in the Two Assets Case(Alternative Approach)	65
3.4.3	RI Entropy in the Three Assets Case	67
3.5	Compound Entropy Maximization	68
3.6	Analytic Compound Density	70
3.7	Generating the Histogram	83
3.8	Discrete Entropy	86
3.9	The Characteristic Function and cdf	88
3.10	Entropy Optimization	89
3.10.1	Convex Combination	89
3.10.2	Unilateral Optimization	91
3.11	Option Pricing Using MC-Tree with the Bias-Correction	95
3.12	Numerical Results in the Two Assets Case	96
3.12.1	Histogram and Discrete Entropy	96
3.12.2	The Characteristic Function and the cdf	97

3.12.3	Convex Combination and Unilateral Optimization of Mixing Densities with degree from 1 to 30	107
3.12.4	Convex Combination and Unilateral Optimization of Symmet- ric Monomials	107
3.12.5	Numerical Results of European Options on the Maximum and Minimum of Two Assets	110
3.12.5.1	Comparison To the Recombining Multinomial Tree Based on Pascal's Simplex	110
3.12.5.2	Comparison To Plain Monte Carlo (MC)	112
3.12.6	Numerical Results for American Options on the Maximum and Minimum of Two Assets	112
4	Credit Valuation Adjustment (CVA)	115
4.1	Unilateral CVA Formulation	115
4.2	Default Probability	118
4.3	Expected Exposure	119
4.3.1	Tree Approach for CVA	120
4.3.2	MC-Tree Approach for CVA	121
4.4	Remarks	121
4.5	Numerical Results of CVA Calculations	121
5	Entropy and Kullback-Leibler formulas for rational densities	123
5.1	Introduction	123
5.2	An Application of the Residue Theorem	125

5.3	The Entropy and the Kullback-Leiber (KL) Divergence of Compound Densities	134
5.4	Entropy and the KL Divergence for Cauchy Distributions	151
5.4.1	Entropy	151
5.4.2	The KL Divergence	152
5.5	Entropy and the KL Divergence Between the Rational Student Densities	153
5.5.1	Introduction	153
5.5.2	Entropy	154
5.5.3	The KL Divergence of Rational Student densities	156
5.5.4	Some Specific Cases	160
5.6	Entropy and the KL Divergence of Rational Densities	163
5.6.1	Entropy and the KL Divergence of a Linear Convex Combina- tion of two Cauchy Densities	165
5.6.2	Entropy and the KL Divergence of a Linear Convex Combina- tion of two Rational Student Densities	173
5.7	On Approximate Entropy Calculation for Finite Gaussian Mixtures .	183
6	Conclusion	185
6.1	Concluding Remarks	185
6.2	Further Research	186
	Appendices	189
A	Choices of Mixing Density	189
A.1	Choice of Mixing Density in One Asset Case	190
A.2	Choice of Mixing Density in Two Asset Case	190

References

193

Declaration of Work

I, Yen Thuan Trinh, hereby declare that this Ph.D. thesis entitled *Option Pricing and CVA Calculations Using The Monte Carlo-Tree (MC-Tree) Method* has been written entirely by me under the supervision of Prof. Bernard Hanzon, Chair of Mathematics, University College Cork, National University of Ireland.

This thesis has not previously been submitted for any other award at this university or at any other institution.

Yen Thuan Trinh _____, Monday, July 18th, 2022

Thesis Structure

Chapter one includes the motivation and a summary of related works or methods from the literature.

Chapter two describes the theory and numerical results for pricing single-asset options using the MC-Tree method.

Chapter three describes the MC-Tree in the multi-assets case. More specifically, the theory and numerical results of MC-Tree in a two assets case are presented in this chapter.

Chapter four presents the formulation of unilateral CVA and its numerical results. Numerical results consist of techniques of optimization, option pricing, and algorithms.

Chapter five presents a new analytic formula for the Entropy and KL Divergence of rational densities.

Chapter six concludes a brief synopsis of the thesis and mentions some ideas for future research.

Publications and Working Papers

Posters, Talks, and Abstracts

1. The 5th Vienna Workshop for High-Dimensional Time Series in Macroeconomics and Finance, 9-10 June 2022, Vienna.

Paper Presented: Entropy and Kullback-Leibler formulas for rational densities.

Presented by Prof. Bernard Hanzon.

2. Actuarial and Financial Mathematics Conference, 10-11 February 2022, Brussels.

Abstract: <https://afmathconf.ugent.be/Abstracts/TrinhYenThuan20210929154107.pdf>

3. Online Workshop: Recent Advances in Financial Mathematics, November 2021.

Paper Presented (Invited Speaker): Option Pricing and CVA Calculations Using The MC-Tree Method with the Distribution Correction Factor.

Abstract: https://sites.google.com/ueh.edu.vn/icams2022/workshop-series_1/workshop-01-recent-advances-in-financial-mathematics

4. Irish Mathematical Society Conference (IMS), Sep 2021, Munster Technological University (MTU).

Poster Presented: Option Pricing and CVA Calculations Using The MC-Tree Method with the Distribution Correction Factor.

5. Actuarial and Financial Mathematics Conference, Brussels, Belgium, Feb 2020.

Poster Presented: Pricing American Option Using Monte Carlo-Tree (MC-Tree) method.

Abstract: <https://afmathconf.ugent.be/Abstracts/TrinhYenThuan20190926151528.pdf>

ArXiv Preprint

- Trinh, Y. T., & Hanzon, B. (2022). Option Pricing and CVA Calculations using the Monte Carlo-Tree (MC-Tree) Method. arXiv preprint arXiv:2202.00785.

Working Papers/Under Review

1. Trinh, Y. T., & Hanzon, B. (2022). Option Pricing and CVA Calculations using the Monte Carlo-Tree (MC-Tree) Method.
2. Hanzon, B & Trinh, Y. T.. Entropy and Kullback-Leibler formulas for rational densities.
3. Trinh, Y. T., & Hanzon, B. Pricing Multi-Assets Option Using The MC-Tree Method.

Abstract

The thesis introduces a new method, the MC-Tree method, for pricing certain financial derivatives, especially options with high accuracy and efficiency. Our solution is to combine Monte Carlo (MC) method and Tree method by doing a mixing distribution on the tree, and the output is the compound distribution on the tree. The compound distribution in the tree output (after a logarithmic transformation of the asset prices) is not the ideal Gaussian distribution but has entropy values close to the maximum possible Gaussian entropy. We can get closer using entropy maximization. We introduce two correction techniques: distribution correction and bias correction to improve the accuracy and completeness of the model. The thesis presents an algorithm and numerical results for calculations of CVA on an American put option using the MC-Tree method.

The MC-Tree method with the distribution correction technique significantly improves accuracy, resulting in practically exact solutions, compared to analytical solutions, at the tree depth $N = 50$ or 100 and MC-drawings $M = 10^5$. The bias-correction technique makes the resulting tree model complete in the sense of financial mathematics and obtains the risk-neutral probability. Besides, we have obtained new formulae for the calculations of the entropy and the Kullback-Leibler divergence for rational densities and approximate entropy of finite Gaussian mixture.

Acknowledgements

- I am immensely grateful to my supervisor: Prof. Bernard Hanzon, for the opportunity of working with him. He gradually and elegantly guided my research work and teaching. His valuable continuous feedback made me look at things more wisely and think more critically.
- I would acknowledge constant support from staff at the School of Mathematical Sciences throughout my whole PhD period. I would also wish to thank UCC for awarding me the full scholarship throughout my PhD. I would thank the committee of the Actuarial and Financial Mathematics Conference, Brussels, Belgium offering me an FWO-WOG travel grant to expand my knowledge through the conference.
- I would like to acknowledge Marta Ferrario and Luca Pettinari for their earlier contributions to the theory of MC-Tree in chapter two.
- I am very thankful to internal examiner and external examiner for their recommendations in improving this thesis.

List of Figures

2.1	Evolution of S_t in a two-step additive binomial tree. Here $p_2 = \frac{\tau^2}{1+\tau^2}$, $p_1 = \frac{1}{1+\tau^2}$.	13
2.2	Entropy behavior when m is an odd number between 1 and 50. Top left: $N = 10$. Top right: $N = 15$. Bottom left: $N = 20$. Bottom right: $N = 50$.	38
2.3	Standard Gaussian density versus compound density at $N = 20$, $m = 9$.	39
2.4	Entropy values versus the tree depth N from 1 to 50 when $m = 9$ is fixed.	39
2.5	L_1 distance versus the tree depth N from 10 to 120 when $m = 9$ is fixed.	40
2.6	The distribution correction factor.	41
2.7	The plot of optimal coefficients when $N = 20$. Top: optimal coefficients α_m with m odd numbers from 9 to 35. Bottom: optimal coefficients α_m with odd numbers of m from 11 to 35.	42
2.8	Verification of put-call parity. The y-axis represents $C(S(t), t) - P(S(t), t) - S(t) + KB(t, T)$, where $C(S(t), t)$, $P(S(t), t)$ are the call and put option prices, and $B(t, T) = e^{-r(T-t)}$.	44
2.9	Option prices versus the tree depth N .	50

2.10	Option prices versus the tree depth N	51
2.11	European call prices versus number of MC drawings M and the tree depth N	52
2.12	Option prices versus the tree depth N	55
2.13	Error versus the depth of tree.	56
3.1	Analytic compound density at $N = 2$, $m = 9$	83
3.2	Approximated compound density at $N = 2$, $m = 9$	83
3.3	Generated histogram for the standardized radius using mixing density with $m = 9$. Red points are the mid points of histogram bins.	98
3.4	Generated histogram for the standardized square radius using mixing density with $m = 9$ by the alternative approach. Red points are the mid points of histogram bins.	99
3.5	Entropy values versus the tree depth. The dotted blue line represents the entropy value of standard bivariate Gaussian density at 2.83788.	100
3.6	Entropy values versus bin-width. The dotted blue line represents the entropy value of standard bivariate Gaussian density at 2.83788.	101
3.7	The characteristic function at $N = 10$ and $m = 9$. Red curve: The ideal characteristic function. Blue curve: Approximated characteristic function of the standardized squared radius.	102
3.8	The cdf of the standardized squared radius versus the ideal cdf at $N = 30$ and $m = 30$. Blue curve: Approximated cdf of standardized squared radius. Red curve: Ideal cdf.	103

3.9	Compound density versus standard exponential density of the standardized squared radius at $N = 30$ and $m = 30$. Blue curve: Approximated compound density. Red curve: Ideal density.	104
3.10	The cdf of standardized radius versus the ideal cdf at $N = 30$ and $m = 30$	105
3.11	Compound density versus the ideal density of the standardized radius at $N = 30$ and $m = 30$	106
3.12	The plot of optimal coefficients in the convex combination of mixing densities when $N = 10$	108
3.13	The plot of optimal coefficients in the convex combination of symmetrized monomials when $N = 10$	109

List of Tables

2.1	KL divergence versus N and m	37
2.2	Entropy values versus N and m	37
2.3	Accuracy comparison between MC-Tree and MC method in pricing European call option. Bias: MC-Tree with the usage of bias-correction. Corr: MC-Tree with the usage of distribution correction factor. SD is the estimate of the standard deviation of the random variable of which the MC estimates the mean.	45
2.4	Accuracy comparison between MC-Tree and MC method in pricing European put option. Bias: MC-Tree with the usage of bias-correction. Corr: MC-Tree with the usage of distribution correction factor. . . .	46
2.5	Error and computation time of European call option with similar running time at $N = 50$. MC-Tree Corr: MC-Tree with the distribution correction factor. SE denotes the standard error.	47
2.6	Accuracy comparison between MC-Tree and binomial models in pricing European call option. Bias: MC-Tree with the usage of bias-correction. Corr: MC-Tree with the usage of distribution correction factor. . . .	48

2.7	Accuracy comparison between MC-Tree and binomial models in pricing European put option. Bias: MC-Tree with the usage of bias-correction. Corr: MC-Tree with the usage of distribution correction factor.	49
2.8	Mean-squared error for various models.	52
2.9	Mean and standard deviation of American put option.	53
2.10	Mean and standard error (SE) of American put Option with similar running time.	53
2.11	Option Prices	54
2.12	Accuracy comparison among models.	54
3.1	Accuracy comparison between MC-Tree and Q&P for the European call options on the maximum of the assets.	111
3.2	Accuracy comparison between MC-Tree and Q&P for the European put options on the minimum of the assets.	111
3.3	Accuracy comparison between MC-Tree and plain MC for the European call options on the maximum of the assets.	112
3.4	Accuracy comparison between MC-Tree and plain MC for the European put options on the minimum of the assets.	112
3.5	Accuracy comparison between MC-Tree and Q&P for the American call options on the maximum of assets.	114
3.6	Accuracy comparison between MC-Tree and Q&P for the American put options on the minimum of assets.	114
4.1	CVA values.	122

5.1 Convergence of entropy of linear combination of student distributions
to entropy of finite Gaussian mixture when increasing the degree of
freedom. 184

Glossary

\mathbb{Q} denotes the risk-neutral measure.

τ denotes a parameter characterizing the tree of arbitrary tree depth.

δ denotes the time-step length.

$C(x)$ denotes the distribution correction factor.

\mathbb{O}_d denotes the set of dx d orthogonal matrices.

$R = \|X\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_d^2}$ denotes the radius of a point in \mathbb{R}^d .

$R_{k,m}$ denotes a drawing of the radius at the k^{th} node in the m^{th} tree.

$R_k = \|X_k\|$ denotes the radius of the vectors at the nodes k in the tree.

$P = \frac{R^2}{2N}$ denotes the standardized squared radius.

Counterparty credit risk (CCR) is the risk that a party (a counterparty) in an over-the-counter (OTC) derivatives contract is incapable of following its obligation due to its default, which causes a significant loss to the other party (bank).

Credit Valuation Adjustment (CVA) value shows a change to the market value of derivative instruments after taking into account the possibility of a counterparty's default.

Netting set is a group of transactions with a single counterparty.

Mark to Market (MtM) recovery value is the recovery value based on the current market value of investments.

Chapter 1

Introduction

1.1. Motivation

1.1.1. Benefits and Drawbacks of Monte Carlo Simulation and Tree Methods

Monte Carlo Simulation is popular in pricing single and multi-asset derivatives due to some reasons. Firstly, the error converges to zero at the rate $\frac{1}{\sqrt{M}}$, where M is the number of simulations, which does not depend on the dimension of the problem. A higher number of simulations is required to obtain better accuracy. This is the result of the central limit theorem. Secondly, Monte Carlo simulation is a simple implementation because correlations and path dependency can be easily modelled and handled. The simple workload is to simulate paths and corresponding cash flows, average the payoff and take the present value.

The main advantage of classical methods such as the Tree or Lattice method is its simplicity for implementation in option pricing. A higher tree depth is required to

get good precision. Tree methods have a drawback with high-dimensional pricing problems because the number of nodes grows exponentially with the dimension. The combination with Monte Carlo can handle this drawback of the tree methods. Besides, we use the recombining multinomial tree (binomial tree) based on Pascal's simplex in [40] (Pascal's triangle) to reduce the number of nodes of classical tree methods. As a result, the computational cost is reduced. We refer to [40] for details of the model: recombining multinomial trees based on Pascal's simplex on multi-assets options.

1.1.2. An MC-Tree Method

The research problem is to introduce a fast pricing method for certain financial derivatives, especially options with the requirement of high accuracy. MC-Tree method allows to generate parameters of each tree with a chosen probability distribution. MC-Tree method holds benefits of both the Monte Carlo and the tree method. Also, one advantage of the MC-tree approach is that it allows us to work with the confidence interval of the MC simulations. If we use only the tree method, we can not have the confidence interval and its benefit. The confidence interval depends on the number of simulations. The maximal depth of the tree used in MC-Tree method is expected to be limited due to computational constraints for the multi-asset case. Options form a large class of derivatives on the financial market. The majority of exchange-traded options on a single stock tend to be American style, while options on indexes are European. Unlike a European option, an American option gives the holder the right (but not the obligation) to exercise at any time before the expiration date. The American options have a cash flow that depends on the price path of the

underlying assets and the decisions of the holder. We look at options and financial derivatives for which the payout only depends on the share prices at the expiry date. Note that some other options can still be valued with tree methods, but we do not go into that in this thesis.

After the financial crisis of 2007-2008, there was a requirement for a significant change in financial modeling and risk management. The financial regulators require banks to hold an amount of capital to capture the credit risk in portfolios. The BIS (Bank of International Settlements) issued the Basel Accords-Basel I, Basel II, and Basel III (see [2], [3]) for the calculations of the required capital. The BIS published Basel III to work along with Basel II in response to the deficiencies in banking regulation in Basel I. Basel III presents a new measure, named credit valuation adjustment (CVA), which is the market value of counterparty credit risk (CCR) [3]. This is the motivation for developing the algorithm to calculate CVA on an American option using the MC-Tree method.

This research is significant because we introduce a new method to obtain faster and more accurate option pricing calculations. The method could be very useful for the financial industry, especially the derivative pricing and risk management industry. Besides, in this thesis, as a spin-off of our research into the MC-Tree method, we develop analytic formulae for the entropy and the Kullback-Leibler (KL) divergence of rational densities, which play an important role in mathematical statistics, information theory, mathematical systems theory, and related fields.

All numerical experiments were conducted on the machine I5-10210U, 8GB memory with programming language R i386 3.5.1. We used the R package: "far" for generating the matrix H in chapter three. In this thesis, we have relied on existing routines in Mathematica symbolic computation to find the analytic compound density.

1.2. Literature Review

The pioneering work of Bachelier (1900) is now seen as the forerunner of what would become a massive usage of mathematical models in finance since the last quarter of the 20th century. There is a demand for fast and accurate methods for pricing options as more exotic financial instruments are developed and traded on the market over the years. Black and Scholes [5] derived the Black-Scholes partial differential equation to price simple European options on a single asset. Cox et al. in [18] extended it to the generalized Black-Scholes equation to price derivatives on multiple assets. Subsequently, analytical solutions for quite a number of derivatives have been derived, using either the Black-Scholes partial differential equation or the discounted risk-neutral expectation method, but it is too difficult to solve the equation analytically for many other derivatives. Based on the Black-Scholes model for those cases for which no useful analytical solution is available many methods for approximate option valuation have been developed, but there is still room for improvement. Our contribution is to introduce a new computational approach to find the value of financial derivatives with arbitrary boundary conditions and generalization to the case of American options and other classes of options such as barrier options.

Boyle [6] introduced the Monte Carlo approach to option pricing, and it is still very popular because of its flexibility to approximate all kinds of option prices. Monte Carlo simulation has been extended in pricing American options with popular methods such as the Stochastic Mesh [14], the Least Square Monte Carlo [35], and the State-Space partition method [44], and so on. Longstaff and Schwartz [35] introduced a least square regression method (LSM) to approximate the prices of American options. Much research has been done subsequently to give the analysis and the

convergence properties of the LSM.

Lattice-type models for pricing options were implemented in many works ([7], [8]). Various choices for tree parameters lead to different existing binomial models. The first proposed formulation of the binomial tree in financial mathematics was the research by Cox, Ross, and Rubinstein [19], providing a simplified discrete approach to option pricing on one asset using the recombining binomial trees based on Pascal's triangle. The authors in this article proposed to take $u = 1/d$, where u, d represent upward and downward movements. The work is still very popular today. One benefit of this binomial tree is that it exhibits a unique risk-neutral probability. Sierag and Hanzon [40] extended this to multi-asset option pricing using recombining multinomial trees based on Pascal's simplex. A second popular model was the research by Jarrow and Rudd [29], also known as the equal-probability model. One drawback is that the model is no longer a risk-neutral model although it matches the risk-neutral continuous-time model in the limit for the time-step length going to zero. Authors in ([19], [29]) matched the first two moments of the multiplicative tree to the risk-neutral continuous model, leading to a system of two equations for three unknowns. Tian's approach [43] was to equate the third moment to handle the issue of this free variable. Leisen and Reimer [32] presented the most different choice of the model parameters u, d and p among the approaches mentioned above. They devoted attention to improving the convergence speed for European call options when approximating the Gaussian distribution. Leisen [31] presented a convergence analysis for binomial models in pricing American put options.

Although many research advances have been proposed so far to price American option contracts, the valuation of American options has not been fully solved, and it remains an interesting topic for researchers in financial mathematics. Pricing

American options requires complex and time-consuming methods to ensure high accuracy due to the absence of analytical solutions in a simple form.

CCR is the risk that a party (a counterparty) in an over-the-counter (OTC) derivatives contract is incapable of following its obligation due to its default, which causes a significant loss to the other party (bank). CVA calculations are complicated because modelling CVA consists of at least three components: the exposure, the default probability of the counterparty, and the loss given default (LGD) (see [22]). The default probability depends on the credit quality of the counterparty. If the exposure profile and the credit quality of the counterparty are positively or negatively related, so-called right-way risk (RWR) or wrong-way risk (WWR) occurs. Otherwise, one speaks about unilateral CVA.

Many studies worked on CVA to meet the requirement of the advanced risk management in banks or other financial institutions. There are existing approaches in the literature. Direct approach for CVA works well with derivatives in the portfolio for which one has analytical pricing formulas available. This method consists of the following three steps. Step one is to simulate paths of the risk factors such as the short rate for interest rate derivatives, stock prices for option on the stock, and so on at time grid $t_0 = t, t_1, \dots, t_N = T$. Then they value the exposure at each time step t_i and for each simulated path $k, 1 \leq k \leq M$ at step two. Step three is to calculate the expected positive exposure at each time step t_i . The direct approach does not work well with complex derivatives, which do not have an available analytical pricing formula. An alternative method, Least Squares Monte Carlo (LSMC), was introduced to solve this problem. LSMC approach for CVA has been investigated in a variety of papers. LSMC is the combination of Monte Carlo simulation with least-squares regression. In this thesis, we present a new approach: MC-Tree approach for CVA.

The article by Zhu (see [45]) focused on pricing counterparty risk and modelling credit exposure at the contract and counterparty levels. The article also gave an example of calculating unilateral CVA without WWR for a portfolio of vanilla interest swaps with a counterparty. This example used the classical MC method, which cannot work for products without a closed-form or analytic pricing. The default probability for the counterparty was implied from the current spreads market of its CDS, and a Matlab function was used to generate the cumulative probability of default at each simulation date. Authors in [35] used LSM to price path-dependent derivatives, Cesari et al. [16] recommended the combination with the Monte Carlo simulation for the calculation of expected exposures and CVA. Broadie and Glasserman [13] observed a bias in estimating the conditional expectation function if they use the same set of simulated sample paths. Hence, the first simulation of the underlying asset is added to estimate the regression coefficients before the second simulation in calculations of the continuation values. Authors in [33] calculated the unilateral CVA for a vanilla swap portfolio based on the nested LSMC method. Reference to [38] developed Stochastic Grid Bundling Method (SGBM) to calculate the exposure profile for Bermudan options under the Heston model. The accuracy of SGBM will further be improved using multi-variable monomials under multi-dimensional models such as the Heston Hull-White model and the H1HW model. This reference studied CVA with WWR for European and Bermudan options using an intensity model. The paper presented three models to show the dependency between the underlying assets and the intensity in the affine-jump diffusion (AJD) class (see [39], [20], [27], [11]), and [21]). Reference [27] introduced a simple model for the calculations of CVA with RWR and WWR. The WWR is modeled by a deterministic relation between the underlying market factors and the intensity. The research by Y. Shen (see [39]) cal-

culated CVA with WWR on the exposure profiles for multi-asset Bermudan options under the GBM asset dynamics. Authors in [21] have drawn attention to research in which the intensity is modeled by a CIR jump-diffusion model, and in which there is a correlation between the diffusion term and the short rate for portfolios of swaps. Similarly, reference [11] worked on credit spread options.

In this thesis, unilateral CVA will be computed for an American put option based on standard assumptions, using the MC-Tree method.

Chapter 2

Pricing Single-Asset Options Using MC-Tree Method

2.1. Introduction

Various existing versions of binomial models use different choices for parameters. We aim to improve the accuracy and speed of the binomial method by applying Monte Carlo simulation on these parameters. The idea is to generate the tree directions and probabilities through a parameter. This parameter is drawn from a probability density, called the mixing density. A formula will be provided for the resulting compound density that is then generated by the tree. The goal is to find a mixing density such that the corresponding compound density is close to a Gaussian density when working with the additive representation of the tree. It does not seem to be possible to find a mixing density for which the compound density is exactly Gaussian. We succeed in specifying a class of mixing densities for which the compound density

is rational (and hence smooth). The standard Gaussian density can be approximated by rational densities as follows from the following well-known limit:

$$\lim_{N \rightarrow \infty} \frac{\frac{1}{\sqrt{2\pi}}}{\left(1 + \frac{x^2/2}{N}\right)^N} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Therefore, it is not unreasonable to construct rational compound densities to approximate the standard Gaussian density function. It is well-known from the literature that the standard Gaussian distribution $\mathcal{N}(0, 1)$ with the pdf $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ maximizes the entropy integral

$$Ent(f) := - \int_{-\infty}^{\infty} f(x) \log(f(x)) dx,$$

subject to the constraints

$$\int_{-\infty}^{\infty} f(x) dx = 1, \quad \int_{-\infty}^{\infty} x f(x) dx = 0, \quad \int_{-\infty}^{\infty} x^2 f(x) dx = 1, \quad f(x) \geq 0 \quad \forall x \in \mathbb{R}$$

and more generally a Gaussian density with given mean and variance maximizes the entropy among all probability densities with that same mean and variance.

2.2. Parameters

Set a mesh on the time horizon $[0, T]$: $0 = t_0 < t_1 < \dots < t_N = T$ with $t_n = \delta n$ for $n = 0, 1, \dots, N$ where $\delta := \frac{T}{N}$. Let $S^n \approx S(t_n) = S(\delta n)$ where $S(t_n)$ is the true underlying asset price at time t_n generated by the dynamic process, e.g. the Geometric Brownian motion (GBM). If we have reached time point $n\delta$ and

the corresponding asset price S^n , then the multiplicative binomial tree has the probability $p_1 > 0$ of moving "down" to $S^n e^d$ and $p_2 > 0$ of moving "up" to $S^n e^u$, where $p_1 + p_2 = 1$, $S^n > 0$, $S^0 = S_0 > 0$ fixed. Here, $u > d$; this is actually all that is required, and $u > 0$ and $d < 0$ is not required. So for $n = 0, 1, \dots$, we have

$$S^{(n+1)} = \begin{cases} S^n e^u & \text{with prob } p_2 \\ S^n e^d & \text{with prob } p_1. \end{cases}$$

Equivalently, we have the following additive tree.

$$\log(S^{(n+1)}) = \begin{cases} \log(S^n) + u & \text{with prob } p_2 \\ \log(S^n) + d & \text{with prob } p_1. \end{cases}$$

As the mean and variance can always be adapted using an affine transformation, we will first consider the case in which $S_0 = 1$, mean=0 and variance=1 for each time-step. This implies

$$\begin{aligned} p_2 + p_1 &= 1, \\ p_2 u + p_1 d &= 0, \\ p_2 u^2 + p_1 d^2 &= 1, \\ p_2 > 0, p_1 > 0. \end{aligned}$$

The family of all solutions can be parametrized by an angle θ , with $0 < \theta < \frac{\pi}{2}$, as follows:

- $p_1^{1/2} = \cos(\theta)$,
- $p_2^{1/2} = \sin(\theta)$,

- $u = \sqrt{\frac{p_1}{p_2}}$,
- $d = -\sqrt{\frac{p_2}{p_1}}$.

For a geometric interpretation of the angle θ , we refer to Sierag and Hanzon [40]. Hence, we now consider θ to be a random variable with distribution function P_m , supported on $(0, \pi/2)$, providing us with a mixing distribution on the tree. We can summarize this schematically as follows:

$$\log(S^N) \mid \theta \sim \text{Binomial}, \quad \theta \sim P_m.$$

Here by "Binomial" we mean the distribution of $\log(S^N)$ at the final nodes of the tree. Combining the binomial distribution on the log-asset-prices with the distribution on θ will result in a compound density for $\log(S^N)$. We will say more about how the compound density can be calculated in the next section.

The proposed approach is now to try to compute an expected value of a payoff function defined on $\log(S^N)$, with respect to the compound density by

1. assembling, say, M drawings of the variable θ and
2. for each θ to "run the tree" to obtain an approximation to the expected value of the payoff, and
3. to compute the average and standard deviation of the tree-outcomes to obtain a Monte Carlo estimation of the expected payoff, as well as a confidence interval.

Informally we refer to this procedure as "shaking the tree". In terms of Monte Carlo theory, this method falls under the category "variance reduction by conditioning" [9]. The idea is that because the tree outcomes will already be very close to the true

value (especially for deeper trees), the Monte Carlo outcomes will be very accurate. Instead of θ , we can also use $\tau = \tan(\theta)$ in which case $d = -\tau$, $u = \frac{1}{\tau}$. This is illustrated by the two-step additive tree in Figure 2.1.

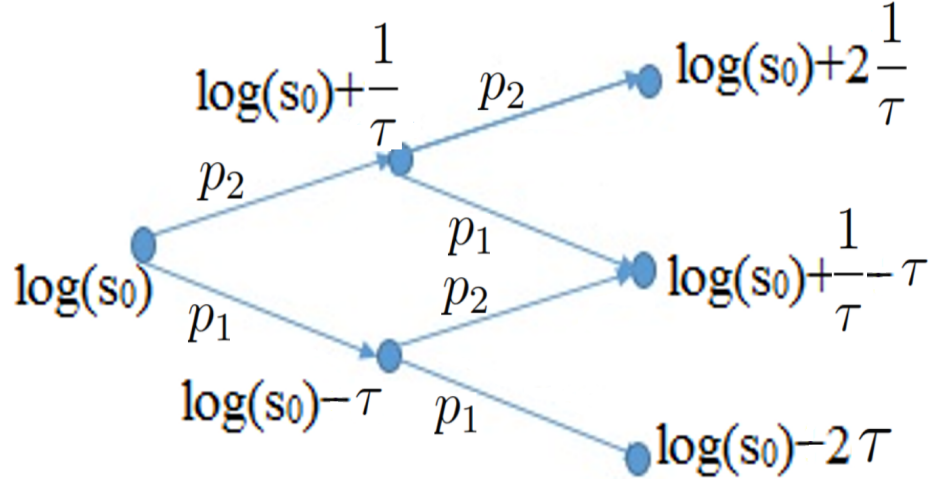


Figure 2.1: Evolution of S_t in a two-step additive binomial tree. Here $p_2 = \frac{\tau^2}{1+\tau^2}$, $p_1 = \frac{1}{1+\tau^2}$.

2.3. Preliminaries to the Derivation of the General Compound Density Formula

To get good approximations of the Gaussian distribution, we try to find mixing densities for which the entropy of the compound density is high. The motivation is that our tree construction forces it to have a distribution on the end-nodes at step N with zero mean and variance equal to N . This will then also hold for the compound density of any such MC-Tree. If τ denotes a parameter characterizing a 1-step tree and hence the corresponding tree of arbitrary tree depth, and X the random variable

resulting from the MC-Tree procedure at time step N , then

$$\mathbb{E}_{X|\tau}(X) = 0, \quad \mathbb{E}_{X|\tau}(X^2) = N.$$

This implies that

$$\mathbb{E}_X(X) = \mathbb{E}_\tau(\mathbb{E}_{X|\tau}(X | \tau)) = 0, \quad \mathbb{E}_X(X^2) = \mathbb{E}_\tau(\mathbb{E}_{X|\tau}(X^2 | \tau)) = N.$$

Let $N \in \mathbb{N}$. Define the log-price of the asset at the final step from the additive tree

$$x_{N,k} = -(N-k)\tau + k\frac{1}{\tau}, \quad \tau > 0, \quad k = 0, 1, 2, \dots, N.$$

Note that

$$\text{Range of } x_{N,k} = \begin{cases} (-\infty, 0) & \text{if } k = 0, \\ \mathbb{R} & \text{if } k = 1, 2, \dots, N-1, \\ (0, \infty) & \text{if } k = N. \end{cases}$$

As the derivative $x'_{N,k}(\tau) = -(N-k) - \frac{k}{\tau^2} < 0 \forall \tau \in (0, \infty)$, it follows that $x_{N,k}$ is monotonically decreasing for each $N \in \mathbb{N}$ and $k \in \{0, 1, \dots, N\}$. Hence, $x_{N,k}(\tau)$ has an inverse function $\tau_k(x)$ with domain $(-\infty, 0)$ if $k = 0$, \mathbb{R} if $k = 1, \dots, N-1$ and $(0, \infty)$ if $k = N$; and range $(0, \infty)$ in all cases.

As can easily be verified, one has

$$\tau_0(x) = -\frac{x}{N} \text{ if } x < 0; \tag{2.1}$$

$$\tau_k(x) = \frac{-x + \sqrt{x^2 + 4k(N-k)}}{2(N-k)} \text{ if } k \neq 0, N; \tag{2.2}$$

$$\tau_N(x) = \frac{N}{x} \text{ if } x > 0. \tag{2.3}$$

Note that $\tau_k(x)\tau_{N-k}(-x) = 1$ holds for all x for which the left hand side is defined. For later reference, we also define

$$y_k(x) = \sqrt{x^2 + 4k(N-k)}, \quad k = 1, 2, \dots, N-1. \quad (2.4)$$

The variable y_k can also be expressed in terms of τ_k , as follows.

$$y_k(x) = (N-k)\tau_k + k\frac{1}{\tau_k} = (N-k)\tau_k(x) + k\tau_{N-k}(-x).$$

Now, for an additive binomial tree with $d = -\tau$ with the probability $\frac{1}{1+\tau^2}$ and $u = \frac{1}{\tau}$ with the probability $\frac{\tau^2}{1+\tau^2}$, the probability distribution of the values at the nodes at the tree depth N can be expressed as:

$$X_{N|\tau} = \begin{cases} x_{N,0}(\tau) & \text{with prob } g_0(\tau) := \frac{1}{(1+\tau^2)^N}, \\ x_{N,k}(\tau) & \text{with prob } g_k(\tau) := \binom{N}{k} \frac{(\tau^2)^k}{(1+\tau^2)^N}, \quad k = \overline{1, N-1}, \\ x_{N,N}(\tau) & \text{with prob } g_N(\tau) := \frac{(\tau^2)^N}{(1+\tau^2)^N}. \end{cases} \quad (2.5)$$

Note that $X_{N|\tau}$ is the sum of N stochastically independent copies of the random variable $X_{1|\tau}$, and hence it has mean $\mathbb{E}[X_{N|\tau}] = 0$, and the variance at $\mathbb{E}[(X_{N|\tau})^2] = N$. So $\frac{1}{\sqrt{N}}X_{N|\tau}$ is a random variable with mean zero and variance one.

Application of the Central Limit Theorem tells us that the cumulative distribution function (cdf) of $\frac{1}{\sqrt{N}}X_{N|\tau}$ converges weakly in distribution to the cdf of a standard Gaussian random variable for $N \rightarrow \infty$ (and $\tau > 0$ fixed) (see [34]). The cdf of $X_{N|\tau}$

can be described as follows.

$$\begin{aligned}
 F(x | \tau) &= P(X_{N|\tau} \leq x | \tau) = g_0(\tau) \mathbb{1}_{\{x \geq x_{N,0}(\tau) \ \& \ x < 0\}} \\
 &+ \sum_{k=1}^{N-1} g_k(\tau) \mathbb{1}_{\{x \geq x_{N,k}(\tau)\}} + g_N(\tau) \mathbb{1}_{\{x \geq x_{N,N}(\tau) > 0\}}.
 \end{aligned}$$

If we now consider the tree parameter τ as random with the probability density function $p_m(\tau)$ (we will call $p_m(\tau)$ the "mixing density") then the resulting compound cdf of X will be $Q(x) := \int_0^\infty F(x | t) p_m(t) dt$.

2.4. General Compound Density Formula

Theorem 2.4.1. *The compound probability density function $q(x) = Q'(x)$ of X satisfies the following formula*

$$q(x) = \sum_{k=0}^N C_k(x), \quad (2.6)$$

where:

- $C_0(x) = \frac{1}{(1+\tau_0^2)^N} p_m(\tau_0) \frac{1}{N} \mathbb{1}_{\{x \leq 0\}}$,
- $C_N(x) = \frac{(\tau_N^2)^N}{(1+\tau_N^2)^N} p_m(\tau_N) \frac{N}{x^2} \mathbb{1}_{\{x > 0\}}$,
- $C_k(x) = \binom{N}{k} \frac{(\tau_k^2)^k}{(1+\tau_k^2)^N} p_m(\tau_k) \frac{\tau_k}{y_k}$, $k = \overline{1, N-1}$.

Proof. Fix N . As for each $k \in \{0, \dots, N\}$, the function $x_{N,k}(\tau)$ is monotonically

decreasing with inverse $\tau_k(x)$, we can write

$$\begin{aligned} Q(x) &= \int_0^\infty F(x | \tau) p_m(\tau) d\tau = \mathbb{1}_{\{x < 0\}} \int_{\{\tau \geq \tau_0(x) \& x < 0\}} g_0(\tau) p_m(\tau) d\tau \\ &+ \sum_{k=1}^{N-1} \int_{\{\tau \geq \tau_k(x)\}} g_k(\tau) p_m(\tau) d\tau + \mathbb{1}_{\{x > 0\}} \int_{\{\tau \geq \tau_N(x) \& x > 0\}} g_N(\tau) p_m(\tau) d\tau. \end{aligned}$$

For $k = 1, \dots, N - 1$, we have

$$1 = \frac{d}{dx} x_{N,k}(\tau_k(x)) = \left\{ -(N - k) - \frac{k}{\tau_k} \right\} \tau_k' = -\left((N - k) \tau_k + k \frac{1}{\tau_k} \right) \frac{\tau_k'}{\tau_k} = -\frac{y_k}{\tau_k} \tau_k'.$$

Therefore,

$$\tau_k'(x) = -\frac{\tau_k}{y_k}, \quad k = 1, \dots, N - 1.$$

Recall $\tau_0(x) = -\frac{x}{N} \Rightarrow \tau_0'(x) = -\frac{1}{N}$, $x < 0$, and $\tau_N(x) = \frac{N}{x} \Rightarrow \tau_N'(x) = -\frac{N}{x^2}$, $x > 0$.

Now taking the derivative of $Q(x)$, we obtain

$$\begin{aligned} q(x) &= \mathbb{1}_{\{x < 0\}} g_0(\tau_0(x)) p_m(\tau_0(x)) \frac{1}{N} + \sum_{k=1}^{N-1} g_k(\tau_k(x)) p_m(\tau_k(x)) \frac{\tau_k(x)}{y_k(x)} \\ &+ \mathbb{1}_{\{x > 0\}} g_N(\tau_N(x)) p_m(\tau_N(x)) \frac{N}{x^2} = C_0(x) + \sum_{k=1}^{N-1} C_k(x) + C_N(x). \end{aligned}$$

□

Remark: It can be proved that the standard Gaussian distribution can not be obtained in a 1-step tree. At $N = 1$, the compound density $q(x) = C_0(x) + C_1(x)$,

where

$$\begin{aligned} C_0(x) &= \frac{1}{1+x^2} p_m(-x) \mathbb{1}_{\{x < 0\}}, \\ C_1(x) &= \frac{1}{x^2(1+x^2)} p_m\left(\frac{1}{x}\right) \mathbb{1}_{\{x > 0\}}. \end{aligned}$$

Set $C_0(x) = f(x)$ on the negative half line, where f is the standard Gaussian pdf. We obtain $p_m(-x) = f(-x)(1+x^2)$, $x < 0$. Then using $p_m(x)$, one can explicitly calculate the corresponding $C_1(x) = \frac{f(\frac{1}{x})}{x^4}$. Observe that $\lim_{x \rightarrow 0} \frac{f(\frac{1}{x})}{x^4} = 0$. Hence, it is itself not the standard Gaussian pdf on the positive half line. In other word, the standard Gaussian pdf can be obtained on the negative half line. This requirement fully determines the mixing density, and this mixing density produces a compound density on the positive half-line that is not a standard Gaussian density at all.

2.5. Mixing Density

As is well-known, due to the convexity of the function $g(y) = y \log(y)$, if $f(x)$ is a pdf on \mathbf{R} then $\frac{f(x)+f(-x)}{2}$ has entropy at least as high as $f(x)$. As we are looking for compound densities with high entropy, the consequence of this is that we can restrict our search to mixing densities that produce an even compound density function. In terms of the mixing distribution, this translates into considering mixing densities which are invariant under a permutation of the two axes in the binomial tree. So we use mixing probabilities on the recombining binomial tree such that any path and its reflection have the same probability.

Proposition 2.5.1. *Consider the case the mixing density is $c(p_1^{1/2})^{m_1}(p_2^{1/2})^{m_2}dp_1$. If $m_1 + m_2$ is odd, then the corresponding pdf of τ is not a rational function. Otherwise, if $m_1 + m_2$ is even, then the pdf of τ is rational, where $\tau = \tan(\theta) = \sqrt{\frac{p_2}{p_1}}$.*

Proof. We have $\tau = \sqrt{\frac{1}{p_1} - 1}$ so $p_1 = \frac{1}{1+\tau^2}$. Hence

$$dp_1 = \left| \frac{-2\tau}{(1+\tau^2)^2} \right| d\tau.$$

Hence,

$$\begin{aligned} c (p_1^{1/2})^{m_1} (p_2^{1/2})^{m_2} dp_1 &= 2c (1+\tau^2)^{-\frac{m_1}{2}} \tau^{m_2+1} (1+\tau^2)^{-\frac{m_2}{2}} (1+\tau^2)^{-2} d\tau \\ &= 2c \tau^{m_2+1} \frac{1}{(1+\tau^2)^{\frac{m_1+m_2}{2}+2}} d\tau. \end{aligned}$$

It follows that if $m_1 + m_2$ is even, the pdf in terms of τ is rational. Otherwise, the pdf is not rational. \square

Our strategy is to use polynomials which are a linear combination of monomials with even total degree.

Proposition 2.5.2. *Every symmetric polynomial of the form $(p_1^{1/2})^{m_1}(p_2^{1/2})^{m_2} + (p_1^{1/2})^{m_2}(p_2^{1/2})^{m_1}$ with $m_1 + m_2 \in 2\mathbb{N}$ and $m_1, m_2 \in \mathbb{N}$ is a linear combination of monomials $(p_1^{1/2} p_2^{1/2})^k$, $k = 1, 2, \dots, m_1 \vee m_2$, where $m_1 \vee m_2 := \max(m_1, m_2)$.*

Proof. First, note that at $p_1 = 0$ as well as at $p_2 = 0$, then polynomial

$$(p_1^{1/2})^{m_1} (p_2^{1/2})^{m_2} + (p_1^{1/2})^{m_2} (p_2^{1/2})^{m_1} = 0.$$

If this is equal to a linear combination of monomials

$$(p_1^{1/2} p_2^{1/2})^k, \quad k \in \{0, 1, 2, \dots, m_1 \vee m_2 := \max(m_1, m_2)\}$$

then the coefficient of $(p_1^{1/2} p_2^{1/2})^0 = 1$ will be zero. Hence, we can restrict the set of values for k to the set $\{1, 2, 3, \dots, m_1 \vee m_2\}$. We will prove this proposition by induction on $m_1 \vee m_2$.

If $m_1 \vee m_2 = 1$ then both $m_1 = 1$ and $m_2 = 1$ because $m_1, m_2 \in \mathbb{N}$. Hence, there is nothing to prove in this case.

Suppose the statement in the proposition is true for $m_1 \vee m_2 \leq m_0$ (Induction Hypothesis).

We want to show that the statement also holds in case $m_1 \vee m_2 = m_0 + 1$.

Without loss of generality, consider

$$(p_1^{1/2})^{m_1} (p_2^{1/2})^{m_2} + (p_1^{1/2})^{m_2} (p_2^{1/2})^{m_1}, \quad m_1 - m_2 = 2\Delta, \quad \Delta \in \mathbb{N}, \quad m_1 = m_0 + 1,$$

then we can write this polynomial as

$$(p_1^{1/2})^{m_2} (p_2^{1/2})^{m_2} \{(p_1^{1/2})^{2\Delta} + (p_2^{1/2})^{2\Delta}\}.$$

Now consider

$$(p_1^{1/2})^{2\Delta} + (p_2^{1/2})^{2\Delta}.$$

If $\Delta = 1$ then we get $p_1 + p_2 = 1$ and hence we are done.

If $\Delta > 1$ then we can write

$$\begin{aligned}
(p_1^{1/2})^{2\Delta} + (p_2^{1/2})^{2\Delta} &= p_1^\Delta + p_2^\Delta \\
&= (p_1 + p_2)^\Delta - \sum_{l=1}^{\Delta-1} \binom{\Delta}{l} \{p_1^l p_2^{\Delta-l}\} \\
&= 1 - \sum_{l=1}^{\Delta-1} \binom{\Delta}{l} \{p_1^l p_2^{\Delta-l}\} \\
&= 1 - \frac{1}{2} \sum_{l=1}^{\Delta-1} \binom{\Delta}{l} \{p_1^l p_2^{\Delta-l} + p_1^{\Delta-l} p_2^l\} \\
&= 1 - \frac{1}{2} \sum_{l=1}^{\Delta-1} \binom{\Delta}{l} \{(p_1^{1/2})^{2l} (p_2^{1/2})^{2(\Delta-2l)} + (p_1^{1/2})^{2\Delta-2l} (p_2^{1/2})^{2l}\};
\end{aligned}$$

and for each index $l \in \{1, 2, \dots, (\Delta - 1)\}$ in the summation, we have

$((2l) \vee 2(\Delta - l)) \leq 2\Delta - 2$, so for each term we can apply the induction hypothesis to conclude that the proposition is true as $2\Delta - 2 + m_2 = m_1 - 2 = m_0 + 1 - 2 = m_0 - 1 \leq m_0$. Note that we have used $\binom{\Delta}{l} = \binom{\Delta}{\Delta-l}$. \square

Remark:

1. If $m_1 = 0$ or $m_2 = 0$ then statement in the proposition still holds if we allow $k = 0$ as well.
2. So it follows that we can write $(p_1^{1/2})^{m_1} (p_2^{1/2})^{m_2} + (p_1^{1/2})^{m_2} (p_2^{1/2})^{m_1}$ as a linear combination of $(p_1^{1/2})^k (p_2^{1/2})^k$. Now note that in the reduction process the total degree never increases. Therefore $2k \leq m_1 + m_2 \Leftrightarrow k \leq \frac{m_1 + m_2}{2}$.

We can conclude that it suffices if we want to restrict the mixing densities in terms of p_1, p_2 to symmetric polynomials in $p_1^{1/2}, p_2^{1/2}$ corresponding to rational pdfs in terms of $\tau = \sqrt{\frac{p_2}{p_1}}$, to linear combination of monomials of the type $(p_1^{1/2} p_2^{1/2})^k, k \in \mathbb{N}_0$. For

given k , such a monomial corresponds to a pdf in τ that can be derived as we will see shortly.

A relatively simple class of mixing densities satisfying this invariance is given, in terms of the parameter p_1 , by

$$\frac{1}{2}c_m(p_1^{1/2}p_2^{1/2})^{m-2}dp_1, \quad m \in \mathbb{N},$$

where $p_1 > 0$, $p_2 > 0$, $p_1 + p_2 = 1$ and c_m is a normalization constant.

Recalling from section 2.2, $p_1 = \cos^2(\theta)$, $p_2 = \sin^2(\theta)$, the transformation $\tau = \tan(\theta)$ leads to $p_1 = \frac{1}{1+\tau^2}$, $p_2 = \frac{\tau^2}{1+\tau^2}$, so

$$\frac{1}{2}c_m(p_1^{1/2}p_2^{1/2})^{m-2}dp_1 = \frac{1}{2}c_m\left(\frac{1}{1+\tau^2}\right)^{(m-2)/2}\left(\frac{\tau^2}{1+\tau^2}\right)^{(m-2)/2}\left|d\left(\frac{1}{1+\tau^2}\right)\right| = c_m\frac{\tau^{m-1}}{(1+\tau^2)^m}d\tau, \quad (2.7)$$

where the constant c_m is given by

$$c_m = \frac{1}{\int_0^\infty \frac{\tau^{m-1}}{(1+\tau^2)^m}d\tau}. \quad (2.8)$$

In terms of the parameter θ we obtain a third representation of these mixing densities:

$$\frac{1}{2}c_m(p_1^{1/2}p_2^{1/2})^{m-2}dp_1 = c_m(\cos(\theta)\sin(\theta))^{m-1}d\theta. \quad (2.9)$$

The idea is now to apply the MC technique and draw τ from this probability distribution on $(0, \infty)$. We will also make use of the transformation

$$\tau = \tan(\theta) \Leftrightarrow \theta = \arctan(\tau), \quad \theta \in (0, \pi/2).$$

Drawing τ can then be replaced by drawing θ and using $\tau = \tan(\theta)$.

2.6. Monte Carlo Drawing

In order to carry out the Monte Carlo simulations, we need to be able to draw independent samples from the mixing distribution. A general technique in case the cumulative distribution function (cdf) is available is to draw random samples from the uniform distribution on the interval $[0, 1]$ and to use the inverse function of the cdf to obtain the desired samples. Note that as our mixing densities are everywhere positive that the inverse of its cdf exists, and given any drawing from the uniform distribution, the corresponding sample from the mixing distribution can be found, for instance, by a bisection method. Therefore what remains is to find the cdf of our mixing distributions. One way to do that is to work out the cdf of the mixing density in terms of the angle θ . This can be done as follows:

Note that

$$\begin{aligned}
p_m(\theta) &= c_m \cos(\theta)^{m-1} \sin(\theta)^{m-1} \\
&= c_m \left(\frac{e^{i\theta} + e^{-i\theta}}{2}\right)^{m-1} \left(\frac{e^{i\theta} - e^{-i\theta}}{2i}\right)^{m-1} \\
&= c_m 2^{-2(m-1)} \operatorname{Re}[(-i)^{m-1} (e^{i2\theta} - e^{-i2\theta})^{m-1}] \\
&= c_m 2^{-2(m-1)} \operatorname{Re}[(-i)^{m-1} e^{-i2\theta(m-1)} (e^{i4\theta} - 1)^{m-1}] \\
&= c_m 2^{-2(m-1)} \operatorname{Re}[(-i)^{m-1} e^{-i2\theta(m-1)} \sum_{s=0}^{m-1} \binom{m-1}{s} e^{i4s\theta} (-1)^{m-1-s}] \\
&= c_m 2^{-2(m-1)} \operatorname{Re}[(-i)^{m-1} \sum_{s=0}^{m-1} \binom{m-1}{s} e^{i(4s-2(m-1))\theta} (-1)^{m-1-s}].
\end{aligned}$$

- In case m is odd, this has the following primitive function

$$\begin{aligned}
& c_m 2^{-2(m-1)} \operatorname{Re} [(-i)^{m-1} \sum_{s=0, s \neq \frac{m-1}{2}}^{m-1} \binom{m-1}{s} \frac{(-1)^{m-1-s}}{i(4s-2(m-1))} e^{i(4s-2(m-1))\theta} \\
& \quad + (-i)^{m-1} \binom{m-1}{(m-1)/2} \theta (-1)^{(m-1)/2} + \tilde{C}_m] \\
& = c_m 2^{(-2m+1)} \operatorname{Re} [(-i)^m \sum_{s=0, s \neq \frac{m-1}{2}}^{m-1} \binom{m-1}{s} \frac{(-1)^{m-1-s}}{(2s-(m-1))} e^{i(4s-2(m-1))\theta} \\
& \quad + (-1)^{(m-1)/2} 2 \binom{m-1}{(m-1)/2} \theta (-1)^{(m-1)/2} + 2\tilde{C}_m] \\
& = c_m 2^{(-2m+1)} [(-1)^{(m-1)/2} \sum_{s=0, s \neq \frac{m-1}{2}}^{m-1} \binom{m-1}{s} \frac{(-1)^{m-1-s}}{(2s-(m-1))} \sin((4s-2(m-1))\theta) \\
& \quad + 2 \binom{m-1}{(m-1)/2} \theta + 2\tilde{C}_m],
\end{aligned}$$

where \tilde{C}_m , $m = 1, 2, \dots$ are real integration constants. As the cdf $F_m(\theta)$ has its support on $(0, \frac{\pi}{2})$, we have $F_m(0) = 0$. Hence, in case m is odd, $\tilde{C}_m = 0$.

It follows that

$$c_m = \frac{2^{2m-1}}{\binom{m-1}{(m-1)/2} \pi}.$$

- In case m is even, we can perform similar calculations, as follows:

$p_m(\theta)$ has the following primitive function

$$\begin{aligned}
& c_m 2^{-2(m-1)} \operatorname{Re} [(-i)^{m-1} \sum_{s=0}^{m-1} \binom{m-1}{s} \frac{(-1)^{m-1-s}}{i(4s-2(m-1))} e^{i(4s-2(m-1))\theta} + \tilde{C}_m] \\
& = c_m 2^{(-2m+1)} \operatorname{Re} [(-i)^m \sum_{s=0}^{m-1} \binom{m-1}{s} \frac{(-1)^{m-1-s}}{(2s-(m-1))} e^{i(4s-2(m-1))\theta} + \tilde{C}_m]
\end{aligned}$$

$$= c_m 2^{(-2m+1)} (-1)^{(m/2)} \sum_{s=0}^{m-1} \binom{m-1}{s} \frac{(-1)^{m-1-s}}{(2s - (m-1))} \cos((4s - 2(m-1))\theta) + \tilde{C}_m.$$

$$\tilde{C}_m = c_m 2^{(-2m+1)} (-1)^{(m/2)+1} \sum_{s=0}^{m-1} \binom{m-1}{s} \frac{1}{(2s - (m-1))} (-1)^{m-1-s}.$$

2.7. A Particular Compound Density

Theorem 2.7.1. *Let m be odd and let the mixing density be given by $p_m(\tau) = c_m \frac{\tau^{m-1}}{(1+\tau^2)^m}$, where c_m is the normalizing constant as described in Equation 2.8. The compound density takes the form $q_m(x) = c_m \frac{A_m(x)}{(x^2+N^2)^{N+m}}$, where $A_m(x)$ is a polynomial with rational coefficients and degree at most $2(N+m-1)$.*

Remark: By abuse of notation, we use $p_m(\tau)$ for the mixing density in term τ .

Note that $p_m(\tau)$ is different from $p_m(\theta)$.

Proof. We distinguish the two following cases.

- N even:

$$q(x) = (C_0 + C_N)(x) + C_{N/2}(x) + \sum_{k=1}^{\frac{N}{2}-1} (C_k + C_{N-k})(x).$$

- N odd:

$$q(x) = (C_0 + C_N)(x) + \sum_{k=1}^{\frac{N-1}{2}} (C_k + C_{N-k})(x).$$

From Theorem 2.4.1 combined with the expression $p_m(\tau)$, we have

$$\begin{aligned}
C_0(x) &= \frac{1}{(1 + \tau_0^2)^N} p_m(\tau_0) \frac{1}{N} \mathbb{1}_{\{x \leq 0\}} \\
&= c_m \frac{\tau_0^{m-1}}{(1 + \tau_0^2)^{N+m}} \frac{1}{N} \mathbb{1}_{\{x \leq 0\}} \\
&= c_m (-1)^{m-1} \frac{x^{m-1} N^{(2N+m)}}{(x^2 + N^2)^{N+m}} \mathbb{1}_{\{x \leq 0\}} \\
&= c_m \frac{x^{m-1} N^{(2N+m)}}{(x^2 + N^2)^{N+m}} \mathbb{1}_{\{x \leq 0\}}.
\end{aligned}$$

Here we have used $(-1)^{m-1} = 1$ as m is odd.

$$\begin{aligned}
C_N(x) &= \frac{(\tau_N^2)^N}{(1 + \tau_N^2)^N} p_m(\tau_N) \frac{N}{x^2} \mathbb{1}_{\{x > 0\}} \\
&= c_m \frac{\tau_N^{2N+m-1}}{(1 + \tau_N^2)^{N+m}} \frac{N}{x^2} \mathbb{1}_{\{x > 0\}} \\
&= c_m \frac{x^{m-1} N^{(2N+m)}}{(x^2 + N^2)^{N+m}} \mathbb{1}_{\{x > 0\}}.
\end{aligned}$$

It follows that the first term $(C_0 + C_N)(x)$ can be written explicitly as follows for all real values of x :

$$(C_0 + C_N)(x) = c_m \frac{N^{2N+m} x^{m-1}}{(x^2 + N^2)^{N+m}} = c_m \frac{A_m(x, 0)}{(x^2 + N^2)^{N+m}},$$

where $A_m(x, 0) := N^{2N+m} x^{m-1}$. Recall that $(m - 1)$ is even, so $(C_0 + C_N)(x)$ is even w.r.t x .

Now we consider the case $1 \leq k \leq N$. Observe that

$$\left(\frac{-x + y_k}{2(N - k)}\right) \left(\frac{x + y_k}{2k}\right) = \frac{-x^2 + y_k^2}{4(N - k)k} = \frac{-x^2 + x^2 + 4k(N - k)}{4k(N - k)} = 1.$$

Hence, as $\tau_k = \frac{-x+y_k}{2(N-k)}$ then $\tau_k^{-1} = \frac{x+y_k}{2k}$ and $\tau_{N-k}^{-1} = \frac{x+y_k}{2(N-k)}$. We rewrite C_{N-k} in terms of τ_{N-k}^{-1} as follows:

$$\begin{aligned}
C_{N-k}(x) &= \binom{N}{N-k} \frac{(\tau_{N-k}^2)^{N-k}}{(1 + \tau_{N-k}^2)^N} P_m(\tau_{N-k}) \frac{\tau_{N-k}}{y_{N-k}} \\
&= c_m \binom{N}{N-k} \frac{(\tau_{N-k}^2)^{N-k}}{(1 + \tau_{N-k}^2)^N} \frac{\tau_{N-k}^{m-1}}{(1 + \tau_{N-k}^2)^m} \frac{\tau_{N-k}}{y_{N-k}} \\
&= c_m \binom{N}{N-k} \frac{(\tau_{N-k}^2)^{N-k + \frac{m-1}{2} + 1} \tau_{N-k}^{-1}}{(1 + \tau_{N-k}^2)^{N+m} y_{N-k}} \\
&= c_m \binom{N}{N-k} \frac{1}{(\tau_{N-k}^{-2})^{N-k + \frac{m-1}{2} + 1} \tau_{N-k}^{-1}} \\
&= c_m \binom{N}{N-k} \frac{1}{(1 + \frac{1}{\tau_{N-k}^2})^{N+m} y_{N-k}} \\
&= c_m \binom{N}{N-k} \frac{(\tau_{N-k}^{-2})^{k + \frac{m-1}{2}} \tau_{N-k}^{-1}}{(1 + \tau_{N-k}^{-2})^{N+m} y_{N-k}}.
\end{aligned}$$

Notice that $y_{N-k} = y_k$ and $\binom{N}{N-k} = \binom{N}{k}$. The term $C_k + C_{N-k}$ can be defined w.r.t x and y_k as follows:

$$\begin{aligned}
(C_k + C_{N-k})(x, y_k) &= c_m \binom{N}{k} \frac{(\tau_k^2)^{k + \frac{m-1}{2}} \tau_k}{(1 + \tau_k^2)^{N+m} y_k} + c_m \binom{N}{k} \frac{(\tau_{N-k}^{-2})^{k + \frac{m-1}{2}} \tau_{N-k}^{-1}}{(1 + \tau_{N-k}^{-2})^{N+m} y_{N-k}} \\
&= c_m \binom{N}{k} \frac{1}{2(N-k)} \left[\frac{(\tau_k^2)^{k + \frac{m-1}{2}}}{(1 + \tau_k^2)^{N+m}} \left(\frac{-x}{y_k} + 1 \right) \right. \\
&\quad \left. + \frac{(\tau_{N-k}^{-2})^{k + \frac{m-1}{2}}}{(1 + \tau_{N-k}^{-2})^{N+m}} \left(\frac{x}{y_k} + 1 \right) \right] \\
&= c_m \binom{N}{k} \frac{1}{2(N-k)} \left[\frac{((\frac{-x+y_k}{2(N-k)})^2)^{k + \frac{m-1}{2}}}{(1 + (\frac{-x+y_k}{2(N-k)})^2)^{N+m}} \left(\frac{-x}{y_k} + 1 \right) \right. \\
&\quad \left. + \frac{((\frac{x+y_k}{2(N-k)})^2)^{k + \frac{m-1}{2}}}{(1 + (\frac{x+y_k}{2(N-k)})^2)^{N+m}} \left(\frac{x}{y_k} + 1 \right) \right] \\
&= c_m \binom{N}{k} \frac{(2(N-k))^{2N-2k+m} N(x, y_k)}{D(x, y_k)},
\end{aligned}$$

where

$$\begin{aligned} N(x, y_k) &= (4(N-k)^2 + (-x + y_k)^2)^{N+m} (x + y_k)^{2k+m} \\ &\quad + (4(N-k)^2 + (x + y_k)^2)^{N+m} (-x + y_k)^{2k+m}. \\ D(x, y_k) &= ((4(N-k)^2 + x^2 + y_k^2)^2 - 4x^2 y_k^2)^{N+m} y_k. \end{aligned}$$

Observe that $N(x, y_k) = -N(x, -y_k)$ and $D(x, y_k) = -D(x, -y_k)$, so

$\frac{N(x, y_k)}{D(x, y_k)} = \frac{N(x, -y_k)}{D(x, -y_k)}$. Hence, $(C_k + C_{N-k})(x, y_k)$ is even in y_k . We can see that

$$N(x, 0) = (4(N-k)^2 + (x)^2)^{N+m} (x)^{2k+m} - (4(N-k)^2 + (x)^2)^{N+m} (x)^{2k+m} = 0$$

Hence, we infer that $N(x, y_k)$ is divisible by y_k , so both numerator and denominator are divisible by y_k . We have $D(x, y_k)/y_k$ and $C_k + C_{N-k}$ both are even in y_k , so $N(x, y_k)/y_k$ is even in y_k . This implies that we can express the new numerator $N(x, y_k)/y_k$ and the new denominator $D(x, y_k)/y_k$ both as polynomial in terms of powers of x and powers of y_k^2 . Replacing $y_k^2 = x^2 + 4k(N-k)$, we can conclude that $(C_k + C_{N-k})(x)$ is rational in x . In a similar way we can see that $(C_k + C_{N-k})(x)$ is an even function of x . Observe that

$$\begin{aligned} &((4(N-k)^2 + x^2 + y_k^2)^2 - 4x^2 y_k^2) \\ &= (4(N-k)^2 + x^2 + x^2 + 4k(N-k))^2 - 4x^2(x^2 + 4k(N-k)) \\ &= 4\{(2(N-k)^2 + x^2 + 2k(N-k))^2 - x^4 - 4k(N-k)x^2\} \\ &= 4\{(2N(N-k) + x^2)^2 - x^4 - 4k(N-k)x^2\} \end{aligned}$$

$$\begin{aligned}
&= 4\{x^4 + 4N(N-k)x^2 + 4N^2(N-k)^2 - x^4 - 4k(N-k)x^2\} \\
&= 16(N-k)\{(N-k)x^2 + N^2(N-k)\} = 16(N-k)^2\{x^2 + N^2\}.
\end{aligned}$$

Hence,

$$(C_k + C_{N-k})(x, y_k) = c_m \frac{\tilde{N}(x, y_k)}{\frac{2^{2N+2k+3m}(N-k)^{2k+m}}{\binom{N}{k}} (x^2 + N^2)^{N+m}},$$

where

$$\tilde{N}(x, y_k) = \frac{N(x, y_k)}{y_k}, \quad d_k := \frac{2^{2N+2k+3m}(N-k)^{2k+m}}{\binom{N}{k}}, \quad A_m(x, k) := \frac{\tilde{N}(x, y_k)}{d_k}.$$

Therefore, the term $(C_k + C_{N-k})(x)$ can hence be written as

$$(C_k + C_{N-k})(x) = c_m \frac{A_m(x, k)}{(x^2 + N^2)^{N+m}},$$

where $A_m(x, k)$ is a polynomial with rational coefficients.

Note that in case $k = N/2$ this formula implies that $C_k(x) = \frac{1}{2}c_m \frac{A_m(x, k)}{(x^2 + N^2)^{N+m}}$.

We have the following cases.

- N even:

$$A_m(x) := \sum_{k=0}^{\frac{N}{2}-1} A_m(x, k) + \frac{1}{2}A_m(x, \frac{N}{2}).$$

- N odd:

$$A_m(x) := \sum_{k=0}^{\frac{N-1}{2}} A_m(x, k).$$

The compound density $q(x)$ is an even rational function with common denominator $(x^2 + N^2)^{N+m}$ because it is the sum of even rational functions with the same denominator. The numerator of q/c_m is the sum of polynomials with rational coefficients,

hence is a polynomial with rational coefficients. Notice that as q has integral one over the real line and each of the $(C_k + C_{N-k})(x)$ functions is non-negative, each such function is integrable and hence its codegree must be at least 2. The same argument holds for q itself and so the numerator degree of q will be less than or equal to $2(N + m - 1)$. \square

To compute $A_m(x, k)$, one could use algebraic manipulation with Euler substitution to eliminate all the occurrences of square roots (see Theorem 3.6.2 for more on usage of Euler substitution). Alternatively, one could compute $A_m(x, k)$ using a Lagrange interpolation technique that we will now explain. We need to take $(N + m)$ interpolation points $x_k^i, i = 1, 2, \dots, N + m$ to approximate $A_m(x, k)$ because the degree of numerator $A_m(x, k)$ is at most $2(N + m - 1)$ and $A_m(x, k)$ is even. Here k is fixed for each term $(C_k + C_{N-k})(x)$. We can calculate the values of $A_m(x_k^i, k)$ by noting that

$$A_m(x_k^i, k) = \frac{1}{C_m} (C_k + C_{N-k})(x_k^i) ((x_k^i)^2 + N^2)^{N+m}.$$

By applying Lagrange interpolation method, we can obtain

$$\begin{pmatrix} A_m(x_k^1, k) \\ A_m(x_k^2, k) \\ \vdots \\ A_m(x_k^{N+m}, k) \end{pmatrix} = \begin{pmatrix} 1 & (x_k^1)^2 & (x_k^1)^4 & \dots & (x_k^1)^{2(N+m-1)} \\ 1 & (x_k^2)^2 & (x_k^2)^4 & \dots & (x_k^2)^{2(N+m-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & (x_k^{N+m})^2 & (x_k^{N+m})^4 & \dots & (x_k^{N+m})^{2(N+m-1)} \end{pmatrix} \begin{pmatrix} a_0 \\ a_2 \\ \vdots \\ a_{2(N+m-1)} \end{pmatrix}$$

The matrix is a $(N + m) \times (N + m)$ Vandermonde matrix of interpolation points. It is known to be non-singular as the interpolation points will be distinct. The Lagrange matrix is the known inverse matrix of this Vandermonde matrix, so we can obtain the solution by using the Lagrange coefficients explicitly. Alternatively we can solve this linear system of equations directly by standard methods.

Recall $A_m(x, k) = \tilde{N}(x, y_k)/d_k$, where $\tilde{N}(x, y) = \frac{N(x, y)}{y}$ is a known two-variable

polynomial in x and y with integer coefficients and d_k is a known integer. To be able to get the rational coefficients of $A_m(x, k)$ exactly, we need to take the interpolation points such that *both* x_k^i and $y_k(x_k^i)$ are rational! That is indeed possible as we will now show. Our approach will be based on an Euler substitution (known from the theory of integration).

Let $z_k := -x + y_k = -x + \sqrt{x^2 + 4k(N - k)}$ then $z_k + x = y_k$. It implies that

$$z_k^2 + x^2 + 2z_k x = x^2 + 4k(N - k) \Leftrightarrow 2z_k x = 4k(N - k) - z_k^2 \Leftrightarrow x = \frac{2k(N - k)}{z_k} - \frac{1}{2}z_k.$$

It follows that $y_k = x + z_k = \frac{2k(N - k)}{z_k} + \frac{1}{2}z_k$.

Note that if we choose z_k rational and non-zero ($z_k \in \mathbb{Q} \setminus \{0\}$) then both x and y_k will be rational.

Furthermore, for any $\hat{x} \in \mathbb{R}$ one can calculate $\hat{z}_k = -\hat{x} + \sqrt{\hat{x}^2 + 4k(N - k)} > 0$ and take a positive rational number z_k arbitrarily close to \hat{z}_k and compute the corresponding rational values of x and y_k . By taking z_k sufficiently close to \hat{z}_k , the corresponding values x and y_k will be as close as is desired to \hat{x} and \hat{y}_k . (Warning: care must be taken for cases in which \hat{z}_k is close to zero).

The Lagrangian interpolation method now requires us to solve a Vandermonde-type linear system of $(N + m)$ equations with only *rational* coefficients. The solution will be a vector of rational numbers in \mathbb{Q}^{N+m} . How to solve such systems in case $N+m$ is large is an active area of research in which considerable advances have been made by authors in [41].

Remark 1.

Consider a random variable X which has the mean at zero and the variance at N with the pdf $q(x)dx$. Then $Z := \frac{X}{\sqrt{N}}$ has the mean at zero and the variance at 1 with

the pdf $\sqrt{N}q(z\sqrt{N})dz$.

Remark 2.

For numerical implementation later on in section 2.10, we only search among odd values of m as these are giving a rational compound density. Based on our numerical results on entropy, but also on the Kullback-Leibler (KL) divergence, as well as L^1 distance, the best choice for the integer m in our class of mixing densities is at $m = 9$.

2.8. Option Pricing

2.8.1. Usage of a Bias-Correction

To compute the value of a European option, a price process S_t is modeled by the geometric Brownian motion under its associated "risk-neutral" measure \mathbb{Q}

$$dS_t = rS_t dt + \sigma S_t d\bar{W}_t, \quad t \geq 0$$

where r is the interest rate, $\sigma \in \mathbb{R}$ represents the diffusion coefficient, and \bar{W} is a standard Brownian motion process under \mathbb{Q} (see [28]). Using Itô's lemma with $f(S) = \log(S)$ gives a classic result, in which the process $\log(S)$ follows the normal distribution $\mathcal{N}((r - \sigma^2/2)T + \log(S_0), \sigma^2 T)$ at the end of any interval $[0, T]$ of length T .

We consider the payoff function $\pi(X) = \max(e^X - K, 0)$ for a call option or $\pi(X) = \max(K - e^X, 0)$ for a put option, where $X = \log(S)$.

For option pricing one typically uses the multiplicative tree (as mentioned before). Using this, a multiplicative upward move, u_1 , and a multiplicative downward move,

d_1 , are defined as

$$\begin{aligned} u_1 &= \exp\{u\sigma\sqrt{\delta} + (r - \sigma^2/2)\delta\}, \\ d_1 &= \exp\{d\sigma\sqrt{\delta} + (r - \sigma^2/2)\delta\}, \end{aligned}$$

where u , d , p as described in section 2.2. Here $\delta > 0$ denotes the time-step length and translation and scaling have been applied to introduce the volatility parameter σ and the drift term r .

Using bias-corrected directions (see [40]) gives

$$\begin{aligned} \tilde{u}_1 &:= u_1 e^{\lambda\delta}, \\ \tilde{d}_1 &:= d_1 e^{\lambda\delta}, \end{aligned}$$

where the real number λ is solved from $p\tilde{u}_1 + (1-p)\tilde{d}_1 = e^{r\delta}$. We obtain

$$\lambda = r - \frac{\log(pu_1 + (1-p)d_1)}{\delta}.$$

This correction amounts to replacing $\hat{\mu} = r - \frac{1}{2}\sigma^2$ by $\hat{\mu} = r - \frac{1}{2}\sigma^2 + \lambda$. The resulting tree model in which u_1 , d_1 are replaced by \tilde{u}_1 , \tilde{d}_1 has risk neutral probability $(p, (1-p))$, and is complete and free of arbitrage (see [40]). Hence, any options (financial contingent claim) can be priced in the market described by this tree.

2.8.2. Usage of a Distribution Correction Factor

The compound densities in the additive trees are close to Gaussian but not equal to a Gaussian exactly. To compensate for that, one can employ a distribution correction

factor. This technique is known from the Monte Carlo method of importance sampling [10]. The distribution correction factor, which we will denote by $C(x)$, can be derived, in the context of option pricing, as follows:

Let P denote the time-zero price of an European option with payoff $\pi(X)$ at time T , where $X = \log(S_T)$, then we have

$$\begin{aligned} P &= e^{-rT} \mathbb{E}_Q[\pi(X)] = e^{-rT} \int_{-\infty}^{\infty} \pi(x) f(x) dx = e^{-rT} \int_{-\infty}^{\infty} [\pi(x) \frac{f(x)}{q(x)}] q(x) dx \\ &= e^{-rT} \int_{-\infty}^{\infty} [\pi(x) C(x)] q(x) dx = e^{-rT} \mathbb{E}_q[\pi(X) C(X)], \end{aligned}$$

where $q(x) > 0$, $\forall x \in \mathbb{R}$ is the compound density, and f is the Gaussian density $f(x) \sim \mathcal{N}((r - \sigma^2/2)T + \log(S_0), \sigma^2 T)$.

2.9. Convex Combination and Convex Optimization

In this section, we aim to optimize coefficients of the convex combination of mixing densities. This optimal mixing density gives the highest entropy of the compound density when fixing the tree depth and the powers of the mixing densities.

2.9.1. Convex Combination

Our strategy is search among linear combination of the rational densities

$p_m(\tau)$, $m = 1, 2, 3, \dots$. The linear combinations are restricted to have coefficients that add up to one. Most natural way would be use to convex combination. Consider a finite convex combination of mixing densities, $\hat{P}_\alpha(\tau) = \sum_m \alpha_m p_m(\tau)$ such that $\alpha_m \geq 0$ and $\sum_m \alpha_m = 1$.

We know that $\hat{P}_\alpha(\tau)$ is again a probability density and still holds properties of the mixing density due to the convex combination of probability densities. Given τ_0, τ_k, τ_N as defined in Equations (2.1)-(2.3), y_k as in Equation 2.4, $C_0(x), C_N(x), C_k(x), p_m(\tau_k)$ as defined in Theorem 2.4.1, and $q_m(x)$ as defined in Theorem 2.7.1, we have

$$\begin{aligned} \hat{P}_\alpha(\tau_k) &:= \sum_m \alpha_m p_m(\tau_k). \\ C_0^\alpha(x) &:= \sum_m \alpha_m C_0(x) = \frac{1}{(1 + \tau_0^2)^N} \left(\sum_m \alpha_m p_m(\tau_0) \right) \frac{1}{N} \mathbb{1}_{\{x \leq 0\}} \\ &= \frac{1}{(1 + \tau_0^2)^N} \hat{P}_\alpha(\tau_0) \frac{1}{N} \mathbb{1}_{\{x \leq 0\}}. \\ C_k^\alpha(x) &:= \sum_m \alpha_m C_k(x) = \binom{N}{k} \frac{(\tau_k^2)^k}{(1 + \tau_k^2)^N} \left(\sum_m \alpha_m p_m(\tau_k) \right) \frac{\tau_k}{y_k} \\ &= \binom{N}{k} \frac{(\tau_k^2)^k}{(1 + \tau_k^2)^N} \hat{P}_\alpha(\tau_k) \frac{\tau_k}{y_k}, \quad k = \overline{1, N-1}. \\ C_N^\alpha(x) &:= \sum_m \alpha_m C_N(x) = \frac{(\tau_N^2)^N}{(1 + \tau_N^2)^N} \left(\sum_m \alpha_m p_m(\tau_N) \right) \frac{N}{x^2} \mathbb{1}_{\{x > 0\}} \\ &= \frac{(\tau_N^2)^N}{(1 + \tau_N^2)^N} \hat{P}_\alpha(\tau_N) \frac{N}{x^2} \mathbb{1}_{\{x > 0\}}. \end{aligned}$$

Due to the linearity property, we can derive the probability density of the tree, $\hat{Q}(x)$, where $\tau \sim \hat{P}_\alpha(\tau)$ as follows.

$$\hat{Q}(x) = \sum_m \alpha_m q_m(x) = \sum_{k=0}^N \left(\sum_m \alpha_m C_k(x) \right) = \sum_{k=0}^N C_k^\alpha(x).$$

The most natural way would be to use convex optimization.

2.9.2. Convex Optimization

Our optimization problem can be stated as follows:

$$\max_{\alpha} f(\alpha) \text{ subject to } \alpha_m \geq 0, \sum_m \alpha_m = 1.$$

$f(\alpha)$ denotes the entropy of the compound density $\hat{Q}(x) = \sum_m \alpha_m q_m(x)$.

This optimization problem can be solved by using the Gradient Ascent Method.

Next, the algorithm for this problem will be presented.

Algorithm

- Define a starting vector $\alpha^{(0)}$. We can choose $\alpha^{(0)}$ to be a vector which is already rather close to the solution to make the algorithm computationally efficient. We suggest a starting vector that gives weight 1 to the mixing density with $m=9$, and other elements of the vector are zero. Then calculate the gradient $\nabla f(\alpha^{(0)})$
- Calculate $\alpha^{(1)} = \alpha^{(0)} + \delta \nabla f(\alpha^{(0)})$. If $f(\alpha^{(1)}) > f(\alpha^{(0)})$, we repeat all steps, where we take $\alpha^{(1)}$ as the new starting vector. Otherwise, we calculate $\alpha^{(1)} = \alpha^{(0)} + \frac{\delta}{2} \nabla f(\alpha^{(0)})$. If this does not work, we take $\alpha^{(1)} = \alpha^{(0)} + \frac{\delta}{2^2} \nabla f(\alpha^{(0)})$, etc until $f(\alpha^{(1)}) > f(\alpha^{(0)})$. If $\nabla f(\alpha^{(0)}) \neq 0$ then there exists a value $k \in \mathbb{N}$ such that $f(\alpha^{(0)} + \frac{\delta}{2^k} \nabla f(\alpha^{(0)})) > f(\alpha^{(0)})$.
- The stopping criterion is that $\nabla f_{\alpha^*}(x) = 0$ and vector α_* is the solution. Due to convexity, the global solution will be found in practice if the actual stopping criterion $|\nabla f_{\alpha}(x)| < \epsilon$ is satisfied for some $\epsilon > 0$ sufficiently small.

2.10. Numerical Results

2.10.1. Numerical Results of Entropy, KL Divergence, and L_1 Distance

We work with various experiments for calculations of KL divergence between the compound density and the Gaussian density, and entropy values for different values of N and m . It is noticeable that this KL divergence and the entropy values obtain the minimal and maximal values, respectively, at $m = 9$ when fixing the tree depth N . Considering our experimental results in Table 2.1 and Table 2.2, the compound distribution is best approximated by a Gaussian distribution at $m = 9$. A good

Table 2.1: KL divergence versus N and m .

	$N = 10$	$N = 15$	$N = 20$	$N = 50$
$m = 7$	0.000756	0.000321	0.000176	0.000026
$m = 9$	0.000719	0.000304	0.000167	0.000025
$m = 11$	0.000724	0.000307	0.000169	0.000026
$m = 13$	0.000741	0.000314	0.000173	0.000027

Table 2.2: Entropy values versus N and m .

	$N = 10$	$N = 15$	$N = 20$	$N = 50$
$m = 7$	1.418172	1.418609	1.418755	1.418908
$m = 9$	1.418213	1.418629	1.418767	1.418910
$m = 11$	1.418211	1.418628	1.418767	1.418910
$m = 13$	1.418195	1.418622	1.418764	1.418910

choice of $m = 9$ can be observed clearly from figure 2.2, when fixing the tree depth and increasing value of m .

Figure 2.3 shows the similarity between compound density at $N = 20$, $m = 9$ and the standard Gaussian density. Figure 2.4 shows that the entropy values increase and converge to the entropy of a standard Gaussian density when fixing m and increasing

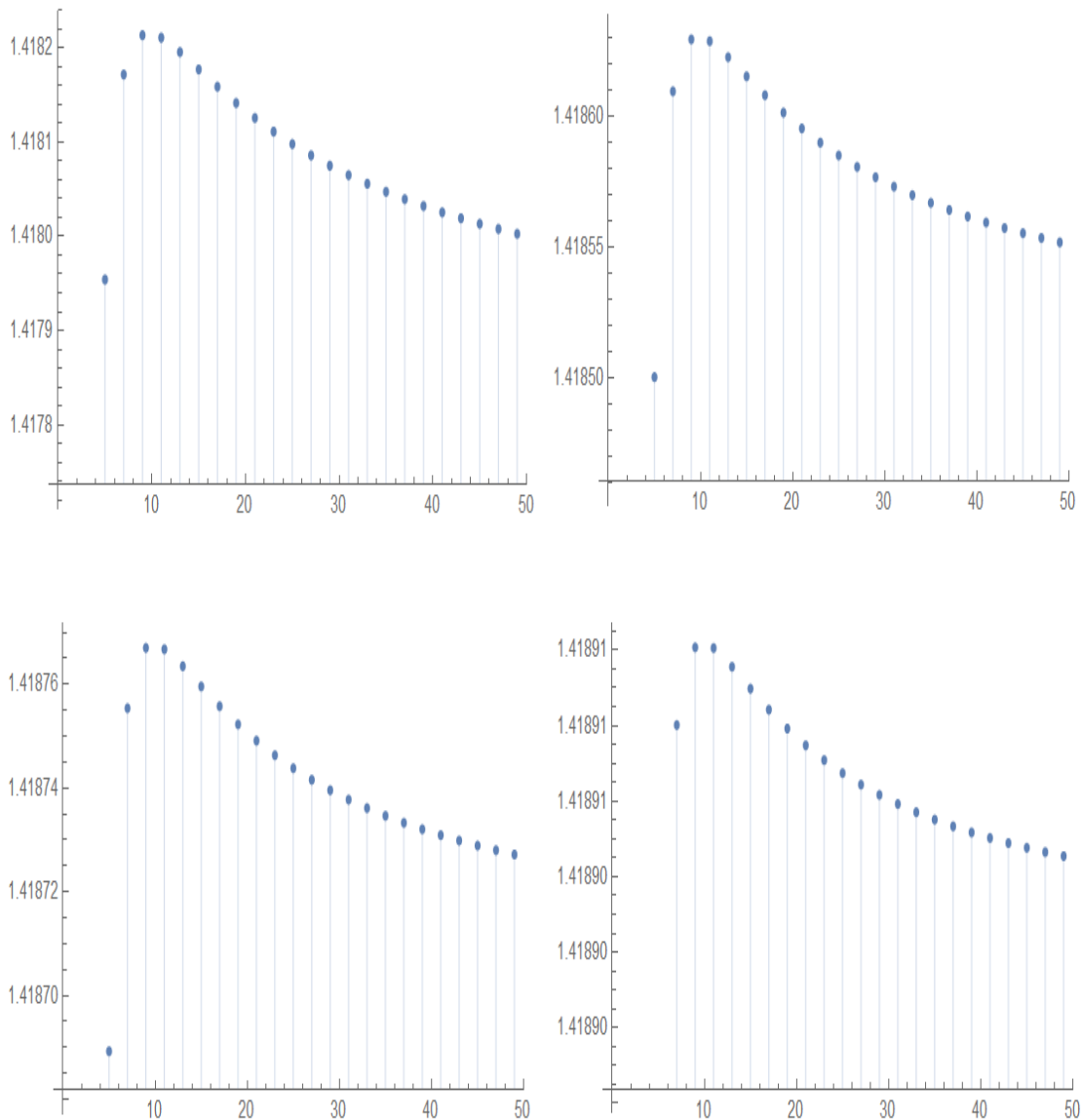


Figure 2.2: Entropy behavior when m is an odd number between 1 and 50. Top left: $N = 10$. Top right: $N = 15$. Bottom left: $N = 20$. Bottom right: $N = 50$.

the tree depth. This is consistent with the maximal entropy theory.

Figure 2.5 shows that the L_1 distance to the Gaussian decreases to zero when fixing m and rising the tree depth from 10 to 120.

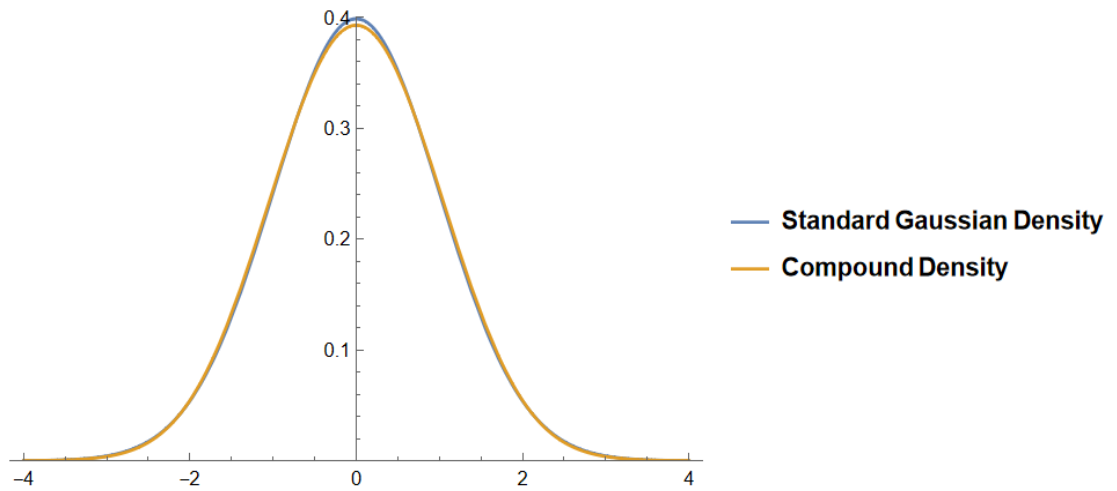


Figure 2.3: Standard Gaussian density versus compound density at $N = 20$, $m = 9$.

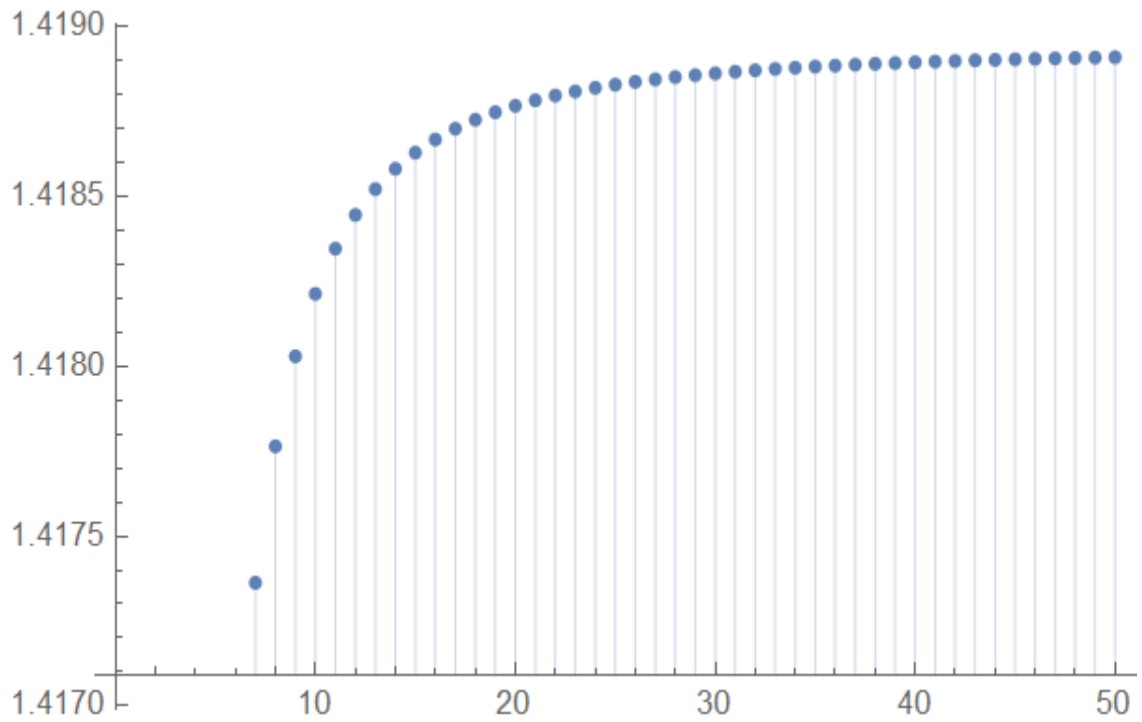


Figure 2.4: Entropy values versus the tree depth N from 1 to 50 when $m = 9$ is fixed.

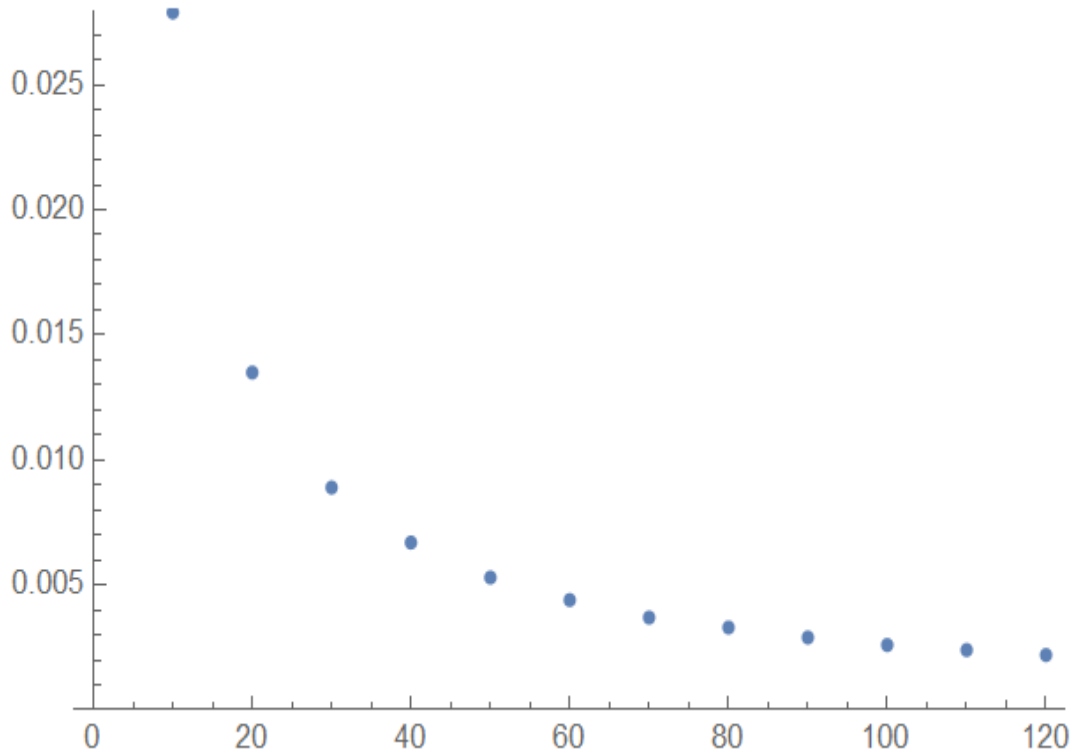


Figure 2.5: L_1 distance versus the tree depth N from 10 to 120 when $m = 9$ is fixed.

2.10.2. Distribution Correction Factor

Figure 2.6 shows the behavior of the distribution correction factor for different tree depths and the corresponding values of m . The distribution correction factor is the ratio of the Gaussian density to the compound density. It is observed that the compound density is closer to the Gaussian density when increasing the tree depth N and fixing m . The compound density and the Gaussian density are almost the

same in the central distribution but dissimilar in the tails.

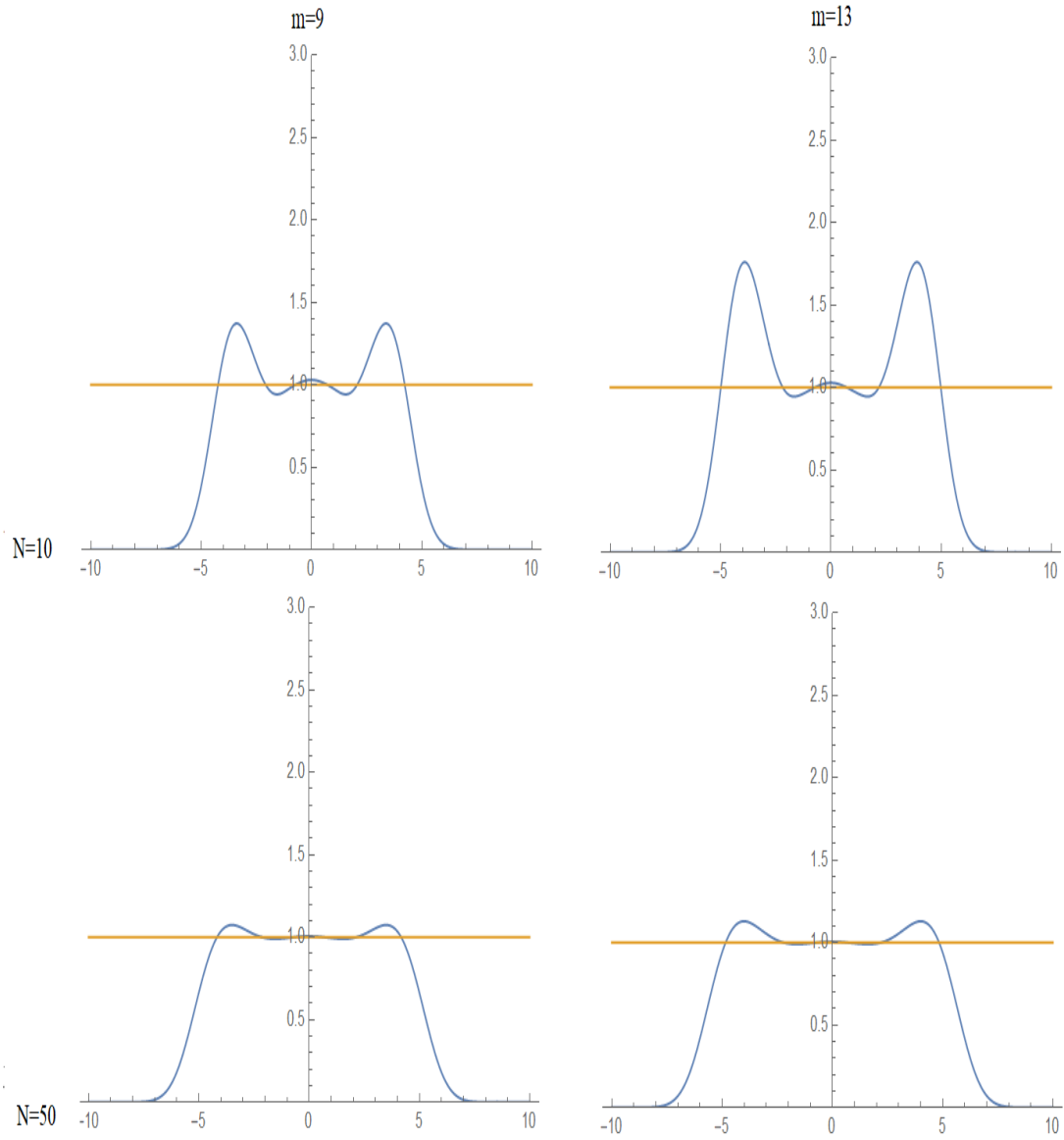


Figure 2.6: The distribution correction factor.

2.10.3. Convex Combination and Convex Optimization

In this section, we implement the Gradient Ascent algorithm in our optimization problem. We want to find optimal coefficients α_m such that the entropy of $\hat{Q}(x)$ is maximized, where $\hat{Q}(x) = \sum_{m=9, \text{ odd}}^{35} \alpha_m q_m(x)$. Figure 2.7 shows the graphical representation of the coefficient α_m of the corresponding mixing densities with odd numbers of m from 9 to 35 when fixing $N = 20$. It is observed that the optimal

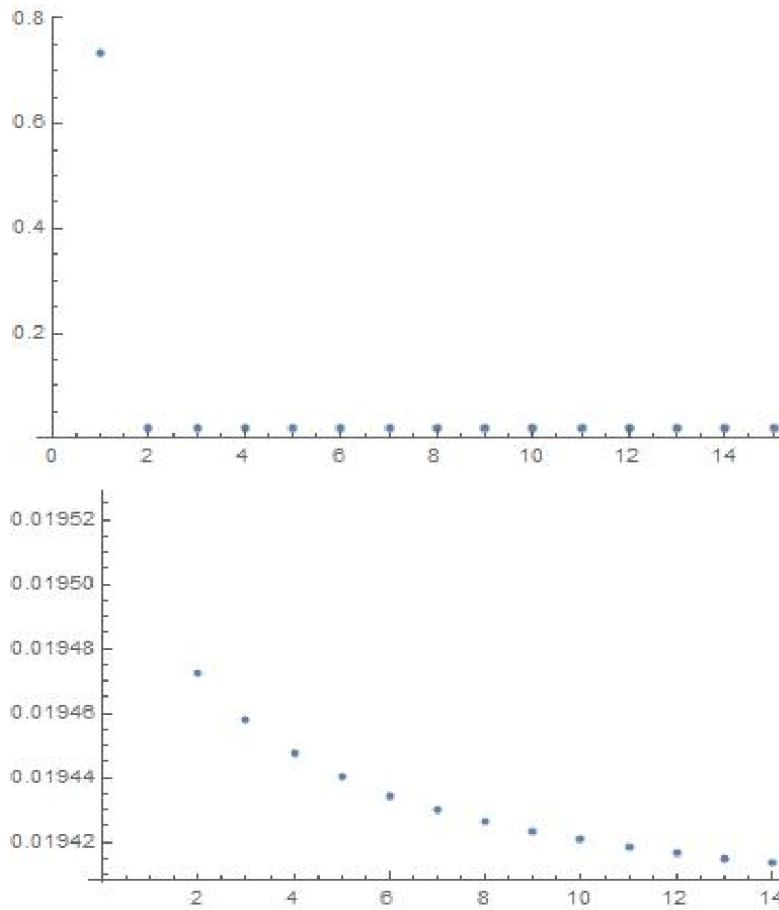


Figure 2.7: The plot of optimal coefficients when $N = 20$. Top: optimal coefficients α_m with m odd numbers from 9 to 35. Bottom: optimal coefficients α_m with odd numbers of m from 11 to 35.

coefficient $\alpha_9 = 0.747382$ while all other coefficients are decreasing from 0.01948. The entropy after optimization is obtained at 1.41878, compared with the entropy at 1.41877 when $N = 20$ and $m = 9$. There is possibly an in-significant improvement in the entropy of 10^{-5} .

2.10.4. Pricing European Options

We will present the numerical results of some experiments to price European call and put options and compare the MC-Tree method with the usage of the bias-correction and the distribution correction factor with the Monte Carlo (MC) method and popular binomial tree models. We can obtain the analytical solution from the well-known Black-Scholes model. The error is the difference between the model value and the analytical solution.

The following parameters are used throughout these numerical experiments.

- Initial stock prices $S_0 = 100$ (if not stated otherwise).
- Strike price $K = 95$.
- Expiration $T = 1$.
- Risk-free rate $r = 0.03$.
- Volatility $\sigma = 0.2$.
- MC-Drawing $M = 10^5$.

It is verified that the put-call parity holds for the MC-Tree with the bias-correction, as shown in Figure 2.8. Given various stock prices $S(t)$ from 1 up to 200 on the

x-axis, the corresponding values on the y-axis are zero. The MC is not involved in the experiment, so confidence intervals are not given.

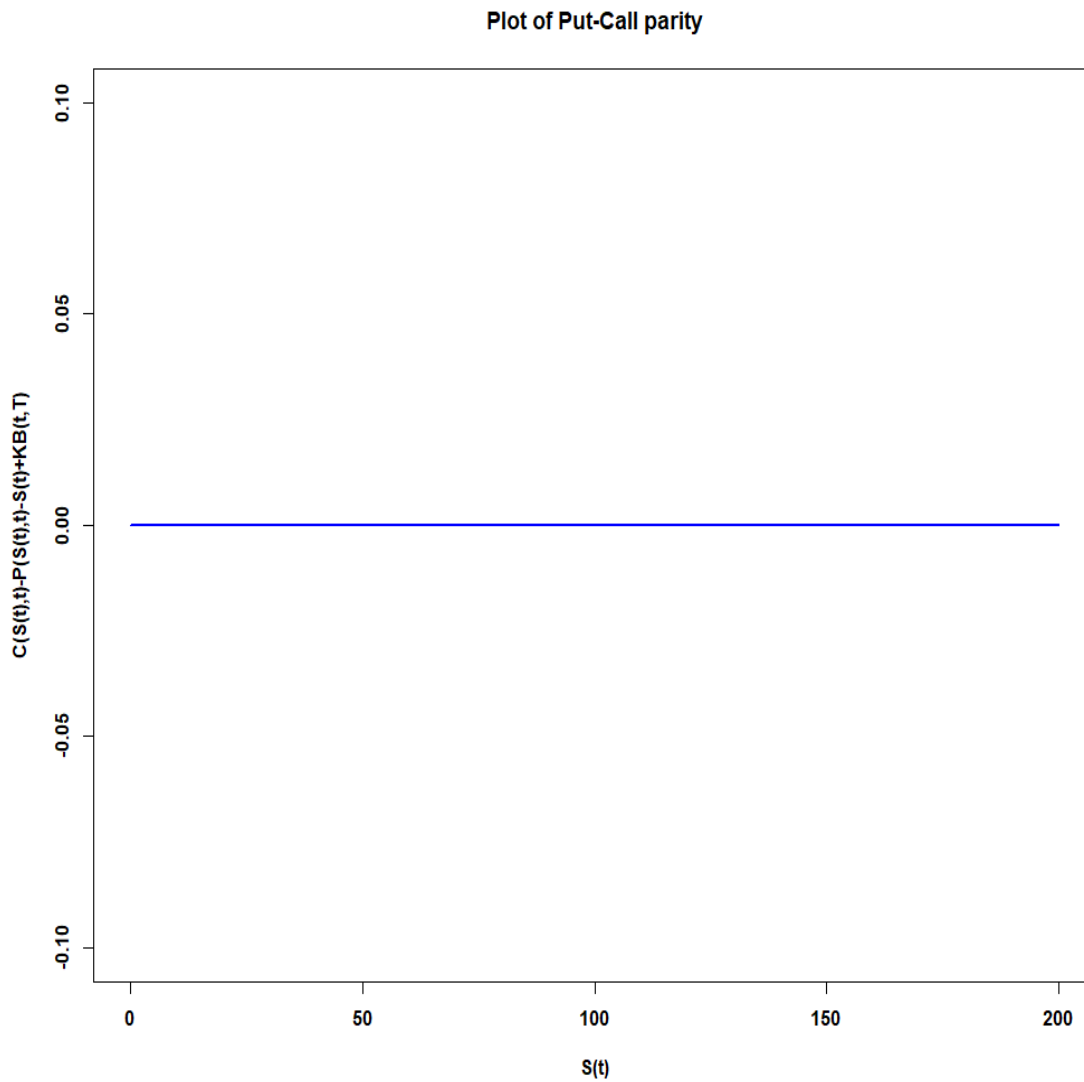


Figure 2.8: Verification of put-call parity. The y-axis represents $C(S(t), t) - P(S(t), t) - S(t) + KB(t, T)$, where $C(S(t), t)$, $P(S(t), t)$ are the call and put option prices, and $B(t, T) = e^{-r(T-t)}$.

Table 2.3: Accuracy comparison between MC-Tree and MC method in pricing European call option. Bias: MC-Tree with the usage of bias-correction. Corr: MC-Tree with the usage of distribution correction factor. SD is the estimate of the standard deviation of the random variable of which the MC estimates the mean.

S	N	Method	Mean	SD	CI	AS
100	50	Corr	12.1798	0.025	(12.17965, 12.17995)	12.1797
	50	Bias	12.1905	0.0279	(12.1903, 12.1907)	
	100	Corr	12.1797	0.0123	(12.17962, 12.17978)	
	100	Bias	12.1851	0.0155	(12.1850, 12.1852)	
	100	MC	12.1867	15.6215	(12.0899, 12.2835)	
95	50	Corr	8.9429	0.0431	(8.9426, 8.9431)	8.9427
	50	Bias	8.9549	0.0323	(8.9547, 8.9551)	
	100	Corr	8.9429	0.0254	(8.942743, 8.943057)	
	100	Bias	8.9489	0.0188	(8.9488, 8.9490)	
	100	MC	8.9469	13.4153	(8.8637, 9.0300)	
90	50	Corr	6.2125	0.071	(6.2121, 6.2130)	6.2125
	50	Bias	6.2230	0.0596	(6.2226, 6.2233)	
	100	Corr	6.2125	0.0463	(6.212213, 6.212787)	
	100	Bias	6.2177	0.0401	(6.2175, 6.2180)	
	100	MC	6.2143	11.1305	(6.1453, 6.2833)	

2.10.4.1. COMPARISON TO PLAIN MC METHOD

Table 2.3 and table 2.4 show that MC-Tree is more accurate than MC method. The usage of the distribution correction factor in option pricing improves accuracy significantly, resulting in the exact analytical solution at $N = 100$.

Table 2.5 shows the results from both methods for similar computation time. It is evident from table 2.5 that the MC-Tree model is still more accurate than the MC method, even with the same computation time.

Table 2.4: Accuracy comparison between MC-Tree and MC method in pricing European put option. Bias: MC-Tree with the usage of bias-correction. Corr: MC-Tree with the usage of distribution correction factor.

S	N	Method	Mean	SD	CI	AS
100	50	Corr	4.3720	0.0324	(4.3718, 4.3722)	4.3720
	50	Bias	4.3828	0.0279	(4.3827, 4.3830)	
	100	Corr	4.3720	0.0185	(4.3719, 4.3721)	
	100	Bias	4.3774	0.0155	(4.3773, 4.3775)	
		MC	4.4107	7.6584	(4.3633, 4.4582)	
95	50	Corr	6.1352	0.0278	(6.135, 6.1353)	6.1351
	50	Bias	6.1473	0.0323	(6.1471, 6.1475)	
	100	Corr	6.1352	0.0155	(6.1351, 6.1353)	
	100	Bias	6.1412	0.0188	(6.1411, 6.1413)	
		MC	6.1693	8.9348	(6.1139, 6.2247)	
90	50	Corr	8.4048	0.0503	(8.4045, 8.4051)	8.4048
	50	Bias	8.4153	0.0596	(8.4149, 8.4157)	
	100	Corr	8.4048	0.0345	(8.4046, 8.4050)	
	100	Bias	8.4101	0.0401	(8.4098, 8.4103)	
		MC	8.4352	10.1748	(8.3721, 8.4982)	

2.10.4.2. COMPARISON TO BINOMIAL MODELS

Table 2.6 and table 2.7 show that MC-Tree with the usage of distribution correction factor performs best. Option price from MC-Tree converges quicker than binomial models to the analytical price when increasing the tree depth. It is evident from Figure 2.9 and Figure 2.10 that CRR and JR model is less stable and more volatile than the MC-Tree model as the tree depth increases. Both figures contain the confidence interval (CI) of the MC-Tree method *without* the distribution correction technique.

Then, we can obtain the mean squared error (MSE) for different models from the range of tree depth from 1 to 100 in Table 2.8. MSE from the MC-Tree is the lowest among models.

Table 2.5: Error and computation time of European call option with similar running time at $N = 50$. MC-Tree Corr: MC-Tree with the distribution correction factor. SE denotes the standard error.

	MC-Tree Corr	MC
Option Price	12.1798	12.1800
SE	0.00008	0.00451
CI	(12.17965, 12.17995)	(12.17117, 12.18883)
Error	5.7e-05	3e-04
Computation Time (Seconds)	34.92103	35.50149
M	100,000	12,000,000

Figure 2.11 plots the European call prices with a range of M and the tree depth from 20 to 200. Clearly, the price is more stable with a rather large M .

2.10.5. Pricing American Put Option

We will present the numerical results of some experiments with the American Put option and compare them with the LSM and popular binomial tree models to gain some insight into the performance of the MC-Tree Method. A quadratic polynomial is used in the regression model (see [35]).

All numerical experiments use the same parameters as mentioned in the previous section: "pricing European option", except $M = 2000$. We consider various examples of American option valuation and compare our method with the LSM, CRR, and JR in the next sections.

Table 2.6: Accuracy comparison between MC-Tree and binomial models in pricing European call option. Bias: MC-Tree with the usage of bias-correction. Corr: MC-Tree with the usage of distribution correction factor.

S	Method	$N = 50$			$N = 100$			AS
		Mean	SD	CI	Mean	SD	CI	
100	Corr	12.1798	0.025	(12.1797, 12.1800)	12.1797	0.0123	(12.1796, 12.1798)	12.1797
	Bias	12.1905	0.0279	(12.1903, 12.1907)	12.1851	0.0155	(12.1850, 12.1852)	
	CRR	12.1733			12.1923			
	JR	12.1677			12.1984			
95	Corr	8.9429	0.0431	(8.9426, 8.9431)	8.9429	0.0254	(8.9427, 8.9431)	8.9427
	Bias	8.9549	0.0323	(8.9547, 8.9551)	8.9489	0.0188	(8.9488, 8.9490)	
	CRR	8.9102			8.9265			
	JR	8.9513			8.9533			
90	Corr	6.2125	0.071	(6.2121, 6.2130)	6.2125	0.0463	(6.2122, 6.2128)	6.2125
	Bias	6.2230	0.0596	(6.2226, 6.2233)	6.2177	0.0401	(6.2175, 6.2180)	
	CRR	6.1912			6.2283			
	JR	6.2281			6.2084			

Table 2.7: Accuracy comparison between MC-Tree and binomial models in pricing European put option. Bias: MC-Tree with the usage of bias-correction. Corr: MC-Tree with the usage of distribution correction factor.

S	Method	$N = 50$			$N = 100$			AS
		Mean	SD	CI	Mean	SD	CI	
100	Corr	4.3720	0.0324	(4.3718, 4.3722)	4.3720	0.0185	(4.3719, 4.3721)	4.3720
	Bias	4.3828	0.0279	(4.3827, 4.3830)	4.3774	0.0155	(4.3773, 4.3775)	
	CRR	4.3657			4.3846			
	JR	4.3600			4.3907			
95	Corr	6.1352	0.0278	(6.135, 6.1353)	6.1352	0.0155	(6.1351, 6.1353)	6.1351
	Bias	6.1473	0.0323	(6.1471, 6.1475)	6.1412	0.0188	(6.1411, 6.1413)	
	CRR	6.1025			6.1188			
	JR	6.1437			6.1456			
90	Corr	8.4048	0.0503	(8.4045, 8.4051)	8.4048	0.0345	(8.4046, 8.4050)	8.4048
	Bias	8.4153	0.0596	(8.4149, 8.4157)	8.4101	0.0401	(8.4098, 8.4103)	
	CRR	8.3835			8.4206			
	JR	8.4204			8.4008			

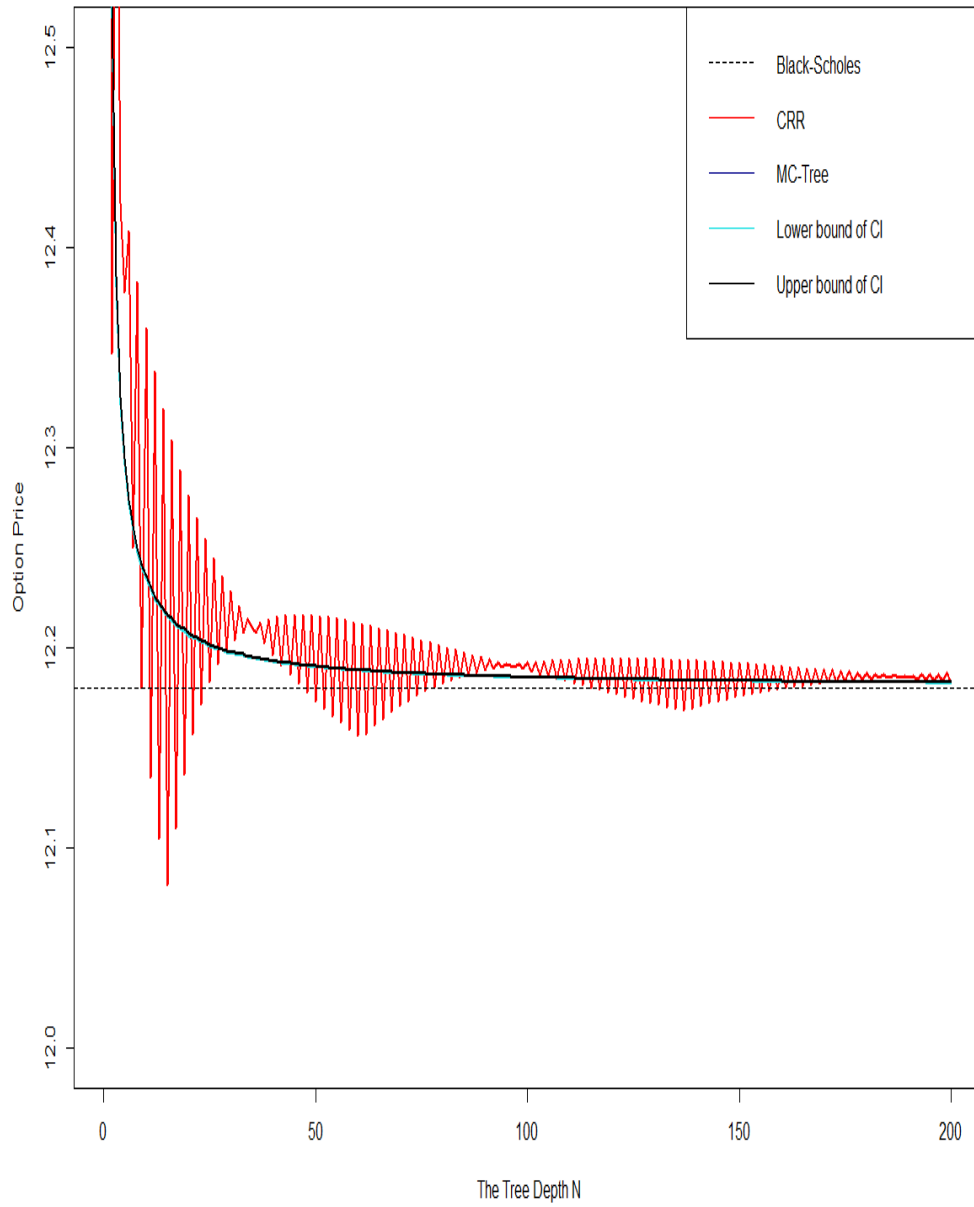


Figure 2.9: Option prices versus the tree depth N .

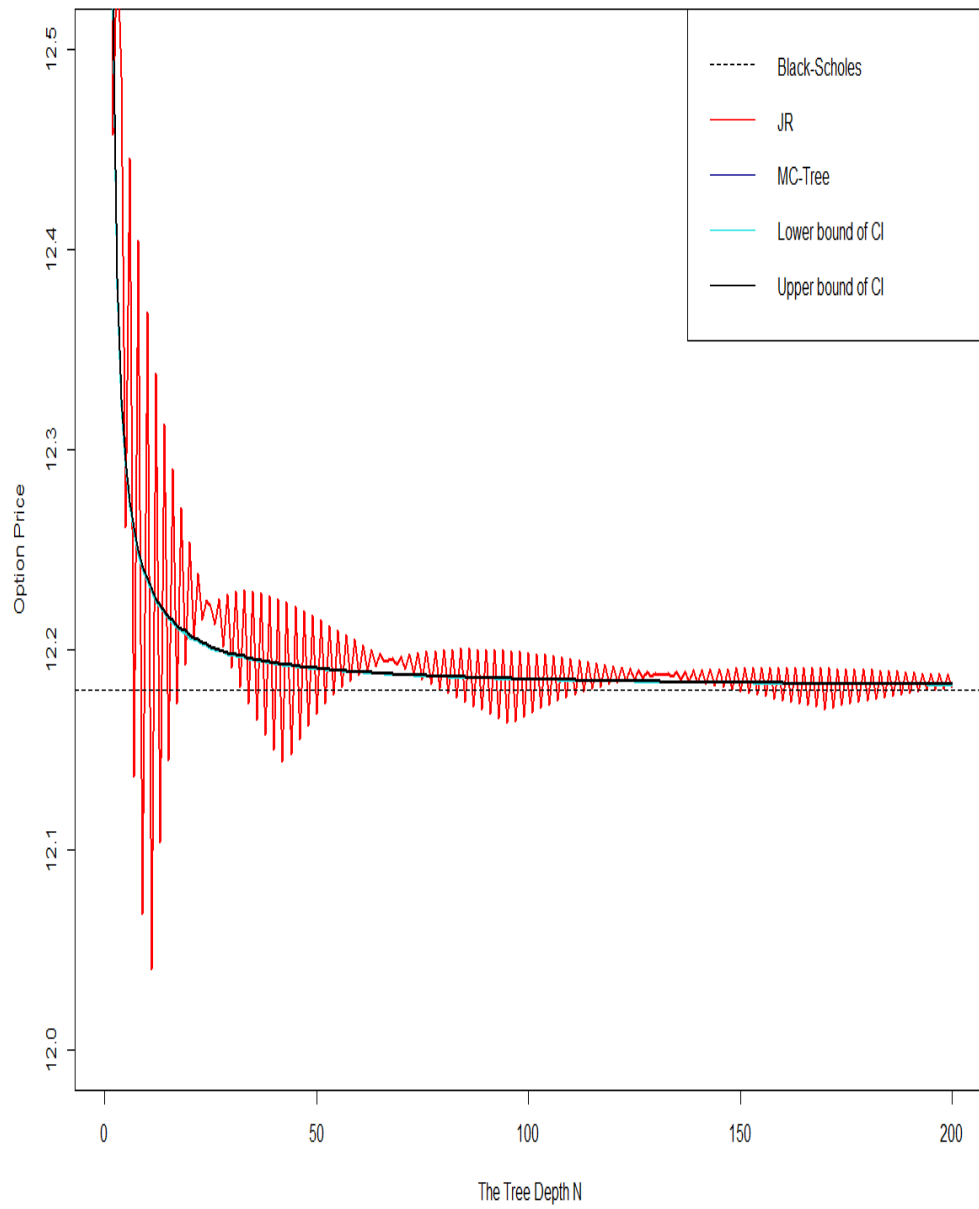
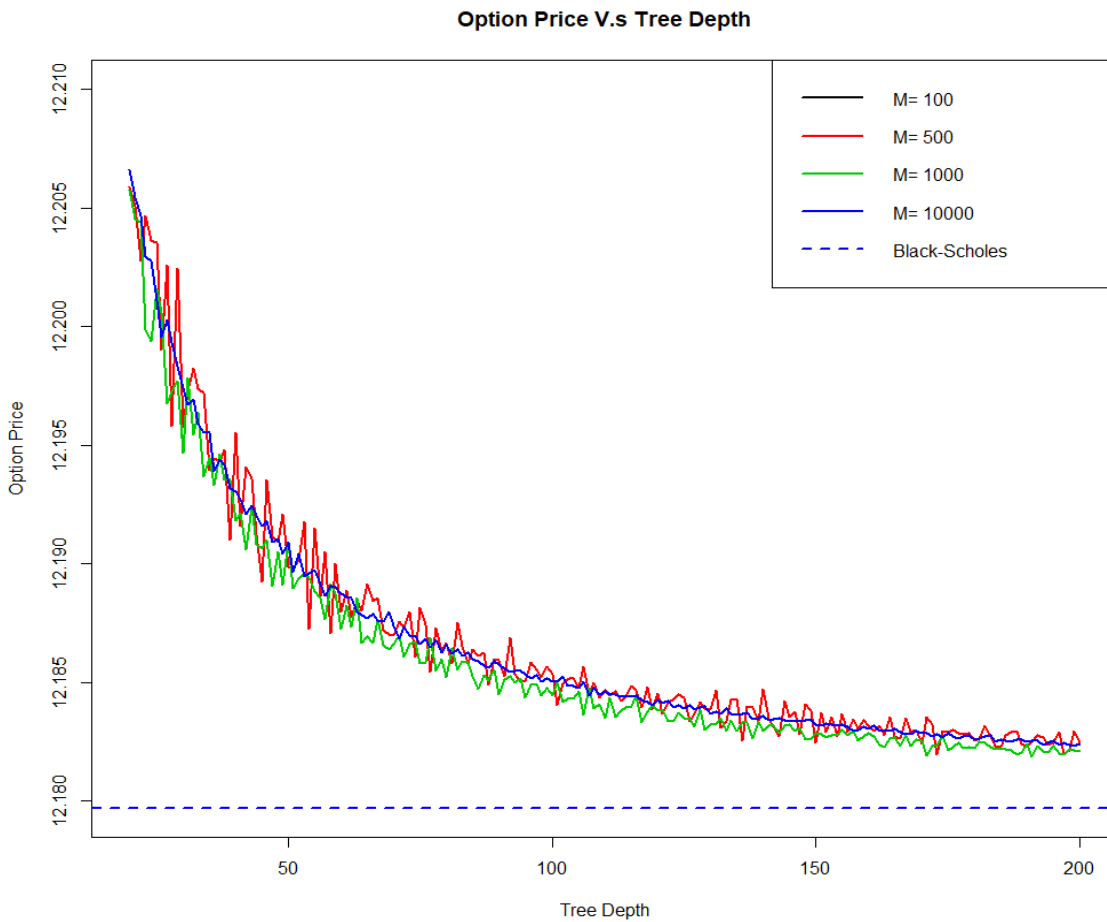


Figure 2.10: Option prices versus the tree depth N .

Table 2.8: Mean-squared error for various models.

Models	MC-Tree	CRR	JR
MSE	0.00015354	0.001074532	0.001015309

**Figure 2.11:** European call prices versus number of MC drawings M and the tree depth N .

2.10.5.1. COMPARISON TO LSM METHOD

We will compare the standard deviation of the MC-Tree and LSM method to understand their accuracy. Table 2.9 shows the mean and the standard deviation from

simulation results with the initial stock price at 100, and the "true" price at 4.5415. The "true" price of an American put option is obtained by the convergent binomial method with the depth of tree at 50,000.

As shown in the table 2.9, the standard deviation of the LSM is much larger than

Table 2.9: Mean and standard deviation of American put option.

	MC-Tree	LSM
Mean	4.5483	4.5782
Standard Deviation	0.0319	7.1828

the one of the MC-Tree method when their means are similar. The LSM needs a significant increase in the number of replications to improve accuracy, which leads to an increase in the computation time per simulation.

It is evident from Table 2.10 that it is not sufficiently good for the LSM to obtain a

Table 2.10: Mean and standard error (SE) of American put Option with similar running time.

	MC-Tree	LSM
Mean	4.5483	4.5274
SE	0.00071	0.01893
Error	0.0068	0.0141
Computation Time (Seconds)	10.4658	10.5423
M	2000	120,000

small standard error as the MC-Tree method, even though the computation time of both are almost the same. It is concluded that the MC-Tree method provides us with more accuracy than the LSM at a similar computational cost.

2.10.5.2. COMPARISON TO BINOMIAL MODELS

We will use the same model parameters, as mentioned in the previous section. As shown in Table 2.11 and Table 2.12, the option prices among models are insignificantly different, and the MC-Tree method produces the smallest error. It can be concluded that the MC-Tree method performs better than other methods: CRR, JR in terms of accuracy using the same tree-depth. Researchers in [1] proved that American option prices of the discrete model also converge to the corresponding value of the continuous-time model under fairly general conditions. It means that the "true" price can be obtained by increasing the tree depth to infinity. Therefore, we also compare results with the "true" prices. Figure 2.12 shows two plots for prices of an

Table 2.11: Option Prices

Stock Price	MC-Tree	CRR	JR	"True" Price
95	6.4140	6.3966	6.4141	6.4058
97	5.6058	5.6148	5.6080	5.5973
100	4.5484	4.5511	4.5583	4.5415
102	3.9409	3.9433	3.9240	3.9338
104	3.4007	3.4034	3.4111	3.3960

Table 2.12: Accuracy comparison among models.

Stock Price	MC-Tree	CRR	JR
95	0.0081	0.0093	0.0083
97	0.0084	0.0175	0.0107
100	0.0080	0.0096	0.0168
102	0.0070	0.0095	0.0098
104	0.0047	0.0073	0.0150

American put option by the MC-Tree, CRR, and JR model when increasing the tree depth. The CRR and JR model both are more volatile than the MC-Tree model. The MC-Tree model is more stable and convergent toward the "true" price as N

increases, compared with the CRR and JR model.

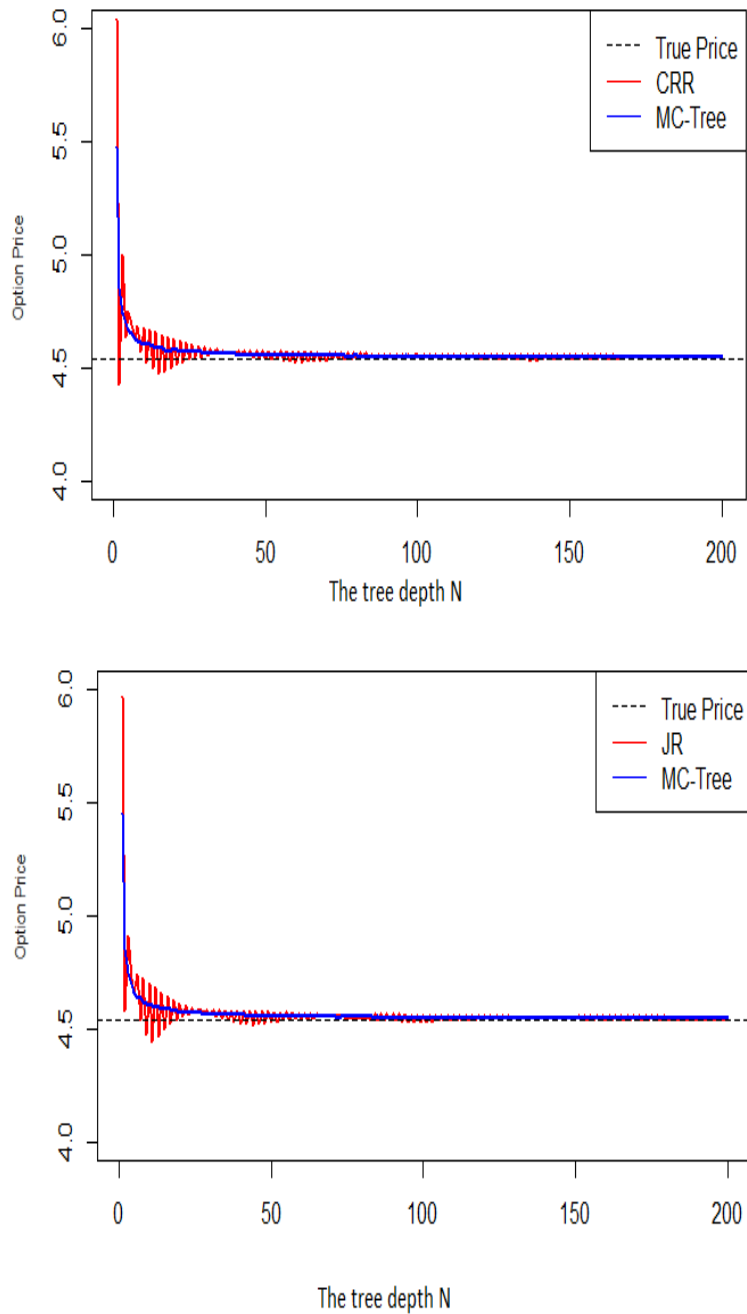


Figure 2.12: Option prices versus the tree depth N .

2.10.5.3. THE OPTIMAL N AND M OF MC-TREE

We also show the range of N and the corresponding M , in which the MC-Tree method performs high accuracy and reasonable efficiency. We found that computation time is approximately proportional to the number of nodes at $\frac{1}{2}N(N+1)$ of an N -step binomial tree. Hence, if we fix the computation time, the optimal choice of the tree depth, N , and the number of simulations, M , are found in the relation $M = \frac{C}{N(N+1)}$, where C is chosen approximately.

We consider an experiment on an American put option with the same input parameters, as mentioned in the previous section.

Figure 2.13 shows that MC-Tree absolute error remains around 0.01 when the

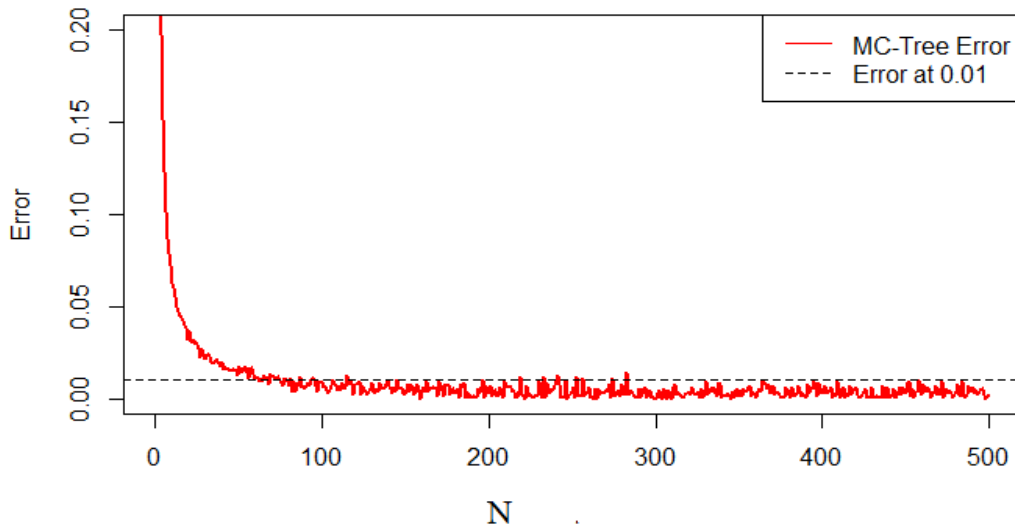


Figure 2.13: Error versus the depth of tree.

computation time is below 2 seconds, N is in the range from around 70 to 500, and M is an integer part of $\frac{2500000}{N(N+1)}$.

Chapter 3

Pricing Multi-Asset Options Using An MC-Tree Method

3.1. Introduction

In chapter two, we presented the theory of MC-Tree that combines MC with the recombining binomial tree based on the Pascal's triangle for pricing single asset options. In this chapter, we generalize the theory of MC-Tree that combines the MC method with the recombining multinomial tree based on Pascal simplex for pricing multi-asset option. In d assets case, we have a probability vector $(p_1, p_2, \dots, p_{d+1})$. Our approach includes applying MC drawings from the Haar measure on the orthogonal group, based on the tree construction in the paper by Sierag and Hanzon in [40]. We will also draw an orthogonal $d \times d$ matrix with each drawing of $(p_1, p_2, \dots, p_{d+1})$, $p_j \geq 0$, $\sum_{j=1}^{d+1} p_j = 1$, where d is the number of risky assets. The orthogonal matrix is generated by drawing the d^2 elements from a $d \times d$ matrix with

d^2 i.i.d standard Gaussian random variables and then apply the Gram-Schmidt orthogonalization procedure on d columns of the matrix. This matrix is denoted by H . The determinant of this orthogonal matrix receives both values -1 and 1.

We will describe the kind of mixing densities for consideration, the role of the radius, compound density entropy, and the derivation of the corresponding compound density analytically and numerically.

3.2. Mixing Density and MC Drawings

The point is that we do mixing density on the tree parameters: probability vector and the direction vectors. For each probability vector, we pick the corresponding direction vectors, then we apply the orthogonal group to rotate and reflect around. We obtain the uniform measure on the orthogonal group. This gives mixing densities on the tree. For each mixing distribution on the vector $(p_1, p_2, \dots, p_{d+1})$, our approach gives a compound density at time step N which is invariant under application of the orthogonal group to the target space \mathbf{R}^d ; i.e, it is invariant under rotations and reflections in the higher dimensional Euclidean space in the multi-asset case ($d \geq 2$). After we have taken the orthogonal group action by MC generation, we need to work out the compound density of the radii of the vectors occurring after N steps in the tree when applying a mixing distribution.

We use the symmetric mixing density because we know that this leads to better or equal entropy value of compound density in relation to application of rotations and reflections to the plane in the 2-risky asset case ($d = 2$) and to higher dimensional Euclidean space in multi-asset case ($d > 2$). In this thesis, we will choose to work

with polynomial mixing density in relation to the parameters p_1, p_2, \dots, p_{d+1} , for which we can describe a great result. We can restrict to symmetric polynomials; $P(\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_{d+1}})$, say. For example, the symmetric polynomials are those for which

$$P(\sqrt{p_1}, \sqrt{p_2}) = P(\sqrt{p_2}, \sqrt{p_1})$$

in case of $d = 1$;

$$\begin{aligned} P(\sqrt{p_1}, \sqrt{p_2}, \sqrt{p_3}) &= P(\sqrt{p_1}, \sqrt{p_3}, \sqrt{p_2}) \\ &= P(\sqrt{p_2}, \sqrt{p_1}, \sqrt{p_3}) \\ &= P(\sqrt{p_2}, \sqrt{p_3}, \sqrt{p_1}) \\ &= P(\sqrt{p_3}, \sqrt{p_1}, \sqrt{p_2}) \\ &= P(\sqrt{p_3}, \sqrt{p_2}, \sqrt{p_1}) \end{aligned}$$

in case of $d = 2$.

Our choice of mixing density is

$$c (p_1^{1/2} p_2^{1/2} p_3^{1/2} \dots p_{d+1}^{1/2})^{m-2} dp_1 dp_2 dp_3 \dots dp_d \quad (3.1)$$

for some $m \in \mathbb{N}$.

The square roots of the probabilities can be described by spherical coordinates as follows:

- $p_1^{1/2} = \cos(\theta_1)$,
- $p_2^{1/2} = \sin(\theta_1) \cos(\theta_2)$,
- $p_3^{1/2} = \sin(\theta_1) \sin(\theta_2) \cos(\theta_3)$,
- \vdots
- $p_d^{1/2} = \sin(\theta_1) \dots \sin(\theta_{d-1}) \cos(\theta_d)$,
- $p_{d+1}^{1/2} = \sin(\theta_1) \dots \sin(\theta_{d-1}) \sin(\theta_d)$,
- $0 < \theta_i < \frac{\pi}{2}, i = \overline{1, d}$.

The mixing density can be factorized into univariate factors in which each is a polynomial in terms of $\sin(\theta_i)$, $\cos(\theta_i)$, $i = \overline{1, d}$. As the antiderivative is known for such functions, so we can implement Monte Carlo Drawing, as presented in section 2.6 of the chapter two for each θ_i .

3.3. The Radius

X_k denotes the vector $\sum_{i=1}^{d+1} k_i d_i$ at node k that is reached by taking k_i times direction $d_i, i = 1, \dots, d + 1$, $k = (k_1, k_2, \dots, k_{d+1})$, $\|k\| = N$ at the final step of the tree. The random variable $R_k = \|X_k\|$ is the radius of the vectors at the node k in the tree.

Theorem 3.3.1. *In a two-asset case,*

$$R_k^2 = \|X_k\|^2 = \left\| \sum_{i=1}^3 k_i d_i \right\|^2 = \sum_{i=1}^3 \frac{k_i^2}{p_i} - |k|^2.$$

Proof. The proof is based on the constructions presented in [40].

Vector $X_k = \sum_{i=1}^3 k_i d_i$ is corresponding to the orthogonal projection of $\sum_{i=1}^3 \frac{k_i}{\sqrt{p_i}} e_i$ along the vector

$$\begin{bmatrix} |k| p_1^{1/2} \\ |k| p_2^{1/2} \\ |k| p_3^{1/2} \end{bmatrix},$$

so to the vector $\tilde{X}_k = \sum_{i=1}^3 \frac{k_i}{\sqrt{p_i}} e_i - \sum_{i=1}^3 |k| \sqrt{p_i} e_i$ to be more precise.

The vector X_k is obtained from \tilde{X}_k by an orthogonal mapping (a rotation), so the length of this vector \tilde{X}_k equal to the length of vector X_k . Due to the orthogonality of the projection, we have

$$\begin{aligned} \|\tilde{X}_k\|^2 + \left\| \sum_{i=1}^3 |k| \sqrt{p_i} e_i \right\|^2 &= \left\| \sum_{i=1}^3 \frac{k_i}{\sqrt{p_i}} e_i \right\|^2 \\ \Leftrightarrow \|\tilde{X}_k\|^2 &= \sum_{i=1}^3 \frac{k_i^2}{p_i} - |k|^2. \end{aligned}$$

□

Remarks

- (i) The radius $R_k = \|X_k\|$ of the vector-nodes after $|k|$ steps in the tree has conditional probability distribution, given (p_1, p_2, p_3) : $\binom{|k|}{k} p_1^{k_1} p_2^{k_2} p_3^{k_3}$, where probabilities runs over all possible k , $k = (k_1, k_2, k_3)$, $|k| = k_1 + k_2 + k_3$, $k_i \in \mathbb{N}_0$, $i = 1, 2, 3$. However, note that R_k can be equal for different values of k , in which case the corresponding probabilities will have to be added to obtain the probability distribution.
- (ii) In general, the squared radius $R_k^2 = \|X_k\|^2$ of the vector-nodes after $|k|$ steps in

the tree is defined by

$$R_k^2 = \left\| \sum_{i=1}^{d+1} k_i d_i \right\|^2 = \sum_{i=1}^{d+1} \frac{k_i^2}{p_i} - |k|^2.$$

The radius $R_k = \|X_k\| = \left\| \sum_{i=1}^{d+1} k_i d_i \right\|$ of the vector-nodes after $|k|$ steps in the tree has conditional probability distribution given $(p_1, p_2, \dots, p_{d+1})$

$$\binom{|k|}{k} p_1^{k_1} p_2^{k_2} \dots p_{d+1}^{k_{d+1}},$$

where k runs over all vectors in \mathbb{N}_0^{d+1} , $k = (k_1, k_2, k_3, \dots, k_{d+1})$, $|k| = \sum_{i=1}^{d+1} k_i$, $k_i \in \mathbb{N}_0$, $i = 1, 2, 3, \dots, d+1$; and $|k| = N$. And X_k denotes the random variable with compound density P_{X_k} if a tree of $|k|$ steps is used; $X_k = X_{(k_1, k_2, \dots, k_{d+1})}$ is the vector value at node $(k_1, k_2, \dots, k_{d+1})$ given a choice of $(p_1^{1/2}, p_2^{1/2}, \dots, p_{d+1}^{1/2})$.

3.4. RI Entropy

In this section, we consider a rotation invariant density $g(x) = f(\|x\|)$ on \mathbb{R}^d for an appropriate function f . Function $f(r)$ depends only on the radius $r = \|x\|_2$ of the vector x ; $S(r)$ is the surface of a sphere with radius r in d dimension.

We have $1 = \int_{x \in \mathbb{R}^d} g(x) dx = \int_{x \in \mathbb{R}^d} f(r) dx = \int_0^\infty f(r) S(r) dr$, where $S(r)$ is the surface of a sphere with radius r in d dimension. Hence $f(r) S(r)$ is a pdf. Let $q(r) := f(r) S(r)$, q is a pdf on $[0, \infty)$. Given q , the RI Entropy of q ($RIEnt(q)$) is

equal to the entropy of g ($Ent(g)$).

$$\begin{aligned} Ent(g) &= - \int_{x \in \mathbb{R}^d} g(x) \log[g(x)] dx = - \int_{x \in \mathbb{R}^d} f(r) \log[f(r)] dx \\ &= - \int_0^\infty q(r) \log\left[\frac{q(r)}{S(r)}\right] dr =: RIEnt(q), \end{aligned}$$

where $f(r) = \frac{q(r)}{S(r)} = g(x)$, $x \in \mathbb{R}^d$.

Definition: RI Entropy

Let $X \in \mathbb{R}^d$ have rotation-invariant distribution, and let q be the pdf of the radius of X . Then the RI Entropy of q is defined by

$$RIEnt(q) = - \int_0^\infty q(r) \log\left[\frac{q(r)}{S(r)}\right] dr.$$

Remark: Note that the RI Entropy of q is equal to the entropy of the pdf of X .

3.4.1. RI Entropy in the Two Assets Case

We have

$$\begin{aligned} 1 &= \int_{-\infty}^\infty \int_{-\infty}^\infty g(x, y) dx dy = \int_{-\infty}^\infty \int_{-\infty}^\infty f(r) dx dy \\ &= \int_0^{2\pi} \int_0^\infty f(r) r dr d\theta = \int_0^\infty 2\pi f(r) r dr, \end{aligned}$$

so $2\pi f(r)r$ is a pdf.

Let $q(r) := 2\pi f(r)r$, q is a pdf on $[0, \infty)$.

Given q , we can compute the entropy of g , the RI entropy of q ($RIEnt(q)$), as

$f(r) = \frac{q(r)}{2\pi r} = g(x, y)$, so

$$\begin{aligned} Ent(g) &= - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \log[g(x, y)] dx dy = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(r) \log[f(r)] dx dy \\ &= - \int_0^{2\pi} \int_0^{\infty} f(r) \log[f(r)] r dr d\theta = -2\pi \int_0^{\infty} f(r) \log[f(r)] r dr \\ &= - \int_0^{\infty} q(r) \log\left[\frac{q(r)}{2\pi r}\right] dr = RIEnt(q). \end{aligned}$$

In the standard Gaussian case: $g(x, y) = f(r) = \frac{1}{2\pi} e^{-r^2/2}$, so

$$q(r) = 2\pi r f(r) = r e^{-r^2/2}.$$

Then RI entropy of $q(r)$ is

$$\begin{aligned} RIEnt(q) &= - \int_0^{\infty} r e^{-r^2/2} \log\left[\frac{r e^{-r^2/2}}{2\pi r}\right] dr = - \int_0^{\infty} e^{-r^2/2} \left(-\frac{r^2}{2}\right) d\left(\frac{r^2}{2}\right) + \log(2\pi) \\ &= \log(2\pi) + 1. \end{aligned}$$

Entropy of standard Gaussian $X \sim \mathcal{N}(0, I_2)$ is $\log(2\pi e)$.

The second moment of R is

$$\mathbb{E}[R^2] = \int_0^{\infty} r^2 q(r) dr = \int_0^{\infty} r^3 e^{-r^2/2} dr.$$

We have $\int r^3 e^{-r^2/2} dr = 2 \int v e^v dv$, where $v = -\frac{r^2}{2}$. We have $\int v e^v dv = (v - 1)e^v + C$, by the integration by parts.

Hence,

$$\int r^3 e^{-r^2/2} dr = -(r^2 + 2)e^{-r^2/2} + C.$$

Therefore, $\mathbb{E}[R^2] = 2$.

3.4.2. RI Entropy in the Two Assets Case(Alternative Approach)

We find RI Entropy in the 2-asset case using polar coordinates, as follows.

Let $g(x, y)dxdy = f(r)dxdy = f(r)rdrd\theta$, where $r = \sqrt{x^2 + y^2}$, then

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\sqrt{x^2 + y^2}) \log(f(\sqrt{x^2 + y^2}))dxdy &= \int_{r=0}^{\infty} \int_{\theta=0}^{2\pi} f(r) \log(f(r))rdrd\theta \\ &= 2\pi \int_{r=0}^{\infty} f(r) \log(f(r))rdr \\ &= 2\pi \int_{\rho=0}^{\infty} f(\sqrt{2\rho}) \log(f(\sqrt{2\rho}))d\rho, \end{aligned}$$

where $\rho = \frac{1}{2}r^2$. Hence, the pdf of the radius is $2\pi f(r)r$. If we use $\rho = \frac{1}{2}r^2$ as a new coordinate instead of r ,

$$\begin{aligned} x &= \sqrt{2\rho}\cos(\theta), \\ y &= \sqrt{2\rho}\sin(\theta) \end{aligned}$$

and $d\rho = rdr$, so $g(x, y)dxdy = f(\sqrt{2\rho})dxdy = f(\sqrt{2\rho})rdrd\theta = f(\sqrt{2\rho})d\rho d\theta$.

Let $h(\rho) := f(\sqrt{2\rho}) = g(x, y)$ then we have $g(x, y)dxdy = h(\rho)d\rho d\theta$. It implies that the pdf of P is $q(\rho) = 2\pi h(\rho)d\rho = e^{-\rho}$.

Following is for the standard Gaussian case.

The entropy for the standard bivariate Gaussian density $X \sim \mathbb{N}(0, I_2)$ is

$$\begin{aligned} - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \log(g(x, y))dxdy &= - \int_{\theta=0}^{2\pi} \int_{r=0}^{\infty} h(\rho) \log(h(\rho))rdrd\theta \\ &= -2\pi \int_0^{\infty} h(\rho) \log(h(\rho))d\left(\frac{1}{2}r^2\right) \\ &= -2\pi \int_0^{\infty} h(\rho) \log(h(\rho))d\rho. \end{aligned}$$

As we see, up to a factor 2π , the entropy of $h(\rho)$ is the same as the entropy of g , and $h(\rho) = g(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} = \frac{1}{2\pi} e^{-\rho}$. Hence,

$$\begin{aligned} -2\pi \int_0^\infty h(\rho) \log(h(\rho)) d\rho &= -2\pi \int_0^\infty \frac{1}{2\pi} e^{-\rho} \log\left(\frac{1}{2\pi} e^{-\rho}\right) d\rho \\ &= \log(2\pi) + 1 = \log(2\pi e). \end{aligned}$$

We used $\int_0^\infty \rho e^{-\rho} d\rho = 1$. It is noticeable that the pdf of random variable P is $q(\rho) = e^{-\rho} d\rho$, and $\int_0^\infty \rho^2 e^{-\rho} d\rho = 2$ so $\mathbb{E}(P^2) = 2$. Hence, the random variable P has the mean at $\mathbb{E}(P) = \int_0^\infty \rho e^{-\rho} d\rho = 1$, and the variance $Var(P) = \mathbb{E}(P^2) - [\mathbb{E}(P)]^2 = 1$. As well-known in the literature, the exponential distribution with mean 1 is the distribution with the maximal entropy at 1 among all distribution on $(0, \infty)$ with mean at 1.

We can conclude that maximizing the entropy over the rotation-and reflection-invariant densities, conditional on the covariance matrix= I_2 is equivalent to maximizing the entropy over all densities for P on $[0, \infty)$, where $P = \frac{R^2}{2N}$, conditional on mean at 1.

Let $k(r)$ denote the radius-density, i.e, the pdf of R , then the relation between the radius-density of the radius R and the density of P is

$$k(r) = f(r)r = f(\sqrt{2\rho})\sqrt{2\rho} = h(\rho)\sqrt{2\rho} \Leftrightarrow h(\rho) = \frac{k(\sqrt{2\rho})}{\sqrt{2\rho}}.$$

We have

$$dr = \frac{1}{r} d\rho = \frac{1}{\sqrt{2\rho}} d\rho,$$

so

$$h(\rho) d\rho = k(r) dr.$$

3.4.3. RI Entropy in the Three Assets Case

We have

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y, z) dx dy dz = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(r) dx dy dz \\ &= \int_0^{\infty} f(r) 4\pi r^2 dr, \end{aligned}$$

so $4\pi r^2 f(r)$ is a pdf.

Let $q(r) := 4\pi r^2 f(r)$, q is a pdf on $[0, \infty)$.

Given q , we can compute the RI entropy of q ($RIEnt(q)$), as follows:

$$f(r) = \frac{q(r)}{4\pi r^2} = g(x, y, z),$$

so $-\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y, z) \log[g(x, y, z)] dx dy dz = -\int_0^{\infty} q(r) \log[\frac{q(r)}{4\pi r^2}] dr = RIEnt(q)$.

In the standard Gaussian case in three dimensions: $g(x, y, z) = f(r) = \frac{1}{\sqrt{(2\pi)^3}} e^{-r^2/2}$,

so $q(r) = 4\pi r^2 f(r) = \sqrt{\frac{2}{\pi}} r^2 e^{-r^2/2}$.

Then the RI entropy of $q(r)$ is

$$\begin{aligned} RIEnt(q) &= -\int_0^{\infty} \sqrt{\frac{2}{\pi}} r^2 e^{-r^2/2} \log\left[\frac{1}{\sqrt{(2\pi)^3}} e^{-r^2/2}\right] dr \\ &= \frac{3}{2} \sqrt{\frac{2}{\pi}} \log(2\pi) \int_0^{\infty} r^2 e^{-r^2/2} dr + \sqrt{\frac{2}{\pi}} \int_0^{\infty} \frac{r^4}{2} e^{-r^2/2} dr \\ &= \frac{3}{2} \sqrt{\frac{2}{\pi}} \log(2\pi) \sqrt{\frac{\pi}{2}} + \sqrt{\frac{2}{\pi}} \frac{3\sqrt{\pi}}{\sqrt{8}} = \frac{3}{2} \log(2\pi e). \end{aligned}$$

The entropy of a standard Gaussian $X \sim \mathcal{N}(0, I_3)$ is $\frac{3}{2} \log(2\pi e)$.

The second moment of R is

$$\mathbb{E}[R^2] = \int_0^\infty r^2 q(r) dr = \int_0^\infty \sqrt{\frac{2}{\pi}} r^4 e^{-r^2/2} dr.$$

We have

$$\int_0^\infty r^4 e^{-r^2/2} dr = 3\sqrt{2} \frac{\sqrt{\pi}}{2} = 3\sqrt{\frac{\pi}{2}}.$$

Therefore, the second moment of R is $\mathbb{E}[R^2] = 3$.

3.5. Compound Entropy Maximization

We want to find the entropy of the compound density which is \mathbb{O}_d invariant in terms of the marginal pdf $P_{R_N}(r)$ of the radius $R_k = \|X_k\|$. Here \mathbb{O}_d denotes the set of dx d orthogonal matrices. We consider the general case of a compound density of the form $P_{X_N}(x) = P_{X_N}(x_1, x_2, \dots, x_d) = f_N(r)$, for some function f_N , where $R = \|X\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_d^2}$ denotes the radius of a point in \mathbb{R}^d . After the way we set it up, the compound density function in the form $f_N(r)$ only depends on the radius R on \mathbb{R}^d space, where $R \geq 0$.

Proposition 3.5.1. *$f(r)r^{d-1}$ is equal to the (marginal) pdf of $R_N = \|X_N\|_2$ up to normalization by a constant factor.*

Proof. The pdf is obtained by computing the marginal density of R_N . Let the space \mathbb{R}^d be parametrized by a pair (R_N, v) , where $v \in S^{d-1} = \{\tilde{v} \in \mathbb{R}^d : \|\tilde{v}\| = 1\}$, the unit sphere in \mathbb{R}^d . We can write the marginal density of R_N as $P_{R_N}(r)dr = \int_{v \in rS^{d-1}} f(r) d\mu_r(v) dr = \int_{v \in S^{d-1}} f_N(r) r^{d-1} d\mu_1(v) dr = f(r) r^{d-1} dr * \text{vol}(S^{d-1})$, where $\mu_r(v)$ is the standard measure on the sphere with radius r . Clearly, the measure scales with the radius according to the rule $\mu_r(rV) = r^{d-1} \mu_1(v)$ if $V \subset S^d$ is any

measurable subset of S^d . Hence rV is a measurable subset of rS^d , and $vol(S^{d-1})$ denotes the volume (area in case $d = 3$, length in case $d = 2$) of the unit sphere in \mathbb{R}^d . Now note that $vol(S^{d-1})f(r)r^{d-1}dr = vol(S^{d-1})f(r)\frac{1}{d}d(r^d)$, so the pdf of R_N is actually $\frac{vol(S^{d-1})}{d}f(r^d)$, which is equal to $f(r^d)$ up to a normalization constant. \square

As the corollary of Proposition 3.5.1, maximizing compound density entropy boils down to maximizing the RI Entropy of R_N given that

$$\mathbb{E}[R_N^2] = \mathbb{E}[\text{trace}X_N X_N'] = \text{trace}(\mathbb{E}X_N X_N') = \text{trace}(NI_d) = Nd.$$

The objective is now to maximize the entropy of compound density

$\int_0^\infty f(r)r^{d-1} \log(f(r))dr$, where $f(r)r^{d-1}$ is the pdf of $R_N = \|X_N\|_2$. The entropy is considered up to an increasing affine transformation, i.e. up to an additive constant and a positive constant factor as this does not affect the maximization.

It is well-known in the literature that the Gaussian density maximizes the entropy integral under the constraints that the mean is zero vector and the covariance matrix is NI_d . As is well known, symmetrizing a density in this way does not decrease the entropy. We will therefore consider only such symmetrized densities. The entropy of symmetrized density is $\int_{T \in \mathbf{O}_d} P_X(Tx)d\mu(T)$, where \mathbf{O}_d denotes the group of $d \times d$ orthogonal matrices, and μ denotes the Haar measure on the orthogonal group \mathbf{O}_d . In a d assets case, we can conclude that maximizing the entropy over the rotation-and reflection-invariant densities, conditional on the mean at the zero vector and the covariance matrix at I_d is equivalent to maximizing the so-called RI Entropy over all densities for R on $[0, \infty)$, where $R = \|X\|_2$, conditional on $\mathbb{E}(R^2) = d$.

3.6. Analytic Compound Density

In this section, we present our approach for finding the analytic compound density in a two asset case. Let the mixing density on the vector $(\sqrt{p_1}, \sqrt{p_2}, \sqrt{p_3})$ be given by formula (3.1) with $d = 2$.

Theorem 3.6.1. *In a two asset case, the compound density of the radius can be written as*

$$q(r) = \sum_{k_3=0}^N \int_0^1 2q_1(\sqrt{A(p_3)r^2 + B(p_3, k_3)})K(p_3, k_3) \frac{A(p_3)r}{\sqrt{A(p_3)r^2 + B(p_3, k_3)}} \rho(p_3) dp_3, \quad (3.2)$$

where

$$\begin{aligned} A(p_3) &:= 1 - p_3, \\ B(p_3, k_3) &:= -(\sqrt{p_3}N - \frac{k_3}{\sqrt{p_3}})^2, \\ \rho(p_3) &:= \hat{c}_m(1 - p_3)^{m-1} p_3^{\frac{m-2}{2}} dp_3, \\ \hat{c}_m &= \frac{1}{\int_0^1 (1 - p_3)^{m-1} p_3^{\frac{m-2}{2}} dp_3}, \\ K(p_3, k_3) &:= \binom{N}{k_3} (1 - p_3)^{N-k_3} p_3^{k_3}. \end{aligned}$$

The function $q_1(x)$ has the form $q_1(x) = c_m \frac{A_m(x)}{(x^2 + (N - k_3)^2)^{N - k_3 + m}}$, where $A_m(x)$ is a polynomial with rational coefficients and even degree at most $2(N - k_3 + m - 1)$.

Proof. By Proposition 3.3.1, we have

$$R^2 = \frac{k_1^2}{p_1} + \frac{k_2^2}{p_2} + \frac{k_3^2}{p_3} - (k_1 + k_2 + k_3)^2$$

$$\Leftrightarrow R^2 - \frac{k_3^2}{p_3} + k_3^2 + 2k_3(k_1 + k_2) = \frac{k_1^2}{p_1} + \frac{k_2^2}{p_2} - (k_1 + k_2)^2.$$

Recall $p_1 > 0$, $p_2 > 0$, $p_3 > 0$ and $p_1 + p_2 + p_3 = 1$. Let $N = |k|$. Define $\tilde{p}_1 := \frac{p_1}{1-p_3}$, $\tilde{p}_2 := \frac{p_2}{1-p_3}$, we have

$$(1-p_3)\left(R^2 - \frac{k_3^2}{p_3} + k_3^2 + 2k_3(k_1 + k_2)\right) = \frac{k_1^2}{\tilde{p}_1} + \frac{k_2^2}{\tilde{p}_2} - (k_1 + k_2)^2(1-p_3)$$

$$\Leftrightarrow (1-p_3)\left(R^2 - \frac{k_3^2}{p_3} + k_3^2 + 2k_3(k_1 + k_2)\right) - p_3(k_1 + k_2)^2 = \frac{k_1^2}{\tilde{p}_1} + \frac{k_2^2}{\tilde{p}_2} - (k_1 + k_2)^2.$$

Denote

$$\hat{R}^2 := (1-p_3)\left(R^2 - \frac{k_3^2}{p_3} + k_3^2 + 2k_3(k_1 + k_2)\right) - p_3(k_1 + k_2)^2 =: A(p_3)R^2 + B(p_3, k_3),$$

where

$$A(p_3) = (1-p_3),$$

$$B(p_3, k_3) = (k_3^2 + 2k_3(N - k_3) - \frac{k_3^2}{p_3})(1-p_3) - p_3(N - k_3)^2 = -(\sqrt{p_3}N - \frac{k_3}{\sqrt{p_3}})^2.$$

Note that for fixed k_3 and given p_3 there is a one-to-one relation between \hat{R}^2 and R^2 .

The probability distribution of the nodes X_k , given $|k| = N$ and given the probability vector $p = (p_1, p_2, p_3)$, can be rewritten as

$$\begin{aligned} \binom{k_1 + k_2 + k_3}{k_1, k_2, k_3} p_1^{k_1} p_2^{k_2} p_3^{k_3} &= \binom{k_1 + k_2}{k_1, k_2} \binom{k_1 + k_2 + k_3}{k_1 + k_2, k_3} \tilde{p}_1^{k_1} \tilde{p}_2^{k_2} (1-p_3)^{k_1+k_2} p_3^{k_3} \\ &= \binom{k_1 + k_2}{k_1, k_2} \tilde{p}_1^{k_1} \tilde{p}_2^{k_2} K(p_3, k_3), \end{aligned}$$

where

$$K(p_3, k_3) := \binom{N}{k_3} (1-p_3)^{N-k_3} p_3^{k_3}.$$

Hence, $\hat{R} = \sqrt{\frac{k_1^2}{\tilde{p}_1} + \frac{k_2^2}{\tilde{p}_2} - (k_1 + k_2)^2}$ has the probability $\binom{k_1+k_2}{k_1, k_2} \tilde{p}_1^{k_1} \tilde{p}_2^{k_2} K(p_3, k_3)$, where k_3 is fixed. The density contribution of \hat{R} is similar to the density of the radius in the one dimensional case, up to a factor $K(p_3, k_3)$ when p_3 is fixed. Recall that $q_1(x)$ is the density of x over the whole real line and $q_1(x)$ is even. Hence, the density of $|x|$ is two times $q_1(|x|)$. The density of \hat{R} is supported on non-negative half line. Therefore, the conditional compound density of \hat{R} , given p_3 , is $q_2(\hat{r}) = 2q_1(\hat{r})K(p_3, k_3)$, where k_3 is fixed, and $q_1(x) = \sum_{k_2=0}^{N-k_3} C_{k_2}(x)$, by applying Theorem 2.7.1. We have

- $N - k_3$ even:

$$q_1(x) = (C_0 + C_{N-k_3})(x) + C_{(N-k_3)/2}(x) + \sum_{k_2=1}^{\frac{N-k_3}{2}-1} (C_{k_2} + C_{N-k_3-k_2})(x).$$

- $N - k_3$ odd:

$$q_1(x) = (C_0 + C_{N-k_3})(x) + \sum_{k_2=1}^{\frac{N-k_3-1}{2}} (C_{k_2} + C_{N-k_3-k_2})(x).$$

Furthermore, from Theorem 2.7.1, we have

$$\begin{aligned} (C_0 + C_{N-k_3})(x) &= c_m \frac{A_m(x, 0)}{(x^2 + (N - k_3)^2)^{N-k_3+m}}, \\ (C_{k_2} + C_{N-k_3-k_2})(x, y_{k_2}) &= c_m \binom{N - k_3}{k_2} \frac{(2(N - k_3 - k_2))^{2(N-k_3)-2k_2+m} N(x, y_{k_2})}{D(x, y_{k_2})}, \end{aligned}$$

where

$$\begin{aligned}
A_m(x, 0) &= (N - k_3)^{2(N-k_3)+m} x^{m-1}, \\
N(x, y_{k_2}) &= (4(N - k_3 - k_2)^2 + (-x + y_{k_2})^2)^{N-k_3+m} (x + y_{k_2})^{2k_2+m} \\
&\quad + (4(N - k_3 - k_2)^2 + (x + y_{k_2})^2)^{N-k_3+m} (-x + y_{k_2})^{2k_2+m}, \\
D(x, y_{k_2}) &= ((4(N - k_3 - k_2)^2 + x^2 + y_{k_2}^2)^2 - 4x^2 y_{k_2}^2)^{N-k_3+m} y_{k_2}, \\
y_{k_2} &= \sqrt{x^2 + 4k_2(N - k_3 - k_2)}.
\end{aligned}$$

The function $q_1(x)$ has the form $q_1(x) = c_m \frac{A_m(x)}{(x^2 + (N - k_3)^2)^{N-k_3+m}}$, where $A_m(x)$ is a polynomial with rational coefficients and even degree at most $2(N - k_3 + m - 1)$.

The mixing density can be rewritten as

$$c(p_1^{1/2} p_2^{1/2} p_3^{1/2})^{m-2} dp_1 dp_3 = \left[\frac{1}{2} c_m (\tilde{p}_1^{1/2} \tilde{p}_2^{1/2})^{m-2} d\tilde{p}_1 \right] [\hat{c}_m (1 - p_3)^{m-1} p_3^{\frac{m-2}{2}} dp_3] \quad (3.3)$$

where $c_m = \frac{1}{\int_0^\infty \frac{\tau^{m-1}}{(1+\tau^2)^m} d\tau}$ by applying formula (2.5), $\hat{c}_m = \frac{1}{\int_0^1 (1-p_3)^{m-1} p_3^{\frac{m-2}{2}} dp_3}$, $c = (\frac{1}{2} c_m) \hat{c}_m$. It follows that the marginal density factorizes and

$\rho(p_3) := \hat{c}_m (1 - p_3)^{m-1} p_3^{\frac{m-2}{2}} dp_3$ is the marginal density of p_3 .

We make the transformation to obtain the density contribution of R, given k_3

$q_3(r, k_3) = q_2(\sqrt{A(p_3)r^2 + B(p_3, k_3)}) \frac{A(p_3)r}{\sqrt{A(p_3)r^2 + B(p_3, k_3)}}$. The compound density of R is

$$\begin{aligned}
q(r) &= \int_{p_3=0}^1 \left\{ \sum_{k_3=0}^N q_3(r, k_3) \rho(p_3) \right\} dp_3 = \sum_{k_3=0}^N \int_{p_3=0}^1 q_3(r, k_3) \rho(p_3) dp_3 \\
&= \sum_{k_3=0}^N \int_{p_3=0}^1 2q_1(\sqrt{A(p_3)r^2 + B(p_3, k_3)}) K(p_3, k_3) \frac{A(p_3)r}{\sqrt{A(p_3)r^2 + B(p_3, k_3)}} \rho(p_3) dp_3.
\end{aligned}$$

□

Theorem 3.6.2. *In a two asset case, the compound density of the radius can be written in another form as*

$$q(r) = \sum_{k_3=0}^N \int_{t=0}^{\infty} f_{k_3}(t; r) dt, \quad (3.4)$$

where for given r the rational density

$$f_{k_3}(t; r) := \frac{2 \binom{N}{k_3} \hat{c}_m (\alpha_2(t))^{N-k_3+m}}{2^{N+\frac{3m-3}{2}} s^{2N+3m-1}} \sum_{i=2}^{N-k_3+m} \left\{ \frac{a_i (vs^2 + yt^2)^{k_3 + \frac{m-1}{2} + i}}{(s^2 + t^2)^{N + \frac{3m+1}{2} - i} \alpha_1^i(t)} \right\}, \quad (3.5)$$

has the degree at $4N - 2k_3 + 5m + 1$ in the denominator, and degree at most $4N - 2k_3 + 5m - 1$ in the numerator. And a_i are known coefficients from partial fraction decomposition,

$$\begin{aligned} b &:= r^2 + 2Nk_3, \\ u &:= u(k_3) = r \sqrt{r^2 + 4k_3(N - k_3)}, \\ v &:= b + u = \frac{(r + \sqrt{r^2 + 4k_3(N - k_3)})^2 + 4k_3^2}{2}, \\ y &:= b - u = \frac{(r - \sqrt{r^2 + 4k_3(N - k_3)})^2 + 4k_3^2}{2}, \\ s &= \sqrt{r^2 + N^2}, \\ \hat{c}_m &= \frac{1}{\int_0^1 (1 - p_3)^{m-1} p_3^{\frac{m-2}{2}} dp_3}. \end{aligned}$$

The polynomial of the degree 4 of t is

$$\alpha_1(t) := (N - k_3)^2 s^4 v + (N - k_3)^2 y t^4 + (2s^2 u^2 + (N - k_3)^2 s^2 v + (N - k_3)^2 s^2 y) t^2,$$

and the quadratic function of t is

$$\alpha_2(t) := t^2(2s^2 - y) + 2s^4 - s^2v.$$

Proof. Our strategy is to apply the Euler's third substitution to make the transformation of the integral in (3.2) into the integral w.r.t the new variable t , $t \in [0, \infty)$, which can be solved.

We have $\hat{r} = \sqrt{A(p_3)r^2 + B(p_3, k_3)} = \sqrt{\frac{ap_3^2 + bp_3 + c}{p_3}}$, where

$$\begin{aligned} a &:= -(r^2 + N^2), \\ b &:= b(k_3) = r^2 + 2Nk_3, \\ c &:= c(k_3) = -k_3^2. \end{aligned}$$

Solving the inequality $\hat{r} > 0$, we obtain $0 < p_3^{(1)} < p_3 < p_3^{(2)} < 1$, where

$$\begin{aligned} p_3^{(1)} &= \frac{r^2 + 2Nk_3 - r\sqrt{r^2 + 4k_3(N - k_3)}}{2(r^2 + N^2)}, \\ p_3^{(2)} &= \frac{r^2 + 2Nk_3 + r\sqrt{r^2 + 4k_3(N - k_3)}}{2(r^2 + N^2)} \end{aligned}$$

are the two zeros of the equation $ap_3^2 + bp_3 + c = 0$. Note that at $p_3 = 0$, we have $ap_3^2 + bp_3 + c = c = -k_3^2 < 0$ and $ap_3^2 + bp_3 + c = c = -k_3^2 = 0$ if $k_3 = 0$ in which case $p_3^{(1)} = 0$.

Similarly, at $p_3 = 1$, we have $ap_3^2 + bp_3 + c = c = -(N + k_3)^2 < 0$.

The parabola opens downward due to $a = -(r^2 + N^2) < 0$. We have

$$\frac{d}{dp_3}(ap_3^2 + bp_3 + c) = 2ap_3 + b.$$

At $p_3 = 0$, we have $\frac{d}{dp_3}(ap_3^2 + bp_3 + c) = b > 0$.

At $p_3 = 1$, we have $\frac{d}{dp_3}(ap_3^2 + bp_3 + c) = 2a + b = -2r^2 - 2N^2 + 2Nk_3 < 0$, as $k_3 \leq N$.

Hence, $p_3^{(1)}, p_3^{(2)} \in [0, 1]$.

Given k_3 , (k_3 runs from 0 to N), we have

$$\begin{aligned} & 2q_1(\sqrt{A(p_3)r^2 + B(p_3, k_3)})K(p_3, k_3)\frac{A(p_3)r}{\sqrt{A(p_3)r^2 + B(p_3, k_3)}}\rho(p_3)dp_3 \\ &= (\sqrt{p_3}(1-p_3)K(p_3, k_3)\rho(p_3))\frac{2q_1(\hat{r})r}{\sqrt{ap_3^2 + bp_3 + c}}dp_3 \\ &= \left(\hat{c}_m \binom{N}{k_3} (1-p_3)^{N-k_3+m} p_3^{k_3+\frac{m-1}{2}}\right) \frac{2q_1(\hat{r})r}{\sqrt{ap_3^2 + bp_3 + c}} dp_3. \quad (1) \end{aligned}$$

Applying Euler's third substitution, we introduce a new variable t which satisfies the following equation.

$$\sqrt{ap_3^2 + bp_3 + c} = \sqrt{a(p_3 - p_3^{(1)})(p_3 - p_3^{(2)})} = (p_3 - p_3^{(1)})t.$$

This yields

$$\begin{aligned} p_3 &= \frac{ap_3^{(2)} - p_3^{(1)}t^2}{a - t^2} = \frac{vs^2 + yt^2}{2s^2(s^2 + t^2)}, \\ dp_3 &= -\frac{2ut}{(s^2 + t^2)^2} dt, \end{aligned}$$

where

$$\begin{aligned} u = u(k_3) &= r\sqrt{r^2 + 4k_3(N - k_3)}, \\ v = b + u &= \frac{(r + \sqrt{r^2 + 4k_3(N - k_3)})^2 + 4k_3^2}{2}, \\ y &= b - u = \frac{(r - \sqrt{r^2 + 4k_3(N - k_3)})^2 + 4k_3^2}{2}, \\ s &= \sqrt{r^2 + N^2}. \end{aligned}$$

We can obtain

$$\begin{aligned} \sqrt{ap_3^2 + bp_3 + c} &= (p_3 - p_3^{(1)})t = \frac{ut}{s^2 + t^2}, \\ \hat{r} &= \sqrt{\frac{ap_3^2 + bp_3 + c}{p_3}} = \frac{\sqrt{2}uts}{\sqrt{(s^2 + t^2)(vs^2 + yt^2)}}. \end{aligned}$$

By partial fraction decomposition, the function $q_1(x)$ from Theorem 3.6.1 has the equivalent form:

$$q_1(x) = \sum_{i=2}^{N-k_3+m} \frac{a_i}{(x^2 + (N - k_3)^2)^i},$$

where a_i are known coefficients from the partial fraction decomposition.

Hence, the function $q_1(\hat{r})$ can be rewritten as

$$q_1(\hat{r}) = \sum_{i=2}^{N-k_3+m} \frac{a_i(s^2 + t^2)^i(vs^2 + yt^2)^i}{(2u^2t^2s^2 + (N - k_3)^2(s^2 + t^2)(vs^2 + yt^2))^i}.$$

We have $2u^2t^2s^2 + (N - k_3)^2(s^2 + t^2)(vs^2 + yt^2) = (N - k_3)^2s^4v + (N - k_3)^2yt^4 + (2s^2u^2 + (N - k_3)^2s^2v + (N - k_3)^2s^2y)t^2$. Denote

$$\alpha_1(t) := (N - k_3)^2s^4v + (N - k_3)^2yt^4 + (2s^2u^2 + (N - k_3)^2s^2v + (N - k_3)^2s^2y)t^2.$$

Hence, the function $q_1(\hat{r})$ can be rewritten as

$$q_1(\hat{r}) = \sum_{i=2}^{N-k_3+m} \frac{a_i(s^2 + t^2)^i (vs^2 + yt^2)^i}{\alpha_1^i(t)}, \quad (2)$$

where a_i are known coefficients from the partial fraction decomposition.

We have

$$(1 - p_3)^{N-k_3+m} p_3^{k_3 + \frac{m-1}{2}} = \frac{(2s^2(s^2 + t^2) - (vs^2 + yt^2))^{N-k_3+m} (vs^2 + yt^2)^{k_3 + \frac{m-1}{2}}}{2^{N + \frac{3m-1}{2}} s^{2N+3m-1} (s^2 + t^2)^{N + \frac{3m-1}{2}}}.$$

Hence

$$\frac{r(1 - p_3)^{N-k_3+m} p_3^{k_3 + \frac{m-1}{2}}}{\sqrt{ap_3^2 + bp_3 + c}} dp_3 = - \frac{r(2s^2(s^2 + t^2) - (vs^2 + yt^2))^{N-k_3+m} (vs^2 + yt^2)^{k_3 + \frac{m-1}{2}}}{2^{N + \frac{3m-3}{2}} s^{2N+3m-1} (s^2 + t^2)^{N + \frac{3m+1}{2}}} dt.$$

We have

$$2s^2(s^2 + t^2) - (vs^2 + yt^2) = t^2(2s^2 - y) + 2s^4 - s^2v.$$

Denote $\alpha_2(t) := t^2(2s^2 - y) + 2s^4 - s^2v$.

Hence,

$$\frac{r(1 - p_3)^{N-k_3+m} p_3^{k_3 + \frac{m-1}{2}}}{\sqrt{ap_3^2 + bp_3 + c}} dp_3 = - \frac{r(\alpha_2(t))^{N-k_3+m} (vs^2 + yt^2)^{k_3 + \frac{m-1}{2}}}{2^{N + \frac{3m-3}{2}} s^{2N+3m-1} (s^2 + t^2)^{N + \frac{3m+1}{2}}} \quad (3)$$

Multiplying (2) by (3), we obtain

$$\begin{aligned} & \left((1 - p_3)^{N-k_3+m} p_3^{k_3 + \frac{m-1}{2}} \right) \frac{q_1(\hat{r})r}{\sqrt{ap_3^2 + bp_3 + c}} dp_3 \\ &= - \frac{(\alpha_2(t))^{N-k_3+m}}{2^{N + \frac{3m-3}{2}} s^{2N+3m-1}} \sum_{i=2}^{N-k_3+m} \left\{ \frac{a_i (vs^2 + yt^2)^{k_3 + \frac{m-1}{2} + i}}{(s^2 + t^2)^{N + \frac{3m+1}{2} - i} \alpha_1^i(t)} \right\}. \quad (4) \end{aligned}$$

Substituting (4) into (1), we obtain

$$2q_1(\sqrt{A(p_3)r^2 + B(p_3, k_3)})K(p_3, k_3)\frac{A(p_3)r}{\sqrt{A(p_3)r^2 + B(p_3, k_3)}}\rho(p_3)dp_3$$

$$= -\frac{2\binom{N}{k_3}\hat{c}_m(\alpha_2(t))^{N-k_3+m}}{2^{N+\frac{3m-3}{2}}s^{2N+3m-1}}\sum_{i=2}^{N-k_3+m}\left\{\frac{a_i(vs^2 + yt^2)^{k_3+\frac{m-1}{2}+i}}{(s^2 + t^2)^{N+\frac{3m+1}{2}-i}\alpha_1^i(t)}\right\}.$$

In summary, given k_3 (k_3 runs from 0 to N), the integrand has the following form

$$\hat{f}_{k_3}(t; r) := -\frac{2\binom{N}{k_3}\hat{c}_m(\alpha_2(t))^{N-k_3+m}}{2^{N+\frac{3m-3}{2}}s^{2N+3m-1}}\sum_{i=2}^{N-k_3+m}\left\{\frac{a_i(vs^2 + yt^2)^{k_3+\frac{m-1}{2}+i}}{(s^2 + t^2)^{N+\frac{3m+1}{2}-i}\alpha_1^i(t)}\right\}.$$

The equation $s^2 + t^2 = 0$ has the pole is in the upper half plane.

We find poles at which $\alpha_1(t) = 0$. Let

$$\begin{aligned} a_1 &= (N - k_3)^2 y, \\ b_1 &= s^2(2u^2 + (N - k_3)^2(v + y)), \\ c_1 &= (N - k_3)^2 s^4 v. \end{aligned}$$

We have a_1, b_1, c_1 are positive real zeros.

We prove that $\Delta = b_1^2 - 4a_1c_1 \geq 0$. We have

$$\Delta \geq 0$$

$$\Leftrightarrow (2u^2 + (N - k_3)^2(v + y))^2 \geq 4(N - k_3)^4vy$$

$$\Leftrightarrow 4u^2 + 4u^2(N - k_3)^2(v + y) + (N - k_3)^4(v^2 + y^2) + (N - k_3)^4 2vy \geq 4(N - k_3)^4vy$$

$$\Leftrightarrow 4u^2 + 4u^2(N - k_3)^2(v + y) + (N - k_3)^4(v^2 + y^2) - 2(N - k_3)^4vy \geq 0$$

$$\Leftrightarrow 4u^2 + 4u^2(N - k_3)^2(v + y) + ((N - k_3)^2v - (N - k_3)^2y)^2 \geq 0,$$

which is always true. We have

$$-b_1 - \sqrt{\Delta} = -(b_1 + \sqrt{\Delta}) \leq 0.$$

We have

$$-b_1 + \sqrt{\Delta} \leq 0.$$

The equation $\alpha_1(t) = 0$ has four complex zeros, where t_1, t_3 are in the upper half plane, and t_2, t_4 in the lower half plane. We have

$$\begin{aligned} t_1 &= \sqrt{\frac{-b_1 + \sqrt{\Delta}}{2a_1}} = id, \\ t_2 &= -\sqrt{\frac{-b_1 + \sqrt{\Delta}}{2a_1}} = -id, \\ t_3 &= \sqrt{\frac{-b_1 - \sqrt{\Delta}}{2a_1}} = ie, \\ t_4 &= -\sqrt{\frac{-b_1 - \sqrt{\Delta}}{2a_1}} = -ie, \end{aligned}$$

where $d = \sqrt{\frac{b_1 - \sqrt{\Delta}}{2a_1}}$,

$e = \sqrt{\frac{b_1 + \sqrt{\Delta}}{2a_1}}$.

Assume $t > 0$, $\frac{dt}{dp_3} < 0$, so $t(p_3)$ is a decreasing function in p_3 . Recall that

$$p_3 = \frac{ap_3^{(2)} - p_3^{(1)}t^2}{a - t^2},$$

so

$$t^2 = \frac{a(p_3 - p_3^{(2)})}{p_3 - p_3^{(1)}}.$$

When $p_3 = p_3^{(1)}$, $t(p_3^{(1)}) = \infty$.

When $p_3 = p_3^{(2)}$, $t(p_3^{(2)}) = 0$.

We have

$$\begin{aligned} & \int_0^1 2q_1(\sqrt{A(p_3)r^2 + B(p_3, k_3)})K(p_3, k_3) \frac{A(p_3)r}{\sqrt{A(p_3)r^2 + B(p_3, k_3)}} \rho(p_3) dp_3 \\ &= \int_{p_3^{(1)}}^{p_3^{(2)}} 2q_1(\sqrt{A(p_3)r^2 + B(p_3, k_3)})K(p_3, k_3) \frac{A(p_3)r}{\sqrt{A(p_3)r^2 + B(p_3, k_3)}} \rho(p_3) dp_3 \\ &= \int_{t(p_3^{(1)})}^{t(p_3^{(2)})} \hat{f}_{k_3}(t; r) dt = \int_{\infty}^0 \hat{f}_{k_3}(t; r) dt = - \int_0^{\infty} \hat{f}_{k_3}(t; r) dt = \int_0^{\infty} f_{k_3}(t; r) dt, \end{aligned}$$

where $f_{k_3}(t; r) := -\hat{f}_{k_3}(t; r)dt$, and $f_{k_3}(t; r)$ is non-negative.

Hence, the density of R is $q(r) = \sum_{k_3=0}^N \int_{t=0}^{\infty} f_{k_3}(t; r) dt$. □

Now we present how to solve the integral in Formula 3.4.

We have $f_{k_3}(t; r)$ is a rational density. Applying partial fraction decomposition over \mathbb{C} , we have $f_{k_3}(t) = g_{k_3}(t) + \bar{g}_{k_3}(t)$, where $g_{k_3}(t)$ has all poles $\alpha_1 = is$, $\alpha_2 = id$, $\alpha_3 = ie$ in the upper half plane, where s, d, e all positive, $s = \sqrt{r^2 + N^2}$, $d = \sqrt{\frac{b_1 - \sqrt{\Delta}}{2a_1}}$, $e = \sqrt{\frac{b_1 + \sqrt{\Delta}}{2a_1}}$, a_1 , b_1 , Δ as in the proof above. And g, and hence \bar{g} are strictly proper. Applying a result from [25], we have

$$\int_{-\infty}^{\infty} f_{k_3}(t; r) dt = 2\pi \lim_{|t| \rightarrow \infty} |tg_{k_3}(t)|.$$

Equivalently,

$$\int_0^{\infty} f_{k_3}(t; r) dt = \pi \lim_{|t| \rightarrow \infty} |t g_{k_3}(t)|.$$

We apply the residue method for calculations of $\int_0^{\infty} f_{k_3}(t; r) dt$. Function $g_{k_3}(t)$ is a rational proper fraction, and can be decomposed into

$$g_{k_3}(t) = \sum_j \sum_i \frac{a_{ij}}{(t - \alpha_i)^j}.$$

We only need $\sum_i \frac{a_{i1}}{(t - \alpha_i)}$ due to $\lim_{|t| \rightarrow \infty} |t \frac{a_{ij}}{(t - \alpha_i)^j}| = 0$ when $j > 1$. Each a_{i1} , $i = 1, 2, 3$ can be obtained using residue theory as follows.

At pole $\alpha_1 = is$ with the multiplicity n_1 , we get the coefficient

$$a_{11} = \frac{1}{(n_1 - 1)!} \lim_{t \rightarrow \alpha_1} \frac{d^{n_1-1}}{dt^{n_1-1}} [(t - \alpha_1)^{n_1} g_{k_3}(t)].$$

At pole $\alpha_2 = id$ with the multiplicity n_2 , we get the coefficient

$$a_{21} = \frac{1}{(n_2 - 1)!} \lim_{t \rightarrow \alpha_2} \frac{d^{n_2-1}}{dt^{n_2-1}} [(t - \alpha_2)^{n_2} g_{k_3}(t)].$$

At pole $\alpha_3 = ie$ with the multiplicity n_3 , we get the coefficient

$$a_{31} = \frac{1}{(n_3 - 1)!} \lim_{t \rightarrow \alpha_3} \frac{d^{n_3-1}}{dt^{n_3-1}} [(t - \alpha_3)^{n_3} g_{k_3}(t)].$$

In the special case that α_i is a simple root, we have $a_{i1} = \frac{N(\alpha_i)}{D'(\alpha_i)}$, $i = 1, 2, 3$, where polynomials $N(t)$, $D(t)$ are the numerator and the denominator of $g_{k_3}(t)$.

Figure 3.1 and Figure 3.2 plot the compound density of the standardized radius versus the ideal density in the analytic approach and a numerical approach, respectively, at

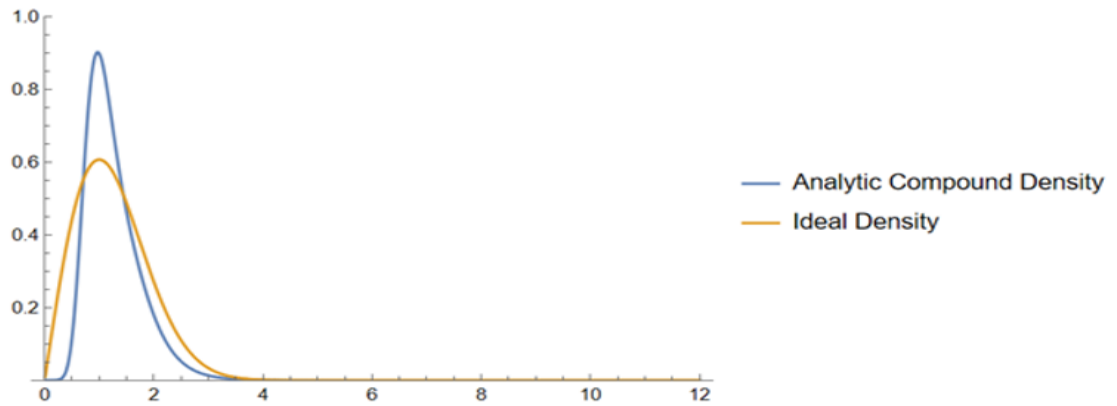


Figure 3.1: Analytic compound density at $N = 2, m = 9$.

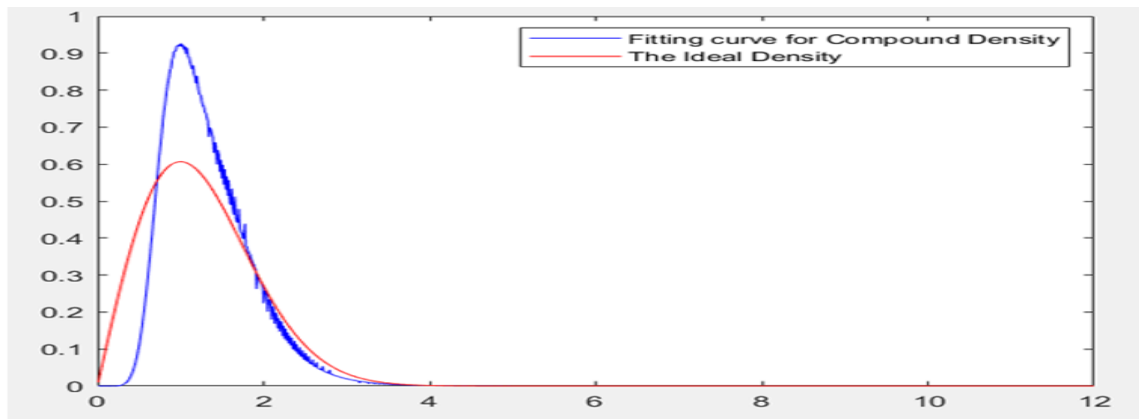


Figure 3.2: Approximated compound density at $N = 2, m = 9$.

$N = 2, m = 9$. The numerical approach will be explained in section 3.12.2. Note that at this low tree depth, the two are still quite distinct.

3.7. Generating the Histogram

We aim to approximate the compound density of the mixture distribution numerically when an analytical solution can not be found easily. We run Monte Carlo on the

variables θ_i , $i = 1, 2, \dots, d$, where d is the number of risky assets, then a histogram can be generated numerically. The algorithm for generating the histogram for the standardized radius is presented as follows:

First, partition along the r -axis $0 = r_1 < r_1 < \dots < r_{I+1}$, such that $r_i \in \mathbb{R}$,

$\forall i \in \{1, \dots, I + 1\}$ and $r_{i+1} - r_i = \Delta$.

Consider the standardized radius $R = \sqrt{\frac{\sum_{i=1}^{d+1} \frac{k_i^2}{p_i} - N^2}{N}} \sim q(r)$ such that $r \geq 0$, and M distinct trees of N time-steps are generated. Here $q(r)$ is the scaled pdf of R . Let $R_{k,m}$ denotes a drawing of the radius at the k^{th} node in the m^{th} tree. Define

$$p_{k,m} = \binom{|k|}{k} (p_1^{k,m})^{k_1} (p_2^{k,m})^{k_2} \dots (p_{d+1}^{k,m})^{k_{d+1}},$$

where $k = (k_1, k_2, k_3, \dots, k_{d+1})$ run over all vectors in \mathbb{N}_0^{d+1} ,

$$\begin{aligned} |k| &= \sum_{i=1}^{d+1} k_i = N, \\ (p_1^{k,m})^{1/2} &= \cos(\theta_1^m), \\ (p_2^{k,m})^{1/2} &= \sin(\theta_1^m) \cos(\theta_2^m), \\ (p_3^{k,m})^{1/2} &= \sin(\theta_1^m) \sin(\theta_2^m) \cos(\theta_3^m), \\ &\vdots \\ (p_{d+1}^{k,m})^{1/2} &= \sin(\theta_1^m) \dots \sin(\theta_{d-1}^m) \sin(\theta_d^m). \end{aligned}$$

Now define

$$p^{(r_i)} = \begin{cases} \sum_k \sum_{m=1}^M \frac{p_{k,m}}{M} \mathbb{1}_{\{r_1 \leq R_{k,m} < r_2\}}, & i = 1 \\ \sum_k \sum_{m=1}^M \frac{p_{k,m}}{M} \mathbb{1}_{\{r_2 \leq R_{k,m} < r_3\}}, & i = 2 \\ \vdots \\ \sum_k \sum_{m=1}^M \frac{p_{k,m}}{M} \mathbb{1}_{\{r_{I+1} \leq R_{k,m}\}}, & i = I + 1 \end{cases}$$

Then, scaling the probabilities here by $\frac{1}{\Delta}$, we obtain the following continuous, piecewise, probability density function, $\hat{q}(r)$, which is a numerical approximation of the density of the mixture distribution $q(r)$.

$$\hat{q}(r) = \begin{cases} \frac{p^{(r_1)}}{\Delta}, & r_1 \leq r < r_2 \\ \frac{p^{(r_2)}}{\Delta}, & r_2 \leq r < r_3 \\ \vdots \\ \frac{p^{(r_{I+1})}}{\Delta}, & r_{I+1} \leq r \end{cases}$$

Note that $\int_0^\infty \hat{q}(r) dr = \frac{p^{(r_1)}}{\Delta} \Delta + \frac{p^{(r_2)}}{\Delta} \Delta + \dots + \frac{p^{(r_{I+1})}}{\Delta} \Delta = 1$.

It can be approximately stated that the random variable $R_N \sim q(r)$. Consider the midpoints of each subintervals: $[r_1, r_2), \dots, [r_{I-1}, r_I), [r_I, r_{I+1}), [r_{I+1}, r_{I+2})$, we get vector $\hat{r} = [(r_1 + \frac{\Delta}{2}), (r_2 + \frac{\Delta}{2}), \dots, (r_{I+1} + \frac{\Delta}{2})]^T$ and the corresponding vector $\hat{p}^r = [\frac{p^{(r_1)}}{\Delta}, \frac{p^{(r_2)}}{\Delta}, \dots, \frac{p^{(r_{I+1})}}{\Delta}]^T$. The first and the second moments of the histogram can be calculated

$$\begin{aligned} \mathbf{E}[R_N] &\approx \langle \hat{r}, \hat{p}^r \rangle, \\ \mathbf{E}[R_N^2] &\approx \langle \text{diag}(\hat{r}^T) \hat{r}, \hat{p}^r \rangle. \end{aligned}$$

Remark:

In the same way, we can generate the histogram for the standardized squared radius $P = \frac{R^2}{2N}$. An example of such a histogram is presented by Figure 3.4 in section 3.12.1.

3.8. Discrete Entropy

Consider a continuous random variable R with a continuous pdf $q(r)$. The length of subintervals of the partition along the r -axis is Δ . We have

$$RIEnt(R) = - \int_0^{\infty} q(r) \log\left(\frac{q(r)}{S(r)}\right) dr = - \int_0^{\infty} q(r) \log(q(r)) dr + \int_0^{\infty} q(r) \log(S(r)) dr.$$

We have

$$\begin{aligned} \int_0^{\infty} q(r) \log(q(r)) dr &\approx \int_{r_1}^{r_2} q(r) \log(q(r)) dr \\ &+ \int_{r_2}^{r_3} q(r) \log(q(r)) dr + \cdots + \int_{r_{I+1}}^{r_{I+2}} q(r) \log(q(r)) dr \\ &\approx \sum_{i=1}^{I+1} q(r_i) \log(q(r_i)) \Delta. \end{aligned}$$

We have

$$\int_0^{\infty} q(r) \log(S(r)) dr \approx \sum_{i=1}^{I+1} q(r_i) \int_{r_i}^{r_{i+1}} \log(S(r)) dr.$$

We define:

$$RIEnt(\hat{R}_{\Delta}) := - \sum_i q(r_i) [\log(q(r_i)) \Delta - W_{r_i}],$$

where $W_{r_i} := \int_{r_i}^{r_{i+1}} \log(S(r)) dr$.

The RI entropy of variable R , $RIEnt(R)$, can be approximated by $RIEnt(\hat{R}_{\Delta})$

Two Assets Case:

RI Entropy of $q(r)$ is

$$RIEnt(R) = - \int_0^{\infty} q(r) \log\left(\frac{q(r)}{2\pi r}\right) dr.$$

In two asset case, we have $S(r) = 2\pi r$. We have

$$\begin{aligned} W_{r_i} &= \int_{r_i}^{r_{i+1}} \log(2\pi r) dr = \log(2\pi)\Delta + \int_{r_i}^{r_{i+1}} \log(r) dr \\ &= \log(2\pi)\Delta + r_{i+1} \log(r_{i+1}) - r_i \log(r_i) - r_i + r_{i+1} \\ &= \log(2\pi)\Delta + r_{i+1} \log(r_{i+1}) - r_i \log(r_i) - \Delta. \end{aligned}$$

$$RIEnt(\hat{R}_\Delta) := - \sum_i q(r_i) [\Delta \log(q(r_i)) - \Delta \log(2\pi) - (r_{i+1} \log(r_{i+1}) - r_i \log(r_i) - \Delta)].$$

Three Assets Case:

RI Entropy of $q(r)$ is

$$RIEnt(R) = - \int_0^\infty q(r) \log\left(\frac{q(r)}{4\pi r^2}\right) dr.$$

In three asset case, we have $S(r) = 4\pi r^2$. We have

$$\begin{aligned} W_{r_i} &= \int_{r_i}^{r_{i+1}} \log(4\pi r^2) dr = \log(4\pi)\Delta + \int_{r_i}^{r_{i+1}} \log(r^2) dr \\ &= \log(4\pi)\Delta + 2(r_{i+1} \log(r_{i+1}) - r_i \log(r_i) - r_i + r_{i+1}) \\ &= \log(4\pi)\Delta + 2(r_{i+1} \log(r_{i+1}) - r_i \log(r_i) - \Delta). \end{aligned}$$

$$RIEnt(\hat{R}_\Delta) := - \sum_i q(r_i) [\Delta \log(q(r_i)) - \Delta \log(4\pi) - 2(r_{i+1} \log(r_{i+1}) - r_i \log(r_i) - \Delta)].$$

The RI entropy of variable R , $RIEnt(R)$, can be approximated by $RIEnt(\hat{R}_\Delta)$.

3.9. The Characteristic Function and cdf

Let the random variable P denote the standardized squared radius at the final nodes of a tree of depth N .

The characteristic function is defined as

$$\begin{aligned}\phi_P(t) &= E[e^{iPt}] = E[E[e^{itP}|\Theta]] = E\left[\sum_{k_1, k_2, k_3} e^{itP} p_{k_1, k_2, k_3}(\theta_1, \theta_2)\right] \\ &= \sum_{k_1, k_2, k_3, k_1+k_2+k_3=N} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} e^{itP_k(\Theta)} p_{(k_1, k_2, k_3)}(\theta_1, \theta_2) p_m(\theta_1, \theta_2) d\theta_1 d\theta_2,\end{aligned}$$

where the mixing density $p_m(\theta_1, \theta_2) = c(p_1^{1/1} p_2^{1/2} p_3^{1/2})^m dp_1 dp_2$, c is a normalization constant, m is the power of the mixing density; $p_{k_1, k_2, k_3}(\theta_1, \theta_2) = \frac{N!}{k_1! k_2! k_3!} p_1^{k_1} p_2^{k_2} p_3^{k_3}$, $|k| = N$, the probabilities of the trinomial distribution.

The cdf can be found as follows:

$$\begin{aligned}F_P(\rho) &= P(P \leq \rho) = E[\mathbb{1}_{\{P \leq \rho\}}] = E[E[\mathbb{1}_{\{P \leq \rho\}}|\Theta]] \\ &= E_\Theta\left[\sum_{k_1, k_2, k_3, k_1+k_2+k_3=N} \mathbb{1}_{\{P_k(\Theta) \leq \rho\}} p_{k_1, k_2, k_3}(\theta_1, \theta_2)\right] \\ &= \sum_{k_1, k_2, k_3, k_1+k_2+k_3=N} \int_{\theta_1} \int_{\theta_2} \mathbb{1}_{\{P_k(\Theta) \leq \rho\}} p_{k_1, k_2, k_3}(\theta_1, \theta_2) p_m(\theta_1, \theta_2) d\theta_1 d\theta_2.\end{aligned}$$

The density can be derived from cdf by taking the derivative. The numerical results of this section can be found in section 3.12.2.

3.10. Entropy Optimization

3.10.1. Convex Combination

In this section, we still want to find mixing density with maximal compound density entropy. We present two approaches. Firstly, the strategy could now be to search through the class of polynomials up to a certain degree as mixing density to find the one that maximizes the compound density entropy. Then we could try various degrees, e.g. m from 1 to 30, to find the best mixing density.

We find the mixing density and corresponding cdf for the range of m from 1 to 30 in the two assets case. Then we form a convex combination of those mixing densities and apply our optimization algorithm. Then we find the optimal coefficients in the convex combination so that the new mixing density maximizes the compound density. Secondly, we present the idea for a convex combination of symmetrized monomials. Let consider the case of 2-assets ($d = 2$). The idea is as follows:

- **Step 1:** Fix the total degree, m .
- **Step 2:** Generate all monomials in $\sqrt{p_1}$, $\sqrt{p_2}$, $\sqrt{p_3}$ which have a total degree at most m .
- **Step 3:** Symmetrize each one of them. We find the symmetrized version of each monomial $(\sqrt{p_1})^{m_1}(\sqrt{p_2})^{m_2}(\sqrt{p_3})^{m_3}$.
- **Step 4:** Consider a linear combination which produces non-negative functions on the set $\{\sqrt{p_1} \geq 0, \sqrt{p_2} \geq 0, \sqrt{p_3} \geq 0, p_1 + p_2 + p_3 = 1\}$.

We could restrict to convex combinations of the normalized versions of these symmetrized monomials. Many monomials will give rise to the same symmetrized polynomials.

To deal with that efficiently, alternative approach would be to use $\max(m_i) \leq m$ as an upper bound in our search at step m ; i.e, $1 \leq m_3 \leq m_2 \leq m_1 \leq m$. Given m , we have $\frac{m(m+1)}{2}$ new monomials in total. Given a value of m , we find the symmetrized version of each monomial $(\sqrt{p_1})^{m_1}(\sqrt{p_2})^{m_2}(\sqrt{p_3})^{m_3}$, as follows:

- If the exponent of each monomials $m_1 = m_2 = m_3 = m$, then the symmetrized version is $(\sqrt{p_1})^m(\sqrt{p_2})^m(\sqrt{p_3})^m$.

- If $m_1 = m_2 = m$, the symmetrized version consists of three terms:

$$\frac{1}{3}(\sqrt{p_1})^m(\sqrt{p_2})^m(\sqrt{p_3})^{m_3} + \frac{1}{3}(\sqrt{p_1})^m(\sqrt{p_2})^{m_3}(\sqrt{p_3})^m + \frac{1}{3}(\sqrt{p_1})^{m_3}(\sqrt{p_2})^m(\sqrt{p_3})^m.$$

If $m_2 = m_3 = m$, the symmetrized version consists of three terms:

$$\frac{1}{3}(\sqrt{p_1})^{m_1}(\sqrt{p_2})^m(\sqrt{p_3})^m + \frac{1}{3}(\sqrt{p_1})^m(\sqrt{p_2})^{m_1}(\sqrt{p_3})^m + \frac{1}{3}(\sqrt{p_1})^m(\sqrt{p_2})^m(\sqrt{p_3})^{m_1}.$$

If $m_1 = m_3 = m$, the symmetrized version consists of three terms:

$$\frac{1}{3}(\sqrt{p_1})^{m_2}(\sqrt{p_2})^m(\sqrt{p_3})^m + \frac{1}{3}(\sqrt{p_1})^m(\sqrt{p_2})^{m_2}(\sqrt{p_3})^m + \frac{1}{3}(\sqrt{p_1})^m(\sqrt{p_2})^m(\sqrt{p_3})^{m_2}.$$

We apply random sampling by generating a vector $U = (u_1, u_2, u_3)$ with the corresponding $p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ for choosing a new monomial among 3 terms.

- If $m_1 \neq m_2 \neq m_3$, the symmetrized version consists 6 terms:

$$\begin{aligned} & \frac{1}{6}(\sqrt{p_1})^{m_1}(\sqrt{p_2})^{m_2}(\sqrt{p_3})^{m_3} + \frac{1}{6}(\sqrt{p_1})^{m_1}(\sqrt{p_2})^{m_3}(\sqrt{p_3})^{m_2} + \frac{1}{6}(\sqrt{p_1})^{m_2}(\sqrt{p_2})^{m_1}(\sqrt{p_3})^{m_3} + \\ & \frac{1}{6}(\sqrt{p_1})^{m_2}(\sqrt{p_2})^{m_3}(\sqrt{p_3})^{m_1} + \frac{1}{6}(\sqrt{p_1})^{m_3}(\sqrt{p_2})^{m_1}(\sqrt{p_3})^{m_2} + \frac{1}{6}(\sqrt{p_1})^{m_3}(\sqrt{p_2})^{m_2}(\sqrt{p_3})^{m_1}, \end{aligned}$$

taking $(m_1, m_2, m_3), (m_1, m_3, m_2), (m_2, m_1, m_3), (m_2, m_3, m_1), (m_3, m_1, m_2), (m_3, m_2, m_1)$ as exponent. We apply random sampling by generating a vector $V = (v_1, v_2, v_3, v_4, v_5, v_6)$ with the corresponding $p = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$ for choosing a new monomial among 6 terms.

For example, given $m = 11$, there are $\frac{11(11+1)}{2} = 66$ new monomials in total. All mixing densities and the corresponding cdf can be generated from the notebook file at <https://gitlab.com/trinhthuanyen2017/mixing-densities-and-the-corresponding-cdf-w.r.t-m>

3.10.2. Unilateral Optimization

We work in the context of using Monte Carlo based histograms.

Consider a partition \mathbf{P} of the nonnegative r-axis: $0 = r_1 < r_2 < \dots < r_{I+1}$. And W_{r_i} is defined as in Section 3.8. Let the (total) probability of probability density $j \in \{1, 2, \dots, J\}$ in the interval $r \in [r_i, r_{i+1}]$, $i = 1, \dots, I + 1$ be denoted by q_{rj} .

Proposition 3.10.1. *If for each subinterval $[r_i, r_{i+1})$ there is at most one $j \in \{1, 2, \dots, J\}$ with $q_{rj} > 0$, then the optimal RI Entropy mixture is given by the formula*

$$c_j = \frac{e^{RIEnt_j}}{\sum_h e^{RIEnt_h}}, \quad RIEnt_j = - \sum_{\{r:q_{rj}>0\}} q_{rj} [\log(q_{rj})\Delta - W_{r_i}],$$

where $c_j > 0$, $\sum_{j=1}^J c_j = 1$

Proof. The RI entropy of the mixture $(\sum_j c_j q_{rj})_{r \in \mathbf{P}}$ is equal to

$$RIEnt(\sum_j c_j q_{rj}) = - \sum_{\{r \in \mathbf{P}\}} \sum_{\{j: c_j > 0\}} c_j q_{rj} [\log(c_j q_{rj})\Delta - W_{r_i}].$$

Due to the fact that each subinterval has at most one j for which $q_{rj} > 0$, we have

$$\begin{aligned}
RIEnt(\sum_j c_j q_{rj}) &= - \sum_{\{r \in \mathbf{P}\}} \sum_{\{j: c_j > 0\}} c_j \log(c_j) q_{rj} \Delta - \sum_{\{r \in \mathbf{P}\}} \sum_{\{j: c_j > 0\}} c_j q_{rj} [\log(q_{rj}) \Delta - W_{r_i}] \\
&= - \sum_{\{j: c_j > 0\}} c_j \log(c_j) (\sum_{\{r \in \mathbf{P}\}} q_{rj}) \Delta + \sum_j c_j (- \sum_{\{r \in \mathbf{P}\}} q_{rj} [\log(q_{rj}) \Delta - W_{r_i}]) \\
&= - \sum_{\{j: c_j > 0\}} c_j \log(c_j) + \sum_j c_j RIEnt_j
\end{aligned}$$

because $\sum_{\{r \in \mathbf{P}\}} q_{rj} = \frac{1}{\Delta}$ and $RIEnt_j = - \sum_{\{r \in \mathbf{P}\}} q_{rj} [\log(q_{rj}) \Delta - W_{r_i}]$. Using Lagrangian technique to find the optimum for this expression w.r.t c_1, c_2, \dots, c_J with $\sum_j c_j = 1$, we have

$$\mathbf{L} = - \sum_j c_j \log(c_j) + \sum_j c_j RIEnt_j - \lambda [\sum_j c_j - 1].$$

Set $\frac{\partial \mathbf{L}}{\partial c_j} = 0$, we obtain

$$- \log(c_j) - 1 + RIEnt_j - \lambda = 0, \quad j = 1, 2, \dots, J.$$

It is implied that

$$\log(c_j) = RIEnt_j - (\lambda + 1) \Rightarrow c_j = \mu e^{RIEnt_j}, \quad \mu = e^{-(\lambda+1)}.$$

We use the constraint $\sum_j c_j = 1$ to determine μ .

$$\sum_j c_j = 1 \Leftrightarrow \sum_j \mu e^{RIEnt_j} = \mu \sum_{j=1}^J e^{RIEnt_j} = 1 \Leftrightarrow \mu = \frac{1}{\sum_{j=1}^J e^{RIEnt_j}},$$

so $c_j = \frac{e^{RIEnt_j}}{\sum_h e^{RIEnt_h}}$. Using this optimal value of c_j , the optimal entropy is

$$\begin{aligned} RIEnt\left(\sum_j c_j q_{rj}\right) &= -\sum_j c_j [\log(c_j) - RIEnt_j] \\ &= -\sum_j \mu e^{RIEnt_j} (\log(\mu) + RIEnt_j - RIEnt_j) \\ &= -\mu \log(\mu) \sum_j e^{RIEnt_j} = -\log(\mu) = \log\left(\sum_j e^{RIEnt_j}\right). \end{aligned}$$

Equivalently, $e^{RIEnt(\sum_j c_j q_{rj})} = \sum_j e^{RIEnt_j}$. \square

We can optimize, using the convexity of the entropy as a function of c_1, \dots, c_J , $\sum_j c_j = 1$, by optimizing "unilaterally" w.r.t each of the variables but respecting the restriction $\sum_j c_j = 1$. The "unilateral" optimization w.r.t c_j with the restriction $\sum_j c_j = 1$ proceeds as follows:

1. Iteration 1

- **Initialization Step:** Fix $(c_1^0, c_2^0, \dots, c_J^0)$, where $c_j^0 = \frac{e^{RIEnt_j}}{\sum_h e^{RIEnt_h}}$.

- **Step 1:** Do "unilateral" optimization for c_1 ($j = 1$).

$$\text{Define } q_{rj}^0 := \frac{\sum_{h \neq j, h \in \{1, 2, \dots, J\}} c_h^0 q_{rh}}{1 - c_j^0}, \quad r \in \mathbf{P}.$$

Note that $\sum_{\{h \neq j, h \in \{1, \dots, J\}\}} c_h^0 = 1 - c_j^0$ and $\sum_{r \in \mathbf{P}} q_{rj}^0 = \frac{1}{\Delta}$. Hence

$$c_j^0 q_{rj} + (1 - c_j^0) q_{rj}^0 = \sum_{h=1}^J c_h^0 q_{rh}.$$

Let $g(c_j) = -\sum_{r \in \mathbf{P}} (c_j^0 q_{rj} + (1 - c_j^0) q_{rj}^0) [\log(c_j^0 q_{rj} + (1 - c_j^0) q_{rj}^0) \Delta - W_{r_i}]$. Then

$$g'(c_j) = -\sum_{r \in \mathbf{P}} (q_{rj} - q_{rj}^0) [\log(c_j^0 q_{rj} + (1 - c_j^0) q_{rj}^0) \Delta - W_{r_i}].$$

Hence, $g''(c_j) = -\Delta \sum_{\rho \in \mathbf{P}} \left\{ \frac{(p_{\rho j} - q_{\rho j}^0)^2}{c_j^0 p_{\rho j} + (1 - c_j^0) q_{\rho j}^0} \right\} \leq 0$. There are three following cases.

- **Case 1:** $c_j^* \in (0, 1)$ is a critical point $g'(c_j) = 0$, $g(c_j)$ is increasing on $[0, c_j^*]$ but decreasing on $[c_j^*, 1]$. $g(c_j)$ is maximized at the critical point c_j^* . The value c_j^* can be found by using the bisection method applied to g' .
- **Case 2:** The critical point $c_j^* = 1$ and $g(c_j)$ is increasing on $[0, 1]$, so $g(c_j)$ is maximized at $c_j^* = 1$ if

$$g'(1) = - \sum_{\rho \in \mathbf{P}} (q_{\rho j} - q_{\rho j}^0) [\log(q_{\rho j}) \Delta - W_{r_i}] \geq 0.$$

- **Case 3:** The critical point $c_j^* = 0$ and $g(c_j)$ is decreasing on $[0, 1]$, so $g(c_j)$ is maximized at $c_j^* = 0$ if

$$g'(0) = - \sum_{\rho \in \mathbf{P}} (q_{\rho j} - q_{\rho j}^0) [\log(q_{\rho j}^0) \Delta - W_{r_i}] \leq 0.$$

The above optimization procedure produces the optimal coefficient $c_1^1 = c_1^*$, and $c_j^1 = \frac{1-c_1^1}{1-c_1^0} c_j^0$, $j = 2, 3, 4, \dots, J$.

- **Step 2:** Update vector $c = (c_1^1, c_2^1, \dots, c_J^1)$.

If $c_2^1 = 1$, no rebalancing of the weights can be carried out, let alone optimal rebalancing, so this step leaves the weights vector c unchanged. Otherwise, we do "unilateral" optimization for c_2 ($j = 2$) by repeating procedure in step 1 with $c = (c_1^1, c_2^1, \dots, c_J^1)$ as the starting vector at the initialization step.

The optimization procedure produces the optimal coefficient $c_2^2 = c_2^*$, and $c_j^2 = \frac{1-c_2^2}{1-c_2^1} c_j^1$, $j = 1, 3, 4, \dots, J$.

⋮

- **Step J:** We repeat our procedure until step J to obtain the optimal vector at this step at $c_j^J = c_j^*$, and $c_j^J = \frac{1-c_j^J}{1-c_j^{J-1}} c_j^{J-1}$, $j = 1, 2, 3, 4, \dots, J - 1$.

2. Iteration 2:

We repeat all the steps of the iteration one. However, the starting vector now is the optimal vector that we obtain in the iteration one.

The process is continued in an analogous fashion until the stopping criteria are reached. The stopping criteria are that all vectors at each iteration or two consecutive iterates are identical. The stopping criteria are that there is no further improvement, and an optimum was reached.

In the case of no further improvement, the gradient of the function on the space $\{c = (c_1, \dots, c_J) : \sum_{j=1}^J c_j = 1\}$ will be zero, and the unique local and hence global maximum will be reached, due to the convexity.

3.11. Option Pricing Using MC-Tree with the Bias-Correction

To compute a European option, a continuous-time price process of d assets Z is modeled by the a geometric Brownian motion with drift, given by

$dZ_j = Z_j \mu_j dt + Z_j \sigma_j dW_j$, for all $1 \leq j \leq d$, where $W = (W_1, W_2, \dots, W_d)^T$ is a vector Brownian motion process with the correlation matrix Γ . We will use the direction vectors $v_i(j)$ in case case $\mu_j = r$, $j = 1, 2, \dots, d$, and corresponding probability vector $p = (p_1, p_2, \dots, p_{d+1})$. The direction vector $v_i(j)$ is defined as follows:

Let $v = \{v_i\}_{i=1}^{d+1}$ be given by

$$v_i(j) := \exp\{\sqrt{\delta}(LHM)_{ij} + \hat{\mu}_j \delta\},$$

for all $1 \leq j \leq d$, for all $1 \leq i \leq d + 1$, where L is a $dx d$ matrix such that $LL^T = \Sigma$, Σ is a covariance matrix, L can be computed by applying Cholesky decomposition; M is a $dx(d+1)$ matrix, and columns of M is the M -vectors (see [40]); H is a $dx d$ matrix, as described in section 3.1.

Using bias-corrected direction vectors v_i with elements $v_i(j)$ give $\tilde{v}_i(j) := v_i(j) \exp\{\lambda_j \delta\}$, $1 \leq i \leq d+1$, $1 \leq j \leq d$, where the real numbers λ_j , $j = 1, 2, \dots, d$ are solved from

$$p_1 \tilde{v}_1(j) + p_2 \tilde{v}_2(j) + \dots + p_{d+1} \tilde{v}_{d+1}(j) = e^{r\delta}.$$

This correction amounts to replacing $\hat{\mu}_j = r - \frac{1}{2}\sigma_j^2$ by $\hat{\mu}_j = r - \frac{1}{2}\sigma_j^2 + \lambda_j$, $j = 1, 2, \dots, d$.

The bias correction in two assets case is defined as

$$\lambda_1 = r - \frac{\log(p_1 v_1(1) + p_2 v_2(1) + p_3 v_3(1))}{\delta},$$

$$\lambda_2 = r - \frac{\log(p_1 v_1(2) + p_2 v_2(2) + p_3 v_3(2))}{\delta},$$

where the three direction vectors are denoted by

$$v_1 := (v_1(1), v_1(2)),$$

$$v_2 := (v_2(1), v_2(2)),$$

$$v_3 := (v_3(1), v_3(2)).$$

The resulting tree model is complete and free of arbitrage.

3.12. Numerical Results in the Two Assets Case

3.12.1. Histogram and Discrete Entropy

In this section, we use the following inputs: range=12, numbers of bins=120 for experiments. Given those inputs, the following histogram in Figure 3.3 is an illustrated example with MC-drawing $M = 10^4$ from the mixing density of the power $m = 9$. The value of RI

Entropy is 2.836196, and the second moment is 2.000878. This can be compared with the theoretical values of 2.83788 for the entropy and 2 for the second moment. Our alternative approach (see remark in section 3.7) produces the following histogram in Figure 3.4 for inputs: range=10, numbers of bins=250. The entropy value is 0.9997302, and the second moment is 1.970511.

Figure 3.5 displays the entropy values as N increases. This experiment tests the effect on the entropy of mixture distribution when changing the tree depth N . In this experiment, the tree depth N is in the set $\{10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50\}$. Figure 3.6 shows effects on entropy values with the changes of the bin-width from mixing density with the power $m = 9$ and the tree depth $N = 10$. Both Figure 3.5 and Figure 3.6 show that entropy values seem to be close and going up, but one could wonder whether we can actually have the convergence to the exact Gaussian entropy. That is an interesting question for future research.

3.12.2. The Characteristic Function and the cdf

We partition t from -12 to 12 with the step size 0.01. Figure 3.7 shows the characteristic function for this partition at $N = 10$ and $m = 9$. Figure 3.8 plots the cdf of the standardized squared radius $\rho = \frac{R^2}{2N}$ and the cdf of standard exponential distribution when partitioning ρ from 0.01 to 8 with 4000 points at $N = 30$ and $m = 30$. Figure 3.9 shows the corresponding PDF.

Figure 3.10 plots cdf of the standardized radius r and cdf of the ideal distribution when partitioning r from 0.01 to 8 with 4000 points at $N = 30$ and $m = 30$. Figure 3.11 shows the corresponding pdf.

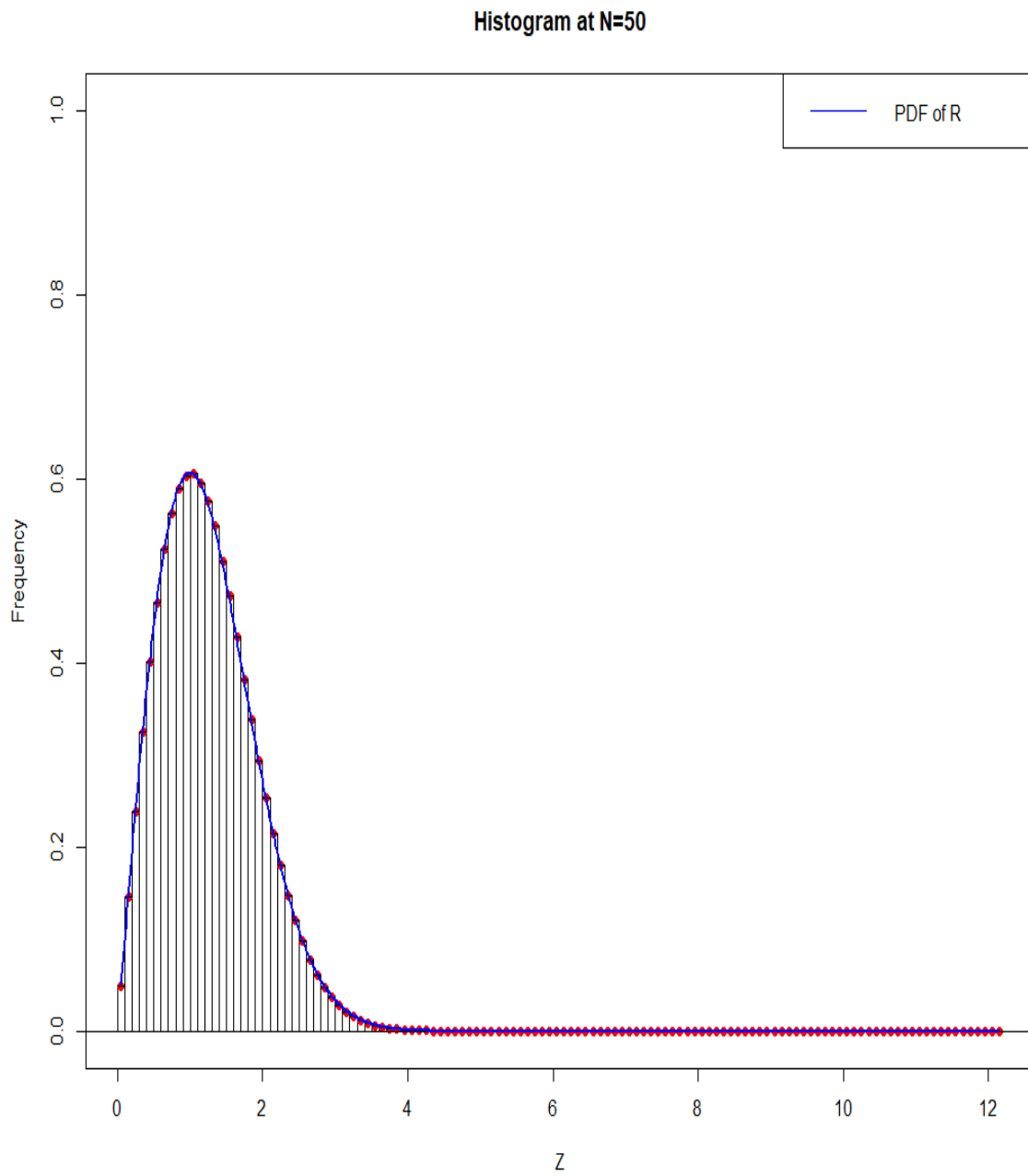


Figure 3.3: Generated histogram for the standardized radius using mixing density with $m = 9$. Red points are the mid points of histogram bins.

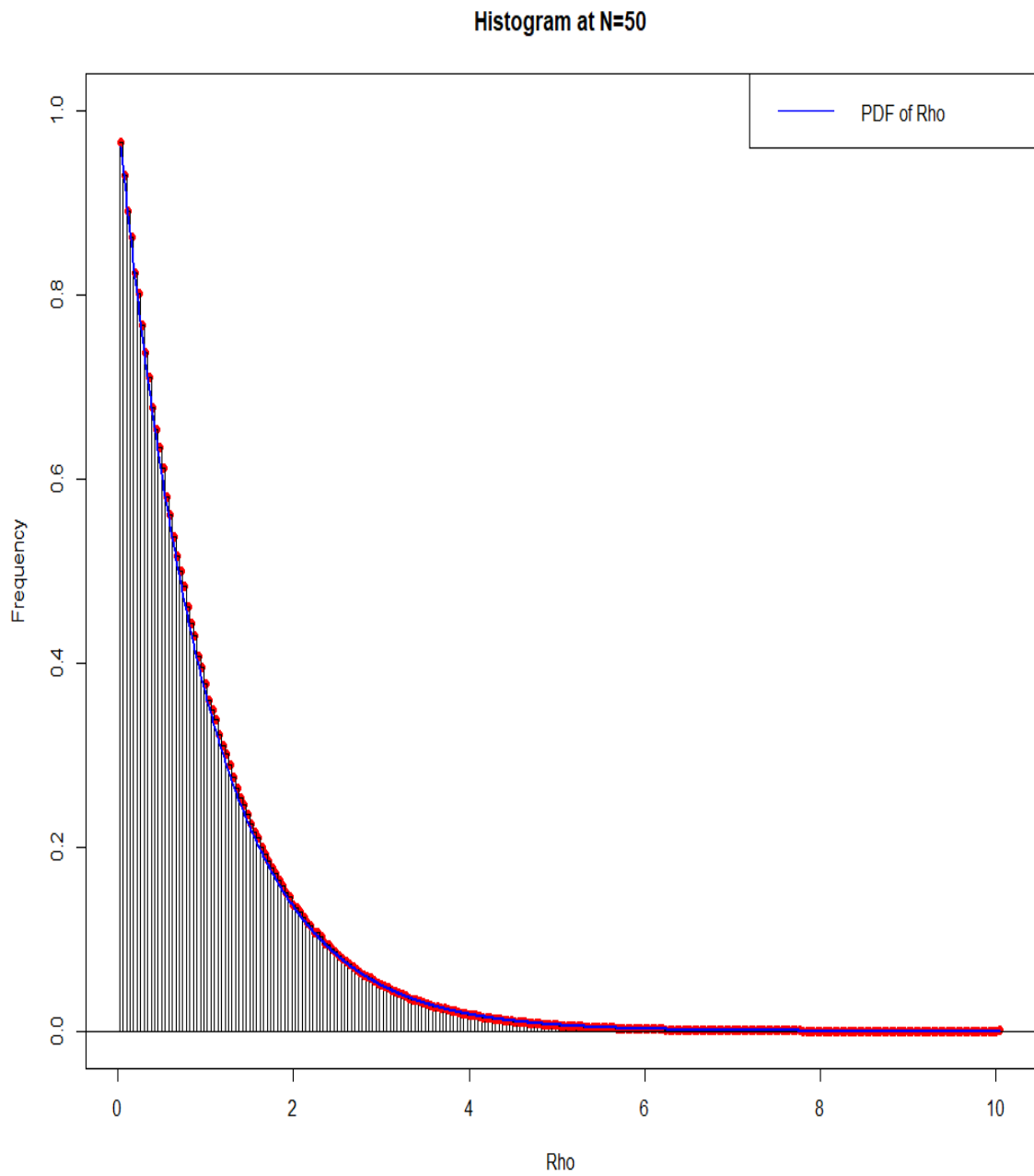


Figure 3.4: Generated histogram for the standardized square radius using mixing density with $m = 9$ by the alternative approach. Red points are the mid points of histogram bins.

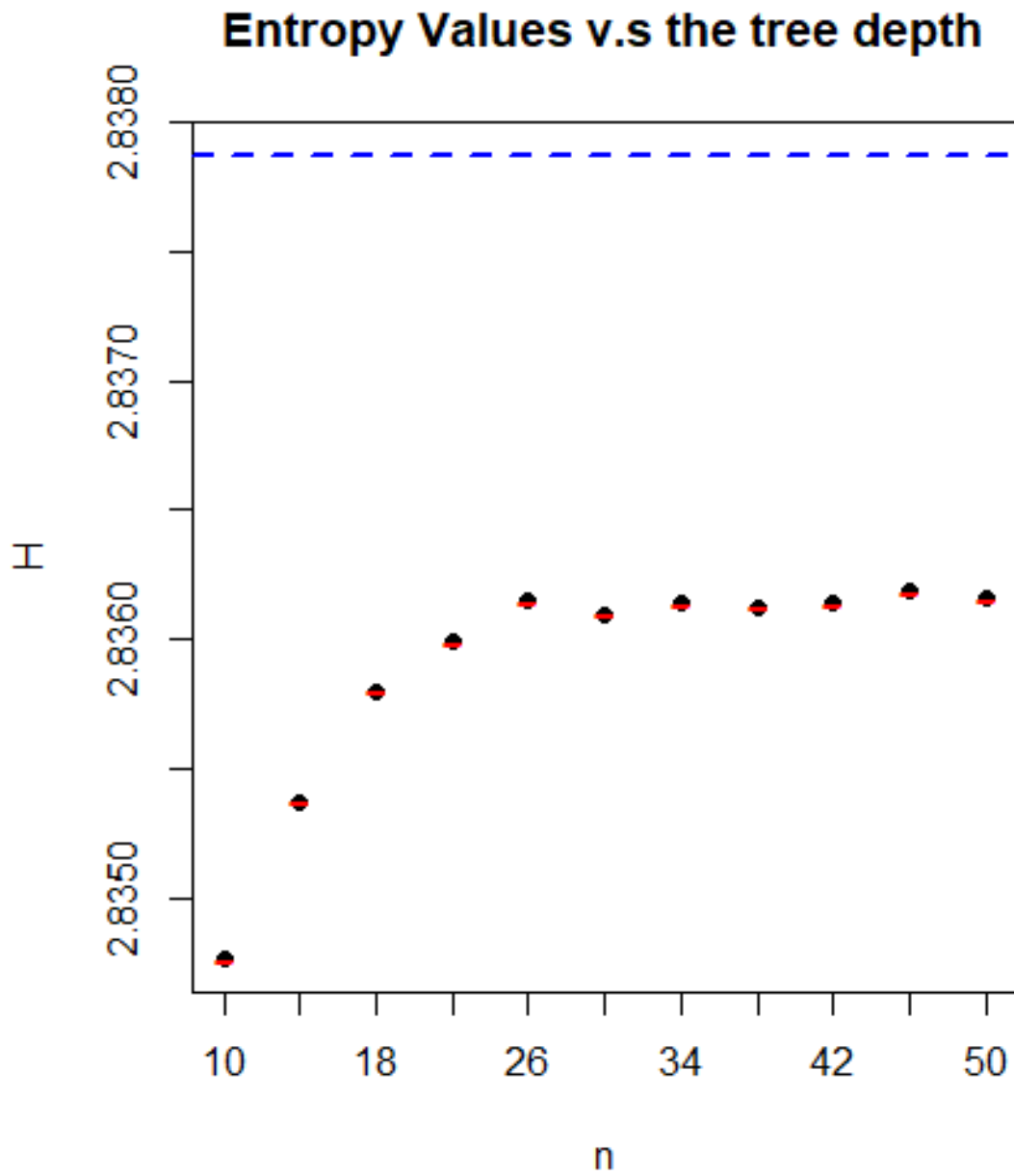


Figure 3.5: Entropy values versus the tree depth. The dotted blue line represents the entropy value of standard bivariate Gaussian density at 2.83788.

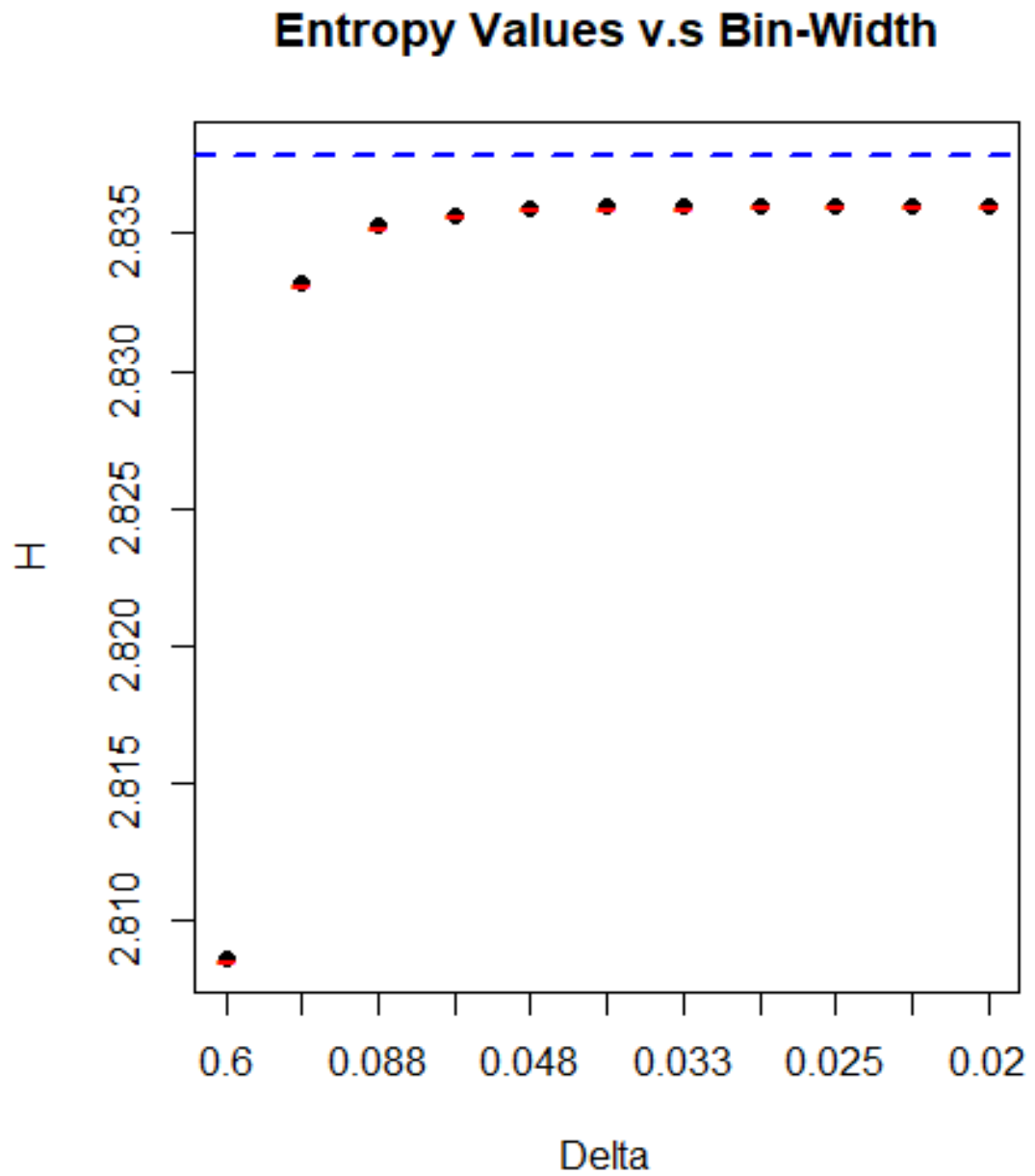


Figure 3.6: Entropy values versus bin-width. The dotted blue line represents the entropy value of standard bivariate Gaussian density at 2.8378.

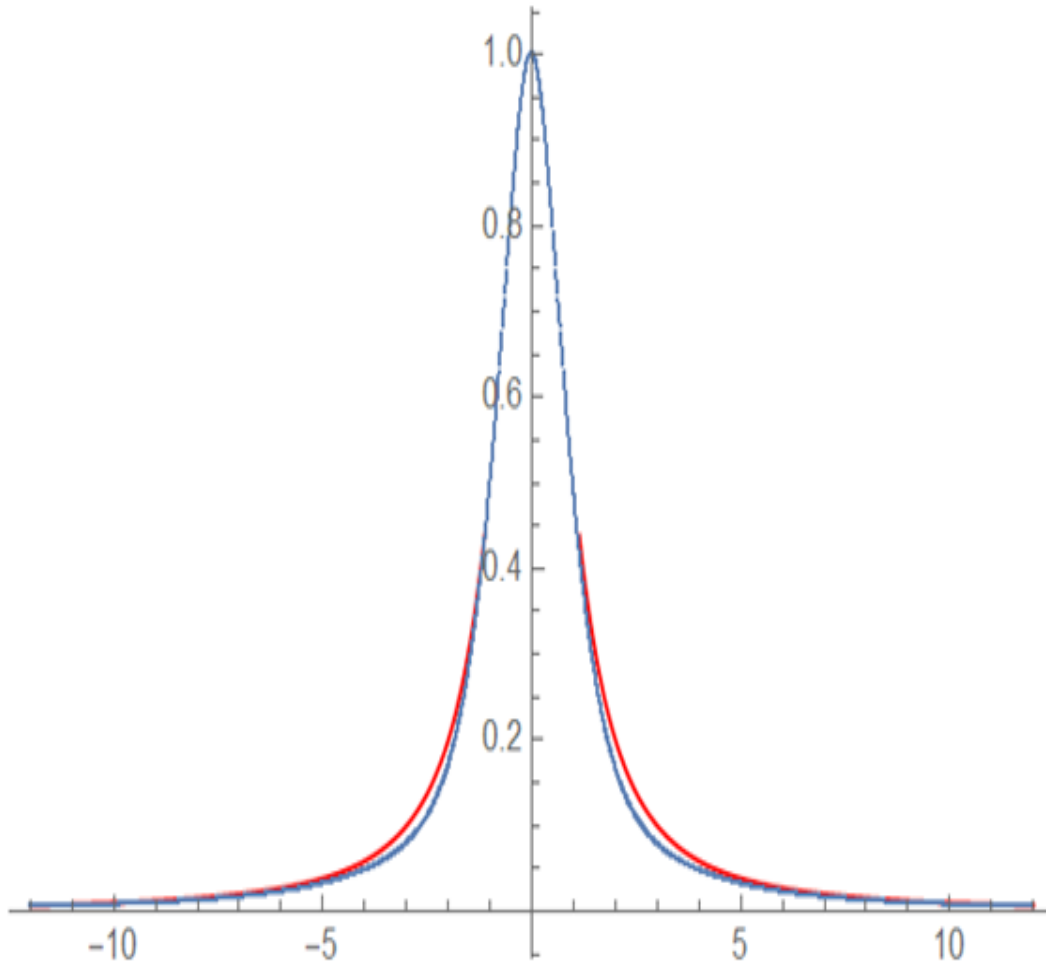


Figure 3.7: The characteristic function at $N = 10$ and $m = 9$. Red curve: The ideal characteristic function. Blue curve: Approximated characteristic function of the standardized squared radius.

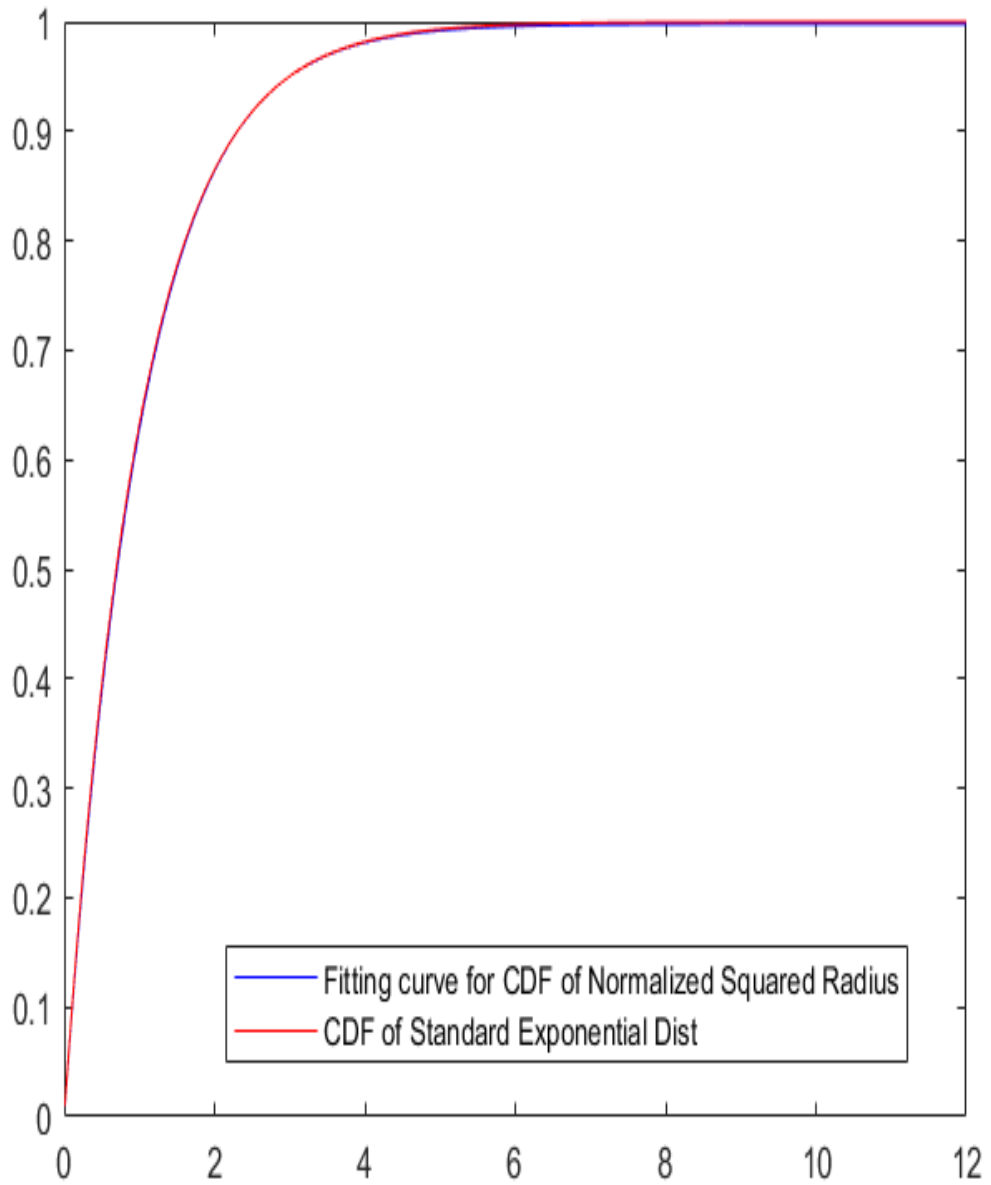


Figure 3.8: The cdf of the standardized squared radius versus the ideal cdf at $N = 30$ and $m = 30$. Blue curve: Approximated cdf of standardized squared radius. Red curve: Ideal cdf.

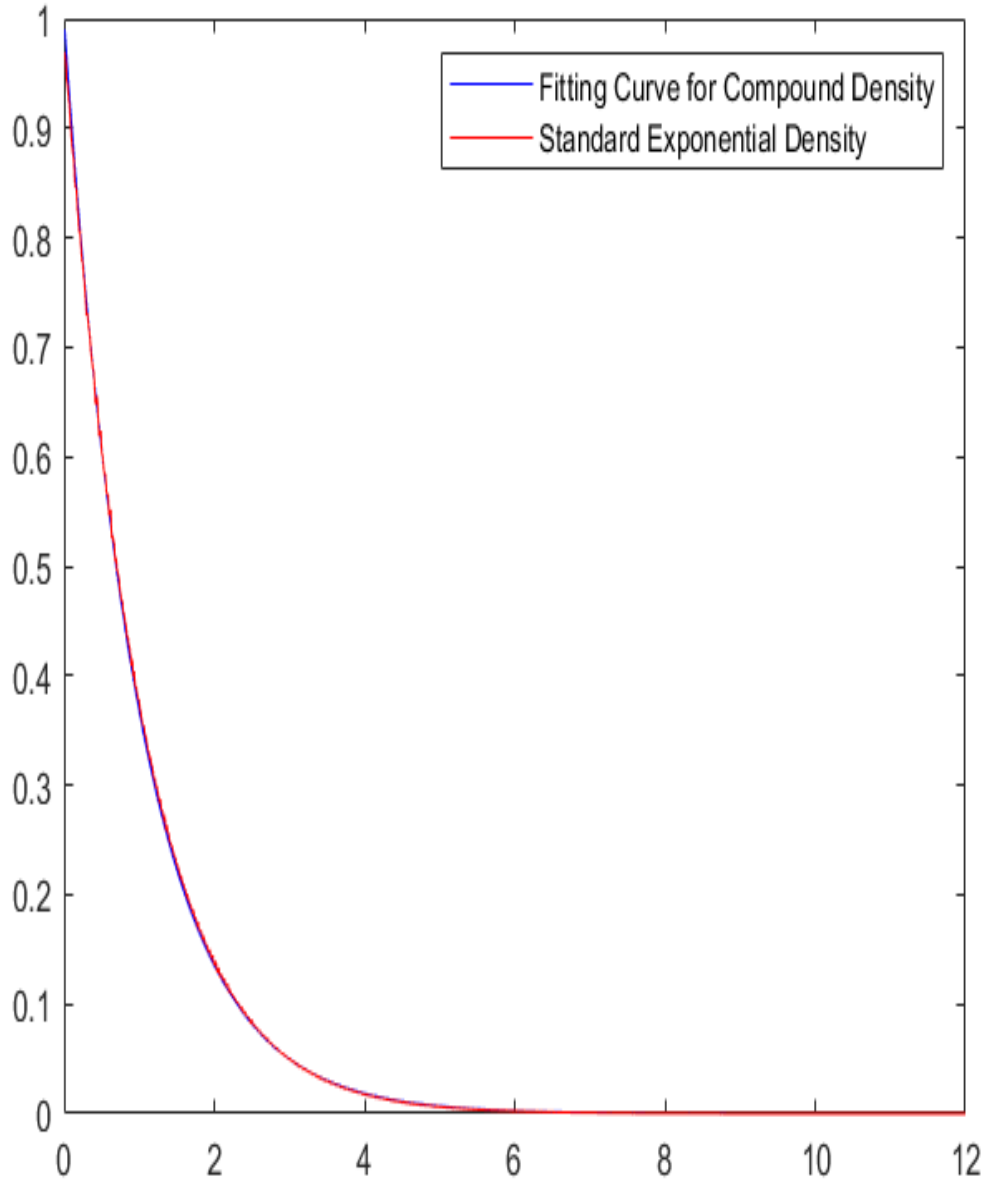


Figure 3.9: Compound density versus standard exponential density of the standardized squared radius at $N = 30$ and $m = 30$. Blue curve: Approximated compound density. Red curve: Ideal density.

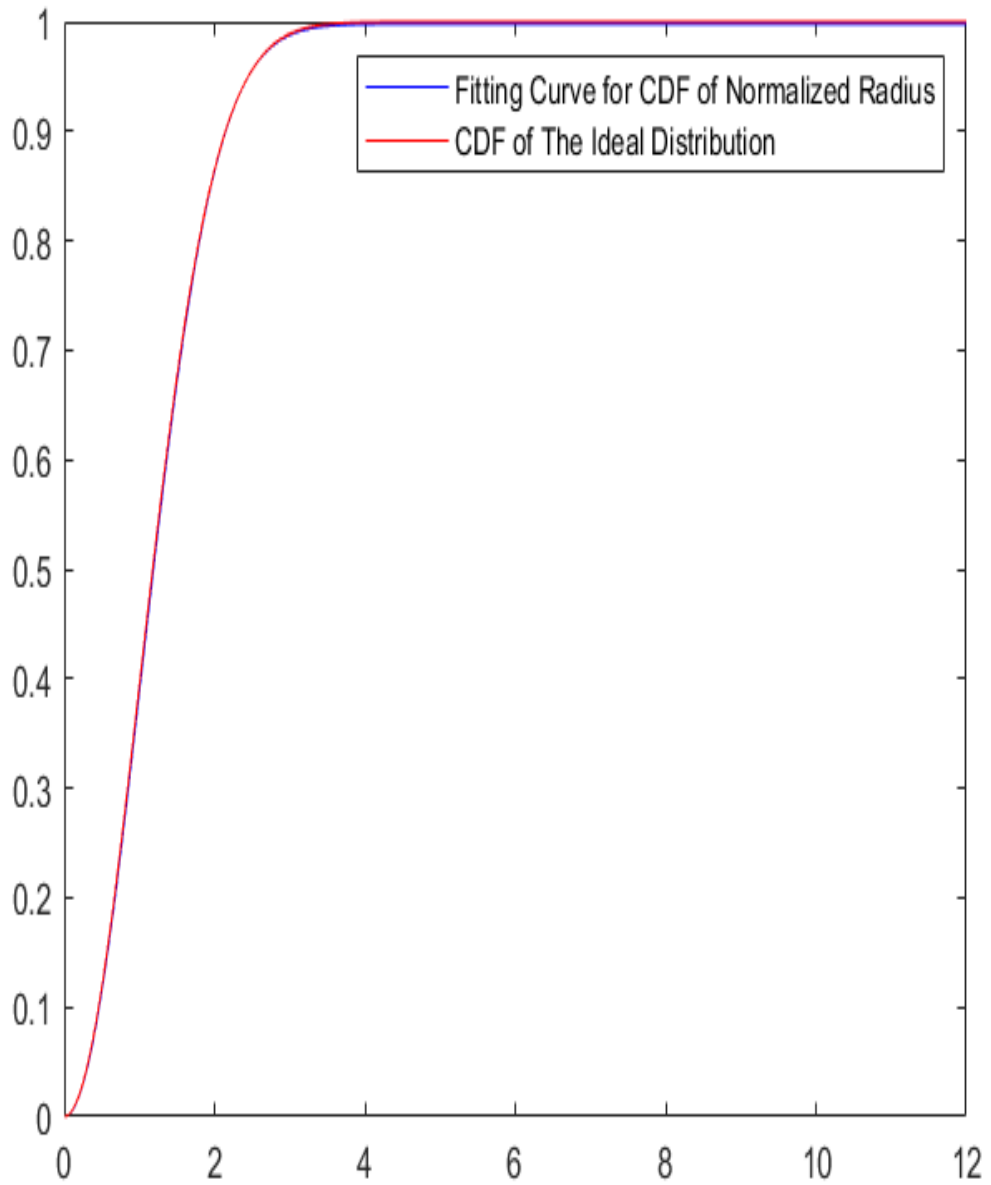


Figure 3.10: The cdf of standardized radius versus the ideal cdf at $N = 30$ and $m = 30$.

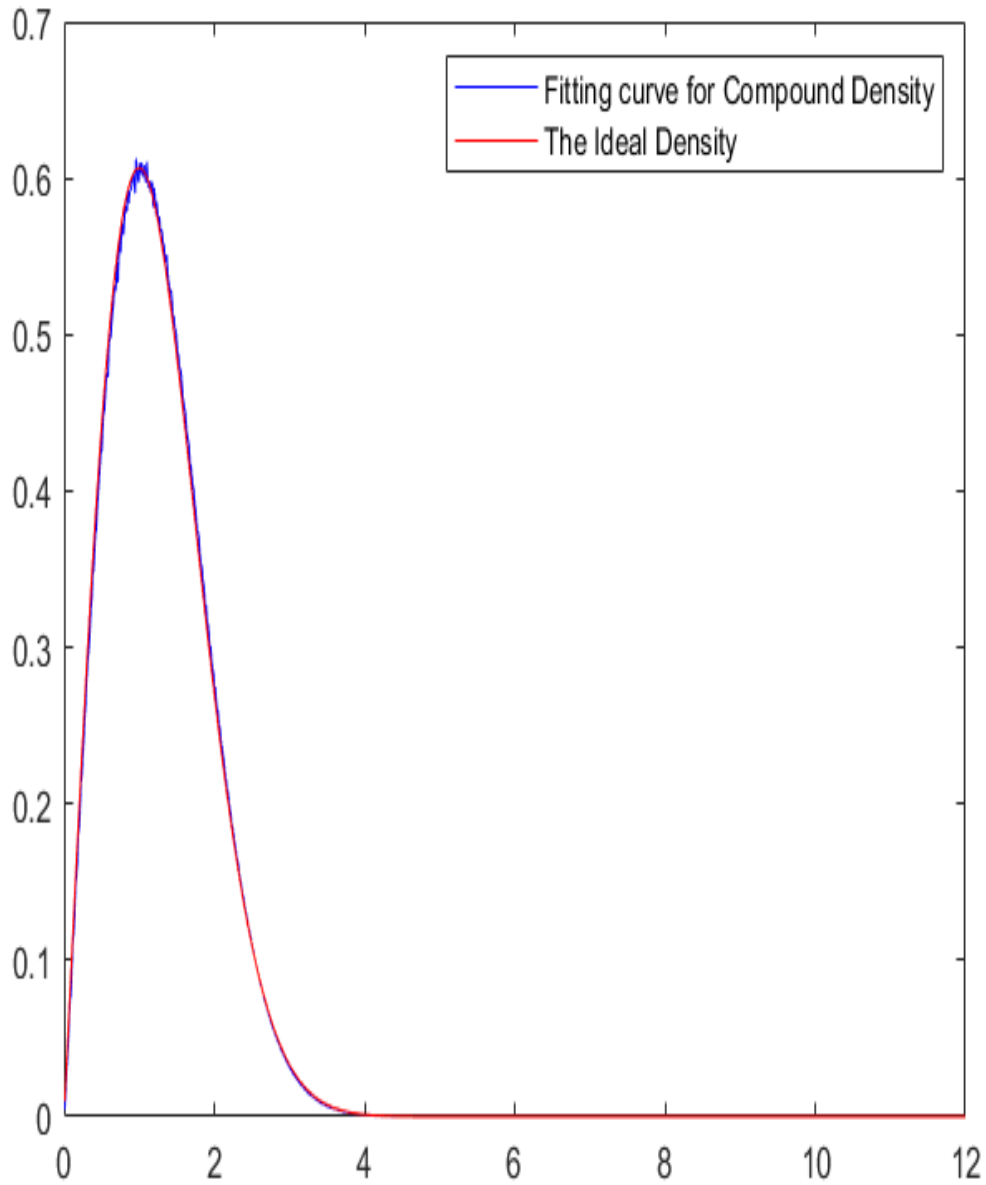


Figure 3.11: Compound density versus the ideal density of the standardized radius at $N = 30$ and $m = 30$

3.12.3. Convex Combination and Unilateral Optimization of Mixing Densities with degree from 1 to 30

In this section, we use the following inputs: range=20, numbers of bins=80 in experiments. Figure 3.12 shows the graphical representation of optimal coefficients α_i of the corresponding mixing densities with a convex combination of mixing densities from $m = 1$ to 30 when fixing $N = 10$. It is observed that the optimal coefficient is $\alpha_{30} = 0.89797246$. The entropy after optimization is 2.829718, compared with the highest entropy of $m = 30$ at 2.829633. There is possibly insignificant improvement in the entropy of 10^{-3} .

3.12.4. Convex Combination and Unilateral Optimization of Symmetric Monomials

In this section, we use the following inputs: range=20, numbers of bins=80 in experiments. Random sampling generates $u_2 = 3$ and $v_6 = 2$. Figure 3.13 shows the graphical representation of optimal coefficients α_i of symmetric monomials at $k = 11$ when fixing $N = 10$. It is observed that the optimal coefficient is $\alpha_{66} = 0.8898784$, followed by $\alpha_{19} = 0.1101216$. The entropy after optimization is obtained at 2.829364, compared with the entropy of the symmetric monomial 11, 11, 11 at 2.829279. There is possibly insignificant improvement in the entropy of 10^{-3} .

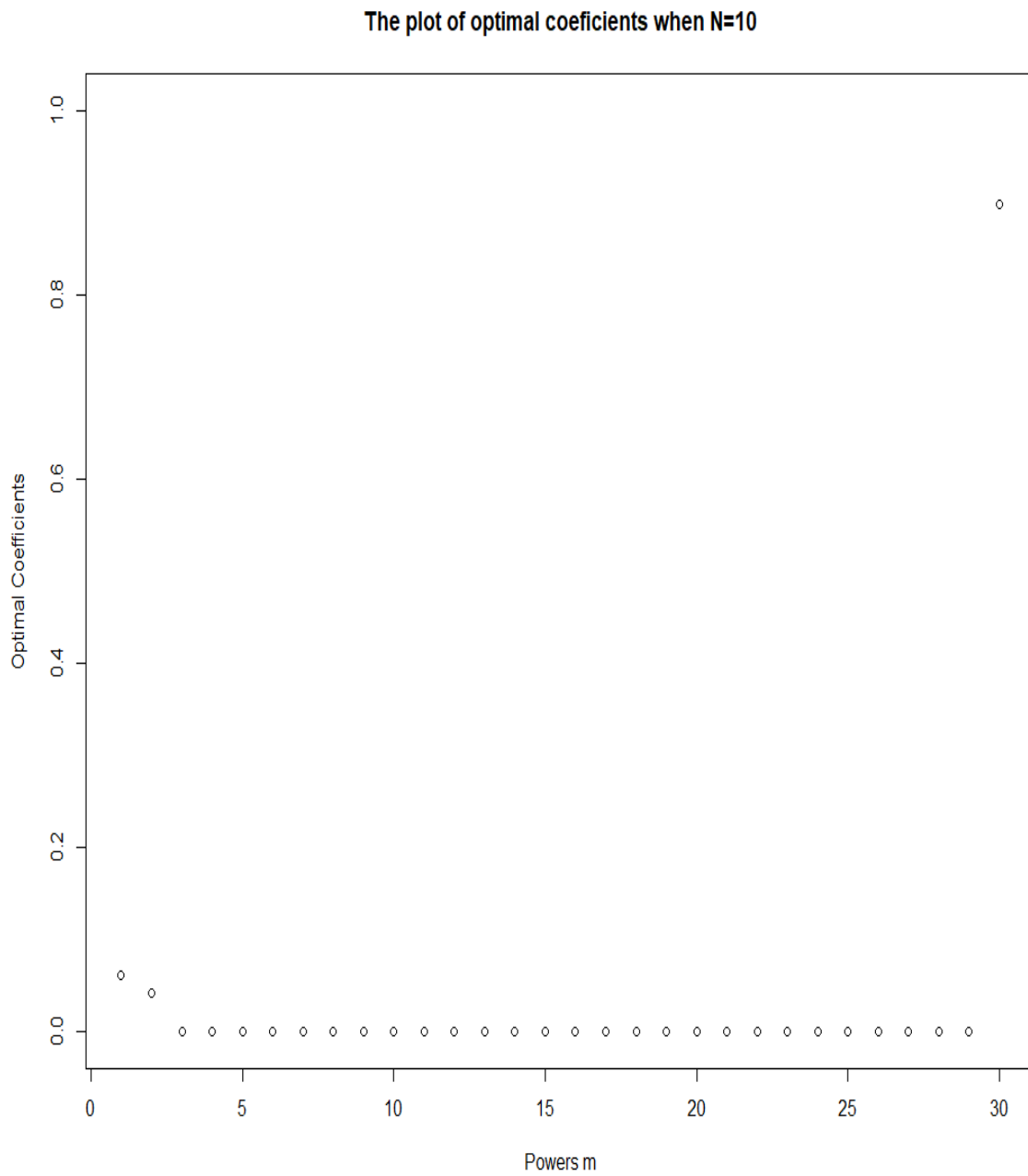


Figure 3.12: The plot of optimal coefficients in the convex combination of mixing densities when $N = 10$.

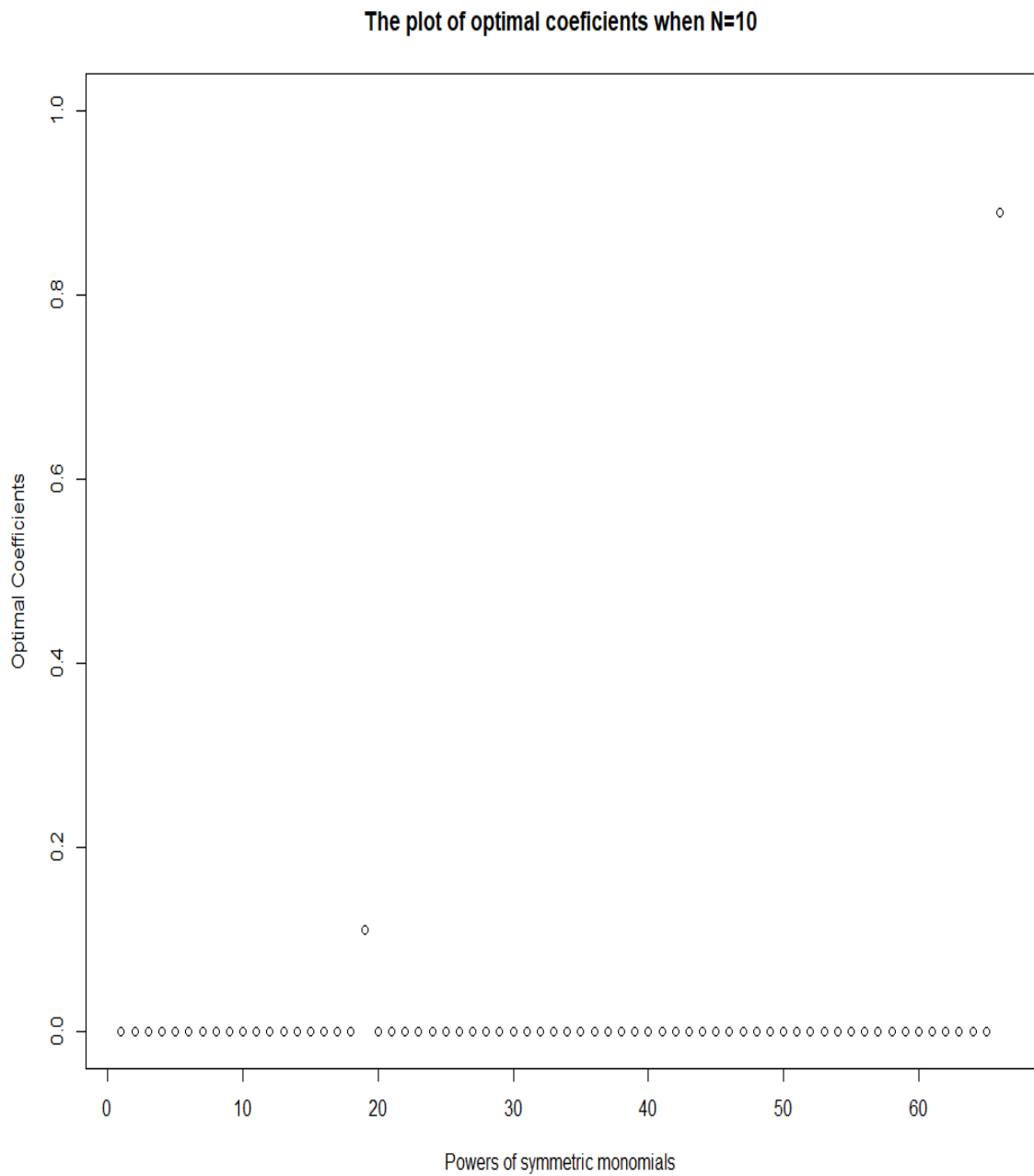


Figure 3.13: The plot of optimal coefficients in the convex combination of symmetrized monomials when $N = 10$.

3.12.5. Numerical Results of European Options on the Maximum and Minimum of Two Assets

We will present numerical results of MC-Tree to the call options on the maximum of assets and the put options on the minimum of the assets. Then we will compare MC-Tree with the plain MC, and recombining multinomial trees based on Pascal's simplex in [40] in terms of accuracy. The following parameters are used through all numerical experiments.

- Number of replication $M = 10^4$.
- Initial stock prices $S_1 = 40$, $S_2 = 40$.
- Expiration $T = 7$ months.
- Risk-free rate $r = 0.05$.
- Volatility $\sigma_1 = 0.2$, $\sigma_2 = 0.3$.
- The correlation $\rho = 0.5$.

3.12.5.1. COMPARISON TO THE RECOMBINING MULTINOMIAL TREE BASED ON PASCAL'S SIMPLEX

Table 3.1 and table 3.2 show that the standard deviation of Q&P is higher than the standard deviation of MC-Tree. Q&P method is from the recombining multinomial tree based on Pascal's simplex (see [40]). The price of both methods are convergent to the analytical solution at the tree depth $N=200$. The analytical solution (AS) is stated by Stulz (1982) in [42]. We notice the small upward bias on Table 3.2. However, the distribution correction technique is not applied here, which may cause this bias because the compound density is not quite Gaussian.

Table 3.1: Accuracy comparison between MC-Tree and Q&P for the European call options on the maximum of the assets.

K	Method	$N = 20$		$N = 50$		$N = 200$			AS
		Mean	SD	Mean	SD	Mean	SD	CI	
35	MC-Tree	9.428	0.064	9.422	0.040	9.420	0.02	(9.4194, 9.4202)	9.420
	Q&P	9.425	0.076	9.421	0.047	9.420	0.023	(9.4191, 9.4200)	
40	MC-Tree	5.501	0.084	5.492	0.053	5.488	0.026	(5.4878, 5.4888)	5.488
	Q&P	5.499	0.101	5.492	0.063	5.488	0.031	(5.4877, 5.4889)	
45	MC-Tree	2.801	0.081	2.797	0.050	2.795	0.025	(2.7943, 2.7953)	2.795
	Q&P	2.801	0.122	2.797	0.077	2.795	0.038	(2.7945, 2.7959)	

Table 3.2: Accuracy comparison between MC-Tree and Q&P for the European put options on the minimum of the assets.

K	Method	$N = 20$		$N = 50$		$N = 200$			AS
		Mean	SD	Mean	SD	Mean	SD	CI	
35	MC-Tree	1.395	0.036	1.390	0.022	1.388	0.011	(1.3881, 1.3886)	1.387
	Q&P	1.392	0.052	1.389	0.032	1.388	0.016	(1.3876, 1.3882)	
40	MC-Tree	3.815	0.051	3.805	0.031	3.801	0.016	(3.8002, 3.8008)	3.798
	Q&P	3.815	0.062	3.806	0.038	3.801	0.019	(3.8001, 3.8009)	
45	MC-Tree	7.512	0.044	7.505	0.027	7.501	0.013	(7.5010, 7.5015)	7.500
	Q&P	7.512	0.075	7.506	0.046	7.502	0.023	(7.5011, 7.5020)	

Table 3.3: Accuracy comparison between MC-Tree and plain MC for the European call options on the maximum of the assets.

K	Method	Mean	SD	AS
35	MC-Tree	9.422	0.040	9.420
	MC	9.345	7.441	
40	MC-Tree	5.492	0.053	5.488
	MC	5.435	6.471	
45	MC-Tree	2.797	0.050	2.795
	MC	2.750	4.963	

Table 3.4: Accuracy comparison between MC-Tree and plain MC for the European put options on the minimum of the assets.

K	Method	Mean	SD	AS
35	MC-Tree	1.390	0.022	1.387
	MC	1.400	2.615	
40	MC-Tree	3.805	0.031	3.798
	MC	3.808	4.29	
45	MC-Tree	7.505	0.027	7.500
	MC	7.480	5.565	

3.12.5.2. COMPARISON TO PLAIN MONTE CARLO (MC)

Table 3.3 and table 3.4 show that the standard deviation of the plain MC is higher than the standard deviation of MC-Tree at the same number of simulations $M = 10^4$ (the tree depth $N = 50$.)

3.12.6. Numerical Results for American Options on the Maximum and Minimum of Two Assets

We will present numerical results of MC-Tree to the American call options on the maximum of assets and the American put options on the minimum of the assets. Then we will compare MC-Tree with recombining multinomial trees based on Pascal's simplex in [40] in terms of accuracy.

The following parameters are used through all numerical experiments.

- Number of replication $M = 10^4$.
- Initial stock prices $S_1 = S_2 = S$.
- Strike price $K = 100$.
- Expiration $T = 1$ year.
- Risk-free interest rate $r = 0.05$.
- Volatility $\sigma_1 = 0.2$, $\sigma_2 = 0.2$.
- The correlation $\rho = 0.3$.
- Dividend rate $\delta = 0.10$.

Table 3.5 and table 3.6 show that the standard deviation of Q&P is higher than the standard deviation of MC-Tree.

Table 3.5: Accuracy comparison between MC-Tree and Q&P for the American call options on the maximum of assets.

S	Method	$N = 10$		$N = 50$	
		Mean	SD	Mean	SD
70	MC-Tree	0.204	0.069	0.235	0.032
	Q&P	0.235	0.137	0.241	0.066
80	MC-Tree	1.229	0.195	1.281	0.086
	Q&P	1.234	0.380	1.286	0.163
90	MC-Tree	4.155	0.341	4.181	0.151
	Q&P	4.161	0.519	4.184	0.237
100	MC-Tree	9.600	0.455	9.594	0.201
	Q&P	9.600	0.600	9.597	0.262
110	MC-Tree	17.306	0.471	17.301	0.208
	Q&P	17.292	0.556	17.302	0.246
120	MC-Tree	26.484	0.389	26.494	0.171
	Q&P	26.462	0.441	26.493	0.197
130	MC-Tree	36.372	0.292	36.400	0.128
	Q&P	36.351	0.332	36.397	0.148

Table 3.6: Accuracy comparison between MC-Tree and Q&P for the American put options on the minimum of assets.

S	Method	$N = 10$		$N = 50$	
		Mean	SD	Mean	SD
70	MC-Tree	37.919	0.089	37.887	0.033
	Q&P	37.918	0.117	37.888	0.043
80	MC-Tree	29.783	0.083	29.752	0.028
	Q&P	29.784	0.135	29.753	0.050
90	MC-Tree	21.967	0.096	21.931	0.033
	Q&P	21.961	0.191	21.932	0.073
100	MC-Tree	14.997	0.167	14.942	0.067
	Q&P	14.992	0.242	14.942	0.098
110	MC-Tree	9.403	0.215	9.335	0.090
	Q&P	9.393	0.265	9.334	0.104
120	MC-Tree	5.399	0.221	5.353	0.092
	Q&P	5.387	0.265	5.349	0.112
130	MC-Tree	2.842	0.196	2.835	0.082
	Q&P	2.821	0.287	2.829	0.124

Chapter 4

Credit Valuation Adjustment (CVA)

4.1. Unilateral CVA Formulation

Institution A, such as Bank, has a netting set with counterparty B at the expiration time at T , where T is the longest-maturity of all of the transactions in the netting set. Institution A is interested in determining the one-sided CVA for this netting set, excluding the possibility of their default.

Define:

- $V(t, T)$ is the risk-free value of the netting set at time t for Institution A.
- $\tilde{V}(t, T)$ is the risky value of the netting set at time t for Institution A.
- $\mathbb{E}_t[\cdot]$ is the conditional expectation under the risk-neutral measure taking account of the information available at time t .
- τ is the default time of the counterparty B.
- LGD is the loss given default to A of the counterparty B. $LGD = 1 - R$, where R is the recovery rate.

- $D(t, \tau)$ is the discount rate from τ back to t , $t \leq \tau$.

If there are no netting agreements, the risk-free value of the portfolio is the sum of individual trades or transactions.

If there are netting agreements, derivatives with negative values at the default time offset the ones with positive values within each netting set.

We can consider two cases.

Case 1: Counterparty does not default before T .

If the counterparty does not default before T , the risky asset is considered as the risk-free asset, and the payoff at time t is $\mathbb{1}_{\{\tau > T\}}V(t, T)$.

Case 2: Counterparty defaults before T .

If counterparty defaults at τ , the payoff is the sum of the value of the payments received before default time τ , $\mathbb{1}_{\{\tau \leq T\}}V(t, \tau)$, and the recovery value upon the default. The MtM recovery value upon the default is defined as $\mathbb{1}_{\{\tau \leq T\}}(R \cdot \max(V(\tau, T), 0) + \min(V(\tau, T), 0))$.

If the MtM value of the trade is positive at default, the bank will receive a portion of the trade value; the bank pays an amount to the counterparty if it is negative.

Therefore, the total payoff for the risky asset is the sum of the values in the two cases, under a risk-neutral assumption. Let $\bar{V}(\tau, T) = D(t, \tau)V(\tau, T)$ be the discounted value, back to time t , of the random variable $V(\tau, T)$.

Then $\bar{V}(t, T) = D(t, t)V(t, T) = V(t, T)$ as $D(t, t) = 1$.

$$\begin{aligned}
 \mathbb{E}_t[\mathbb{1}_{\tau \leq T}(\bar{V}(t, \tau) + \bar{V}(\tau, T))] &= \mathbb{E}[\mathbb{E}_t[(\bar{V}(t, \tau) + \bar{V}(\tau, T)) | \tau \leq T]] \\
 &= \mathbb{E}[\mathbb{E}_t[\bar{V}(t, T) | \tau \leq T]] = \bar{V}(t, T)\mathbb{E}[\mathbb{1}_{\{\tau \leq T\}}] \\
 &= V(t, T)\mathbb{E}[\mathbb{1}_{\{\tau \leq T\}}].
 \end{aligned}$$

We have

$$\begin{aligned}
\tilde{V}(t, T) &= \mathbb{E}[\mathbb{1}_{\{\tau \leq T\}} \bar{V}(t, \tau) + \mathbb{1}_{\{\tau \leq T\}} (R \cdot \max(\bar{V}(\tau, T), 0) + \min(\bar{V}(\tau, T), 0)) \\
&\quad + \mathbb{1}_{\{\tau > T\}} \bar{V}(t, T)] \\
&= \mathbb{E}[\mathbb{1}_{\{\tau \leq T\}} \bar{V}(t, \tau) + \mathbb{1}_{\{\tau \leq T\}} (R \cdot \max(\bar{V}(\tau, T), 0) + \bar{V}(\tau, T) - \max(\bar{V}(\tau, T), 0)) \\
&\quad + \mathbb{1}_{\{\tau > T\}} V(t, T)] \\
&= \mathbb{E}[\mathbb{1}_{\{\tau \leq T\}} (\bar{V}(t, \tau) + \bar{V}(\tau, T)) + \mathbb{1}_{\{\tau \leq T\}} ((R - 1) \max(\bar{V}(\tau, T), 0)) \\
&\quad + \mathbb{1}_{\{\tau > T\}} V(t, T)] \\
&= \mathbb{E}[(\mathbb{1}_{\{\tau > T\}} V(t, T) + \mathbb{1}_{\{\tau \leq T\}} V(t, T)) + \mathbb{1}_{\{\tau \leq T\}} ((R - 1) \max(\bar{V}(\tau, T), 0))] \\
&= \mathbb{E}[V(t, T) + \mathbb{1}_{\{\tau \leq T\}} ((R - 1) \max(D(t, \tau) V(\tau, T), 0))] \\
&= V(t, T) + \mathbb{E}[D(t, \tau) \mathbb{1}_{\{\tau \leq T\}} ((R - 1) \max(V(\tau, T), 0))].
\end{aligned}$$

CVA is the difference between the CCR-risk free portfolio value and the CCR-risky portfolio value when the counterparty may default. CVA at time t is defined as

$$CVA(t) = V(t, T) - \tilde{V}(t, T) = \mathbb{E}[D(t, \tau)(1 - R) \mathbb{1}_{\{t < \tau \leq T\}} \max(V(\tau, T), 0)].$$

CVA formula can be derived in another way. The bank will suffer from a loss in the default event of a counterparty at time $\tau \leq T$, as follows:

$$L = D(t, \tau)(1 - R) \mathbb{1}_{\{\tau \leq T\}} \max(V(\tau, T), 0).$$

CVA at time t is defined as the expectation of the loss.

$$CVA(t) = \mathbb{E}[L] = \mathbb{E}[D(t, \tau)(1 - R) \mathbb{1}_{\{t < \tau \leq T\}} \max(V(\tau, T), 0)].$$

It is difficult for us to calculate the formula $CVA(t)$ above due to some reasons. The default can happen at any time between t and T because time is continuous. $LGD = 1 - R$ is stochastic in general and correlated with other variables.

We can make some standard assumptions. Assume that LGD and discount factors are nonrandom. We assume the possible default event of counterparty and the value of a netting set $V(t, T)$ are uncorrelated, which is an assumption of unilateral CVA. We divide the interval $[t, T]$ into N subintervals, which are not necessarily equally spaced. Given the set $\{t_0, t_1, t_2, \dots, t_N\}$, where $t_0 = t$, $t_N = T$.

$$\begin{aligned} CVA(t) &\approx (1 - R)\mathbb{E}_t^{\mathcal{Q}}\left[\sum_{i=1}^N \mathbb{1}_{\{t_{i-1} < \tau \leq t_i\}} D(t, t_i) \max(V(t_i, T), 0)\right] \\ &= (1 - R)D(t, t_i) \sum_{i=1}^N \mathbb{E}_t^{\mathcal{Q}}[\mathbb{1}_{\{t_{i-1} < \tau \leq t_i\}}] \mathbb{E}_t^{\mathcal{Q}}[\max(V(t_i, T), 0)] \\ &= (1 - R)D(t, t_i) \sum_{i=1}^N P(t_{i-1} < \tau \leq \tau_i) \mathbb{E}_t^{\mathcal{Q}}[\max(V(t_i, T), 0)], \end{aligned}$$

where $P(t_{i-1} < \tau \leq \tau_i)$ is the risk neutral probability of default between t_{i-1} and t_i observed at time t . If we assume that a possible default of counterparty B between t_{i-1} and t_i occurs at the middle time and approximates the discounted positive exposure by the average value corresponding to the extremes of the interval, $CVA(t)$ is defined as $CVA(t) = (1 - R) \sum_{i=1}^N P(t_{i-1} < \tau \leq t_i) \left(\frac{D(t, t_{i-1}) \mathbb{E}_t[\max(V(t_{i-1}, T), 0)] + D(t, t_i) \mathbb{E}_t[\max(V(t_i, T), 0)]}{2} \right)$.

4.2. Default Probability

The first approach is to use an intensity default model. We will use an intensity default model (see [12]) to calculate the default probability of a counterparty. If the counterparty survived up to time t , the default probability in an infinitesimal interval $[t, t + dt]$ is defined as $P(t < \tau \leq t + dt) = \lambda_t dt$, where λ_t is a deterministic default intensity or the hazard

rate. The cumulative distribution function of default time τ of the counterparty is derived from a poisson process, $\Phi(t) = P(\tau \leq t) = 1 - e^{-\int_0^t \lambda_u du}$. We can assume a constant hazard rate or a deterministic piece-wise constant rate. The default probability between t_{i-1} and t_i is given by

$$P(t_{i-1} < \tau \leq t_i) = P(\tau \leq t_i) - P(\tau < t_{i-1}) = e^{\lambda t_{i-1} t_{i-1}} - e^{\lambda t_i t_i}.$$

An equivalent method is to use CDS spread or bond credit spread.

$$P(t_{i-1} < \tau \leq t_i) = \max(0, e^{-\frac{s_{i-1} t_{i-1}}{(1-R)}} - e^{-\frac{s_i t_i}{(1-R)}}),$$

where s_i is the credit spread of the counterparty at t_i . The expression $\frac{s_i}{(1-R)}$ approximates the instantaneous (risk-neutral) intensity of default. We apply a max function to avoid negative probabilities under some rare cases.

4.3. Expected Exposure

The exposure is the amount of loss if a counterparty cannot make future payments upon the option contract obligation due to its default at a particular time. The main part of CVA calculations is the calculations of the expected future exposure $\mathbb{E}_t^{\mathbb{Q}}[\max(V(t_i, T), 0)]$. The exposure calculations depend highly on the complexity of the derivatives in the portfolio. Modelling CVA of an American option is a challenge due to the complexity of CVA calculations and the characteristics of the American option. We present an algorithm for the calculation of the expected exposure in the formula of the unilateral CVA for the American put option using MC-Tree in the next section. Our method of calculations for the CVA of an American option is not known in the literature, to the best knowledge of the authors.

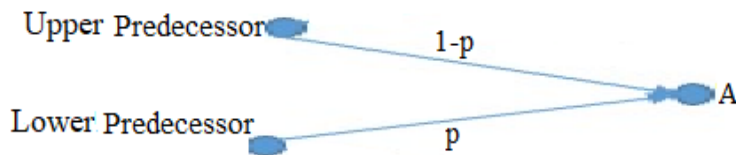
4.3.1. Tree Approach for CVA

We develop the algorithm to calculate CVA on an American put option, using the MC-tree method, as follows:

Step 1: Run the tree backward to compute the American option value at each node. Label each node either C for continuation or N for no continuation. Let

$$\mathbb{1}_C = \begin{cases} 1 & \text{if node label = "C"}, \\ 0 & \text{if node label = "N"}. \end{cases}$$

Step 2: Run the tree forward to compute the probabilities of the American option reaching a given node.



$$P(A) = p * P(\text{Lower Predecessor}) \mathbb{1}_C(\text{Lower Predecessor}) \\ + (1 - p) * P(\text{Upper Predecessor}) \mathbb{1}_C(\text{Upper Predecessor}).$$

Step 3: Run the tree backward to compute the expected exposure at time step in the tree using the probabilities in step 2.

4.3.2. MC-Tree Approach for CVA

The tree approach for CVA will produce a CVA value for any given tree. The CVA value from the MC-Tree approach is the mean of all CVA values of all trees.

4.4. Remarks

- We can short two American options in the portfolio, which means that there are negative values in the netting set. Hence, if there are two American options in the portfolio, we need to expand the MC-Tree approach for CVA to take account of possible cases: 1C & 2C, 1N & 2N, 1C & 2N, 1N & 2C, where 1C & 2C denote the continuation for the first option and the second option, respectively; 1N & 2N denote no continuation for the first option and the second option, respectively.
- If there are a larger number of American options in the portfolio (netting set), CVA calculations become computationally intensive. If all assets in the netting set are non-negative, we can just compute the probabilities using a tree for each asset separately, then add up to get the total result.
- For a portfolio with only European options, the probabilities in the tree are unchanged for European options, and we can do the netting set for each node without taking account of cases "alive" or "dead" for each option.

4.5. Numerical Results of CVA Calculations

The following parameters are used to estimate CVA of an American put option.

Initial stock price $S_0=80$, Strike price $K = 100$, Expiration $T = 1$, Interest rate $r = 0.03$,

Volatility $\sigma = 0.2$, Dividend rate=0, Recovery rate $R = 0.4$, Intensity of default $\lambda = 0.03$.

CVA value is convergent to 0.34 when we increase the number of simulation M , and the

Table 4.1: CVA values.

N	M	CVA
50	100	0.2447
75	150	0.3013
100	250	0.3104
250	700	0.3282
250	10000	0.3392
250	100000	0.3440
2000	700	0.3414
4000	700	0.3414

tree depth N , respectively.

Chapter 5

Entropy and Kullback-Leibler formulas for rational densities

5.1. Introduction

We consider integrable univariate rational density functions on the real line, then refer to them as rational densities henceforth. After normalization, we have rational probability density functions. This chapter presents an analytic method to compute the entropy of rational densities and the KL divergence between two rational densities. The entropy and KL divergence are defined by the usual entropy and KL divergence integrals. The method will give exact answers only if the poles are known or can be computed exactly. If the spectral factorization is done by an approximative numerical procedure, the answer will also be an approximation. There are many works on spectral factorization in the literature due to its popularity. Here, we mention two noticeable approaches. The study in the article [24] presented five numerical algorithms for spectral factorization of Laurent polynomials. Authors in [4] showed an efficient algorithm based on Graeffe iteration for

polynomial spectral factorization.

Rational probability density functions play an important role in classical statistical inference, applications in science, and in time series analysis as spectral densities. Although the Gaussian distribution is dominant in applications of statistical modeling in many areas in which there is a growing interest in the so-called heavy-tailed distribution. Rational densities of such heavy-tailed distribution follow important subclass, other types of distributions have been emphasized in the field.

Hanzon and Ober in [25] investigated a solution to the filtering problem for a linear time-varying system whose noise inputs have rational probability density functions. They obtained important results on rational densities and representation of rational densities. The case of Cauchy noise is also treated in that paper. Hanzon and Scherrer [26] solved the filtering problem for a class of discrete-time stochastic volatility models. The disturbances have rational probability density functions such as a Cauchy density or a Student t-density with odd number of degrees of freedom. Entropy computation is crucial in probability theory, for instance, in entropy maximization problems over a class of densities. In the literature, the KL divergence is also applied to integrable spectral densities of finite-dimensional Gaussian systems (see [23]). Authors in [37] derived mathematical expressions for the Shannon entropy and cross-entropy of exponential families. The authors concluded that entropy and the cross-entropy of exponential families are not always available in a closed-form solution. Reference [17] presented a closed-form expression for the KL divergence between Cauchy distributions. However, the formula is derived from using the Mgfund package of the computer algebra system Maple to calculate a parametric definite integral with six parameters. The formula for the entropy of Student distribution is mentioned in [30]. However, the authors presented a numerical approximation for the exact expression because it is difficult to evaluate the effect of the number of degrees of freedom on the entropy from the formula.

Entropy and KL divergence between Cauchy densities and Student rational densities are crucial in many applications, but an analytical derivation is not available in the literature, to the best of our knowledge.

The chapter is organized as follows. Section 5.2 presents the main theorem which is used to derive other theorems, and propositions in this chapter. Then we apply those theorems to derive entropy and KL divergence of the compound density of an MC-Tree in section 5.3, between Cauchy densities in section 5.4, and between rational Student densities in section 5.5. Section 5.6 extends our approach to calculate entropy and KL divergence for any (pair of) integrable and non-negative rational function(s). Some examples are illustrated in subsections 5.6.1 and 5.6.2. Finally, some remarks on possible usage to finite Gaussian mixture are presented in section 5.7.

5.2. An Application of the Residue Theorem

Let us consider integrals of the type $\int_{-\infty}^{\infty} r(x) \log(q(x)) dx$, where $r(x)$ is any absolutely integrable rational function, and $q(x)$ is a non-negative, not identically zero rational function. Then $q(x)$ is equal to the quotient of two non-negative polynomials. Of course, $\log(q(x))$ can be written as the log of the numerator minus the log of the denominator. Therefore, we can concentrate on the case where we have a nonnegative polynomial in the log term. As is well-known and not difficult to show q is nonnegative if and only if there exist a complex polynomial $\psi(z)$ with all its zeros in the closed upper half plane (i.e. the imaginary parts of the zeros of $\psi(z)$ are non-negative) such that $q(x) = |\psi(x)|^2$ for all real values of x . Such ψ will be called a spectral factor. The rational function $r(z)$ when viewed as a function on the closed lower half plane in \mathbb{C} has singularities at its poles in the open lower half plane, say z_1, z_2, \dots, z_k , with $Im(z_j) < 0$, $j = \overline{1, k}$ with the corresponding multiplicities $\mu_1, \mu_2, \dots, \mu_k; \mu_j \in \mathbb{N}$, and $\sum_{j=1}^k \mu_j = n$, n is called the McMillan degree

(order) of $r(x)$ (see [25]). Our main theorem is as follows:

Theorem 5.2.1. *Let r , q , ψ be as just described. Let $r(x) \not\equiv 0$,*

$$\begin{aligned} \int_{-\infty}^{\infty} r(x) \log(q(x)) dx &= 2\operatorname{Re}\{(-2\pi i) \sum_{j=1}^k \operatorname{Res}(r(z) \log(\psi(z)), z_j)\} \\ &= 2\operatorname{Re}\left\{(-2\pi i) \sum_{j=1}^k \frac{1}{(\mu_j - 1)!} \lim_{z \rightarrow z_j} \frac{d^{\mu_j - 1}}{dz^{\mu_j - 1}} \left((z - z_j)^{\mu_j} r(z) \log(\psi(z)) \right)\right\}, \end{aligned}$$

where $\operatorname{Res}(r(z) \log(\psi(z)), z_j)$ denotes the residue of the function $r(z) \log(\psi(z))$ at $z = z_j$.

Proof. Note that q and hence ψ , may have real zeros. To deal with this, we consider $q_\epsilon = q + \epsilon$, $\epsilon > 0$. As $q_\epsilon(x) > 0 \forall x \in \mathbb{R}$ it has a spectral factor ψ_ϵ , say, with all its zeros in the open upper half plane. So $q_\epsilon = \psi_\epsilon \psi_\epsilon^* > 0$, $\forall x \in \mathbb{R}$. Its leading coefficient will be positive. We take ψ_ϵ to be the spectral factor with real positive leading coefficient. Then for $\epsilon \downarrow 0$, we have $\psi_\epsilon \rightarrow \psi$, as the leading coefficient of ψ_ϵ converges to the leading coefficient of ψ and the zero locations of ψ_ϵ converge to those of ψ (and ψ_ϵ, ψ are polynomials, hence completely determined by their zero locations and their leading coefficient).

The residue formula is continuous in terms of the leading coefficient and the zeros of ψ , hence it will suffice to prove the formula for the case in which $q(x) > 0$, $\forall x \in \mathbb{R}$.

So assume $q(x) > 0$, $\forall x \in \mathbb{R}$. Then $q(x) = \psi(x)\psi^*(x)$, $\forall x \in \mathbb{R}$, $\psi(z)$ polynomial with real positive leading coefficient and with all the zeros of ψ in the open upper half plane.

Now note that

$$\begin{aligned} \int_{-\infty}^{\infty} r(x) \log(q(x)) dx &= \int_{-\infty}^{\infty} r(x) \log(|\psi(x)|^2) dx = \int_{-\infty}^{\infty} 2r(x) \log(|\psi(x)|) dx \\ &= 2\operatorname{Re}\left\{\int_{-\infty}^{\infty} r(z) \log(\psi(z)) dz\right\}. \end{aligned}$$

Now note that $\psi(z) \neq 0$, $\forall z \in \mathbb{G}$, where \mathbb{G} is an open simply connected set containing the closed lower half plane and is such that \mathbb{G} does not contain any of the zero locations

of ψ . (So for instance, if $m = \min\{Im(z); \psi(z) = 0\}$ then $m > 0$ and we can take $\mathbb{G} = \{z : Im(z) < m\}$). We claim that $\log(\psi(z))$ is analytic on \mathbb{G} . To be more specific, the log function is multivalued and determined up to an additive term of the form $k2\pi i$, $k \in \mathbb{Z}$. Let $d := deg(\psi)$, then $\lim_{x \rightarrow \infty, x \in \mathbb{R}} \frac{\psi(x)}{x^d} := \lambda > 0$, λ the leading coefficient of ψ . Taking the branch of the logarithm for which $\lim_{x \rightarrow \infty, x \in \mathbb{R}} Im(\log(\psi(x))) = 0$, we obtain a univalued analytical function $\log(\psi(z))$, $z \in \mathbb{G}$. The reason is that $\psi(z)$ is polynomial, hence analytic and $\psi(z) \neq 0, \forall z \in \mathbb{G}$. Hence, $\log(\psi(z))$ has no singularities on \mathbb{G} and \mathbb{G} is simply connected, so $\log(\psi(z))$ will indeed be univalued and analytic. As $0 < \int_{-\infty}^{\infty} |r(x)|dx < \infty$ (by assumption in the theorem), and $r(x)$ rational, it follows that $r(x)$ can be written as $r(x) = \frac{r_N(x)}{r_D(x)}$, where r_N, r_D are real polynomials and coprime, with $deg(r_D) \geq deg(r_N) + 2$. It now follows that we can apply the well-known $M-l$ method to our integral. For each $R > 0$, consider the complex integral $\oint_{C_R} r(z) \log(\psi(z))dz = \int_{-R}^R r(z) \log(\psi(z))dz + \int_{S_R} r(z) \log(\psi(z))dz$, where S_R is a semi-circle in the lower half plane with centre at the origin and radius R , and orientation of decreasing angle. Then C_R is a closed curve winding clockwise. We take $R > 0$ sufficiently large such that all poles (there are finitely many) of $r(z)$ in the open lower half plane have modulus less than R . Then the residue theorem can be applied to $\oint_{C_R} r(z) \log(\psi(z))dz$. (Note that the clockwise orientation, gives an extra minus sign). To show that $\oint_{C_R} r(z) \log(\psi(z))dz = \int_{-\infty}^{\infty} r(z) \log(\psi(z))dz$, we use the $M-l$ method to show that $\lim_{R \rightarrow \infty} \int_{S_R} r(z) \log(\psi(z))dz = 0$. The argument runs as follows.

$$\begin{aligned} \lim_{R \rightarrow \infty} \left| \int_{S_R} r(z) \log(\psi(z))dz \right| &\leq \lim_{R \rightarrow \infty} (\pi R) \text{Max}_{z \in S_R} |r(z) \log(\psi(z))| \\ &= \pi \lim_{R \rightarrow \infty} \text{Max}_{z \in S_R} |z r(z) \log(\psi(z))| \\ &\leq \pi \lim_{R \rightarrow \infty} \text{Max}_{z \in S_R} |z^2 r(z)| \cdot \lim_{R \rightarrow \infty} \text{Max}_{z \in S_R} \left| \frac{\log(\psi(z))}{z} \right| \\ &= 0. \end{aligned}$$

The last equality is due to the facts that $\lim_{R \rightarrow \infty} \text{Max}_{z \in S_R} |z^2 r(z)|$ is a finite limit, and

$$\lim_{R \rightarrow \infty} \text{Max}_{z \in S_R} \left| \frac{\log(\psi(z))}{z} \right| = 0.$$

So the residue formula can indeed be applied. \square

In the following this formula will be worked out further in various special cases.

Theorem 5.2.2. *Suppose that the monic polynomial $q(x)$ is nonnegative on the real line and has degree n (n is even). Then $q(z)$ has the spectral factor*

$\psi(z) = (z - z_1)(z - z_2) \dots (z - z_{n/2})$, where $z_k, k = \overline{1, \frac{n}{2}}$ are zeros in the upper half plane.

Let $l \in \mathbb{N}$, we have the general formula for $\int_{-\infty}^{\infty} \frac{\log(q(x))}{(x^2+1)^l} dx$, as follows:

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\log(q(x))}{(x^2+1)^l} dx &= \sum_{k=1}^{n/2} \left\{ \frac{2^{3-2l} \pi (2l-2)! (-1)^{2-2l}}{((l-1)!)^2} \log |z_k + i| \right. \\ &+ \pi \sum_{j=1}^{l-1} \left[\binom{2l-j-2}{l-1} \frac{1}{j} (-1)^{-2l+2j+1} 2^{-2l+j+3} \frac{1}{((1+b_k)^2 + a_k^2)^j} \right] \left[(1+b_k)^j \right. \\ &\left. \left. + \sum_{m=2, \text{even}}^j \binom{j}{m} (-1)^{m/2} (a_k)^m (1+b_k)^{j-m} \right] \right\}, \end{aligned}$$

where $z_k = a_k + ib_k, b_k > 0$.

Proof. We have

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\log(q(x))}{(x^2+1)^l} dx &= 2\text{Re} \left\{ \frac{-2\pi i}{(l-1)!} \lim_{z \rightarrow -i} \frac{d^{l-1}}{dz^{l-1}} \left[\frac{\log(\psi(z))}{(z-i)^l} \right] \right\} \\ &= \sum_{k=1}^{n/2} 2\text{Re} \left\{ \frac{-2\pi i}{(l-1)!} \lim_{z \rightarrow -i} \frac{d^{l-1}}{dz^{l-1}} \left[\frac{\log(z - z_k)}{(z-i)^l} \right] \right\}. \end{aligned}$$

Applying general Leibniz rule, we have

$$\begin{aligned}
\frac{d^{l-1}}{dz^{l-1}} \frac{\log(z - z_k)}{(z - i)^l} &= \sum_{j=0}^{l-1} \binom{l-1}{j} \frac{d^j}{dz^j} (\log(z - z_k)) \frac{d^{l-1-j}}{dz^{l-1-j}} ((z - i)^{-l}) \\
&= \log(z - z_k) (-1)^{l-1} l(l+1) \dots (l+l-2) (z - i)^{-2l+1} \\
&+ \sum_{j=1}^{l-1} \binom{l-1}{j} (-1)^{l-2} (j-1)! (z - z_k)^{-j} l(l+1) \dots (2l-j-2) (z - i)^{-2l+1+j} \\
&= \log(z - z_k) (-1)^{l-1} \frac{(2l-2)!}{(l-1)!} (z - i)^{-2l+1} \\
&+ \sum_{j=1}^{l-1} \frac{(l-1)!}{j!(l-1-j)!} (-1)^{l-2} (j-1)! (z - z_k)^{-j} \frac{(2l-j-2)!}{(l-1)!} (z - i)^{-2l+1+j} \\
&= \log(z - z_k) (-1)^{l-1} \frac{(2l-2)!}{(l-1)!} (z - i)^{-2l+1} \\
&+ \sum_{j=1}^{l-1} \binom{2l-j-2}{l-1} \frac{(l-1)!}{j} (-1)^{l-2} (z - z_k)^{-j} (z - i)^{-2l+1+j}.
\end{aligned}$$

We have

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{\log(q(x))}{(x^2 + 1)^l} dx &= \sum_{k=1}^{n/2} 2Re \left\{ \frac{-2\pi i}{(l-1)!} \lim_{z \rightarrow -i} \frac{d^{l-1}}{dz^{l-1}} \left[\frac{\log(z - z_k)}{(z - i)^l} \right] \right\} \\
&= \sum_{k=1}^{n/2} \left\{ 2Re \left\{ (-2\pi i) \left[\frac{\log(-i - z_k) (-1)^{l-1} (2l-2)! (-2i)^{-2l+1}}{((l-1)!)^2} \right. \right. \right. \\
&+ \left. \left. \left. \sum_{j=1}^{l-1} \left[\binom{2l-j-2}{l-1} \frac{1}{j} (-1)^{l-2} (-i - z_k)^{-j} (-2i)^{-2l+1+j} \right] \right] \right\} \right\} \\
&= \sum_{k=1}^{n/2} \left\{ 2Re \left\{ (-2\pi i) \left[\frac{\log(-i - z_k) (-1)^{-l} (2l-2)! (2i)^{-2l+1}}{((l-1)!)^2} \right. \right. \right. \\
&+ \left. \left. \left. \sum_{j=1}^{l-1} \left[\binom{2l-j-2}{l-1} \frac{1}{j} (-1)^{-l-1+j} (-i - z_k)^{-j} (2i)^{-2l+1+j} \right] \right] \right\} \right\}.
\end{aligned}$$

We have used

$$\begin{aligned}\frac{d^j}{dz^j} \log(z - z_k) &= (-1)^{j-1} (j-1)! (z - z_k)^{-j}. \\ \frac{d^{l-1-j}}{dz^{l-1-j}} [(z-i)^{-l}] &= (-1)^{l-1-j} l(l+1) \dots (2l-1-j-1) (z-i)^{-(2l-1-j)}.\end{aligned}$$

We consider the first part

$$\begin{aligned}2\operatorname{Re}\{(-2\pi i)\left[\frac{\log(-i - z_k)(-1)^{-l}(2l-2)!(2i)^{-2l+1}}{((l-1)!)^2}\right]\} &= \frac{2^{3-2l}\pi(2l-2)!(-1)^{2-2l}}{((l-1)!)^2} \operatorname{Re}[\log(-i - z_k)] \\ &= \frac{2^{3-2l}\pi(2l-2)!(-1)^{2-2l}}{((l-1)!)^2} \log|z_k + i|\end{aligned}$$

because $z_k = a_k + ib_k$, $\operatorname{Im}(z_k) > 0 \Rightarrow \operatorname{Im}(-z_k) < 0 \Rightarrow \operatorname{Im}(-i - z_k) < 0$.(1)

Next, we consider the second part

$$\begin{aligned}& 2\operatorname{Re}\left\{(-2\pi i)\left\{\sum_{j=1}^{l-1} \binom{2l-j-2}{l-1} \frac{1}{j} (-1)^{-l-1+j} (-i - z_k)^{-j} (2i)^{-2l+1+j}\right\}\right\} \\ &= 2\operatorname{Re}\left\{(-\pi)\left\{\sum_{j=1}^{l-1} \binom{2l-j-2}{l-1} \frac{1}{j} (-1)^{-l-1+j} \left(\frac{-i - z_k}{2i}\right)^{-j} (2i)^{-2l+2}\right\}\right\} \\ &= 2\operatorname{Re}\left\{(-\pi)\left\{\sum_{j=1}^{l-1} \binom{2l-j-2}{l-1} \frac{1}{j} (-1)^{-l-1+j} \left(-\frac{1}{2} + \frac{1}{2}iz_k\right)^{-j} (2)^{-2l+2} (-1)^{-l+1}\right\}\right\} \\ &= 2\operatorname{Re}\left\{(-\pi)\left\{\sum_{j=1}^{l-1} \binom{2l-j-2}{l-1} \frac{1}{j} (-1)^{-2l+j} \left(-\frac{1}{2} + \frac{1}{2}iz_k\right)^{-j} (2)^{-2l+2}\right\}\right\} \\ &= \pi \sum_{j=1}^{l-1} \left[\binom{2l-j-2}{l-1} \frac{1}{j} (-1)^{-2l+j+1} (2)^{-2l+3} \frac{1}{\left(\left(\frac{1}{2} + \frac{1}{2}b_k\right)^2 + \frac{1}{4}a_k^2\right)^j} \operatorname{Re}\left[\left(-\frac{1}{2} - \frac{1}{2}b_k - \frac{1}{2}a_k i\right)^j\right] \right] \\ &= \pi \sum_{j=1}^{l-1} \left[\binom{2l-j-2}{l-1} \frac{1}{j} (-1)^{-2l+2j+1} (2)^{-2l-j+3} \frac{1}{\left(\left(\frac{1}{2} + \frac{1}{2}b_k\right)^2 + \frac{1}{4}a_k^2\right)^j} \operatorname{Re}\left[(1 + b_k + a_k i)^j\right] \right]\end{aligned}$$

$$= \pi \sum_{j=1}^{l-1} \left[\binom{2l-j-2}{l-1} \frac{1}{j} (-1)^{-2l+2j+1} (2)^{-2l+j+3} \frac{1}{((1+b_k)^2 + a_k^2)^j} \operatorname{Re}[(1+b_k + a_k i)^j] \right].$$

We have $z_k = a_k + ib_k$, where $b_k > 0$. We have used

$$\left(-\frac{1}{2} + \frac{1}{2}i z_k\right)^{-j} = \left(-\frac{1}{2} - \frac{1}{2}b_k + \frac{1}{2}a_k i\right)^{-j} = \left[\frac{-\frac{1}{2} - \frac{1}{2}b_k - \frac{1}{2}a_k i}{\left(-\frac{1}{2} - \frac{1}{2}b_k\right)^2 + \frac{1}{4}a_k^2}\right]^j.$$

We have

$$\begin{aligned} (1 + b_k + a_k i)^j &= \sum_{m=0}^j \binom{j}{m} (a_k i)^m (1 + b_k)^{j-m} \\ &= (1 + b_k)^j + \sum_{m=1, \text{ odd}}^j \binom{j}{m} (-1)^{(m-1)/2} (a_k)^m (1 + b_k)^{j-m} i \\ &\quad + \sum_{m=2, \text{ even}}^j \binom{j}{m} (-1)^{m/2} (a_k)^m (1 + b_k)^{j-m}. \end{aligned}$$

Hence,

$$\operatorname{Re}[(1 + b_k + a_k i)^j] = (1 + b_k)^j + \sum_{m=2, \text{ even}}^j \binom{j}{m} (-1)^{m/2} (a_k)^m (1 + b_k)^{j-m}.$$

Hence,

$$\begin{aligned} 2\operatorname{Re} \left\{ (-2\pi i) \left\{ \sum_{j=1}^{l-1} \binom{2l-j-2}{l-1} \frac{1}{j} (-1)^{-l-1+j} (-i - z_k)^{-j} (2i)^{-2l+1+j} \right\} \right\} &= \\ \pi \sum_{j=1}^{l-1} \left[\binom{2l-j-2}{l-1} \frac{1}{j} (-1)^{-2l+2j+1} (2)^{-2l+j+3} \frac{1}{((1+b_k)^2 + a_k^2)^j} \right] & \left[(1+b_k)^j + \right. \\ \left. \sum_{m=2, \text{ even}}^j \binom{j}{m} (-1)^{m/2} (a_k)^m (1+b_k)^{j-m} \right] & \quad (2) \end{aligned}$$

Combining (1) and (2), we are done. □

We include some following results from Theorem 5.2.2.

Theorem 5.2.3. $\int_{-\infty}^{\infty} \frac{1}{1+x^2} \log(q(x)) dx = \pi \log(|\psi(-i)|^2)$.

Proof.

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{1+x^2} \log(q(x)) dx &= \int_{-\infty}^{\infty} \frac{1}{1+x^2} \log(|\psi(x)|^2) dx = 2\operatorname{Re} \left(\int_{-\infty}^{\infty} \frac{1}{1+x^2} \log(\psi(x)) dx \right) \\ &= 2\operatorname{Re} \left\{ (-2\pi i) \operatorname{Res} \left(\frac{\log(\psi(z))}{z^2+1}, z = -i \right) \right\} \\ &= 2\operatorname{Re} \left\{ (-2\pi i) \lim_{z \rightarrow -i} \frac{\log(\psi(z))}{z-i} \right\} \\ &= 2\operatorname{Re}(\pi \log(\psi(-i))) = 2\pi \log |\psi(-i)| = \pi \log(|\psi(-i)|^2). \end{aligned}$$

□

Remark: Theorem 5.2.3 is still hold if $q(x)$ has real zeros. This follows from a limiting argument with

$$q(x) = \lim_{\epsilon \rightarrow 0} q(x) + \epsilon = \lim_{\epsilon \rightarrow 0} |\psi(x) + \epsilon|^2.$$

And

$$\lim_{\epsilon \rightarrow 0} \pi \log(|\psi(-i) + \epsilon|^2) = \pi \log(|\psi(-i)|^2).$$

The following proposition is an immediate result from Theorem 5.2.3.

Proposition 5.2.4. *i. Let b, c be real numbers such that $D = b^2 - 4c \leq 0$. Then*

$$\int_{-\infty}^{\infty} \frac{1}{1+x^2} \log(x^2 + bx + c) dx = \pi \log(c + 1 + \sqrt{|D|}).$$

ii. Let $u, v \in \mathbb{R}$ such that $\Delta = v^2 - 4w < 0$ and let ψ be the spectral factor of q .

$$\int_{-\infty}^{\infty} \frac{1}{w+vx+x^2} \log(q(x)) dx = \frac{\pi}{\frac{1}{2}\sqrt{|\Delta|}} \log(|\psi(-\frac{v}{2} - i\frac{1}{2}\sqrt{|\Delta|})|^2).$$

Proof. i. We apply Theorem 5.2.3 with $\psi(z) = (z + \frac{b}{2}) - i\frac{\sqrt{|D|}}{2}$.

We have

$$\begin{aligned} \psi(-i) &= (-1 - \frac{\sqrt{|D|}}{2})i + \frac{b}{2}, \\ |\psi(-i)|^2 &= (\frac{\sqrt{|D|}}{2} + 1)^2 + \frac{b^2}{4} = c + 1 + \sqrt{|D|}. \end{aligned}$$

Hence,

$$\int_{-\infty}^{\infty} \frac{1}{1+x^2} \log(x^2 + bx + c) dx = \pi \log(|\psi(-i)|^2) = \pi \log(c + 1 + \sqrt{|D|}).$$

ii. We have $w + vx + x^2 = (x + \frac{v}{2})^2 + \frac{|\Delta|}{4}$. Let $y = \frac{x + \frac{1}{2}v}{\frac{1}{2}\sqrt{|\Delta|}}$, so $x = \frac{1}{2}\sqrt{|\Delta|}y - \frac{1}{2}v$.

$$\int_{-\infty}^{\infty} \frac{1}{w + vx + x^2} \log(q(x)) dx = \frac{1}{\frac{1}{2}\sqrt{|\Delta|}} \int_{-\infty}^{\infty} \frac{\log(q(\frac{1}{2}\sqrt{|\Delta|}y - \frac{1}{2}v))}{y^2 + 1} dy.$$

We apply Theorem 5.2.3,

$$\int_{-\infty}^{\infty} \frac{\log(q(\frac{1}{2}\sqrt{|\Delta|}y - \frac{1}{2}v))}{y^2 + 1} dy = \pi \log(|\psi(\frac{-v}{2} - \frac{i}{2}\sqrt{|\Delta|})|^2).$$

Hence,

$$\int_{-\infty}^{\infty} \frac{1}{w + vx + x^2} \log(q(x)) dx = \frac{\pi}{\frac{1}{2}\sqrt{|\Delta|}} \log(|\psi(-\frac{v}{2} - \frac{i}{2}\sqrt{|\Delta|})|^2).$$

□

Theorem 5.2.5. $\int_{-\infty}^{\infty} \frac{1}{(1+x^2)^2} \log(q(x)) dx = \pi \log |\psi(-i)| - \pi \operatorname{Im}(\frac{\psi'(-i)}{\psi(-i)}).$

Proof.

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{(1+x^2)^2} \log(q(x)) dx &= 2\operatorname{Re}\{(-2\pi i) \lim_{z \rightarrow -i} \frac{d}{dz} [\frac{\log(\psi(z))}{(z-i)^2}]\} \\ &= 2\operatorname{Re}\{(-2\pi i) \lim_{z \rightarrow -i} (\frac{-2}{(z-i)^3} \log(\psi(z)) + \frac{1}{(z-i)^2} \frac{\psi'(z)}{\psi(z)})\} \\ &= 2\operatorname{Re}\{(-2\pi i) (\frac{i}{4} \log(\psi(-i)) - \frac{1}{4} \frac{\psi'(-i)}{\psi(-i)})\} \\ &= 2\operatorname{Re}\{\frac{2\pi}{4} \log(\psi(-i)) + \frac{\pi i}{2} \frac{\psi'(-i)}{\psi(-i)}\} \\ &= \pi \log(|\psi(-i)|) + \pi \operatorname{Re}(\frac{i\psi'(-i)}{\psi(-i)}) \\ &= \pi \log(|\psi(-i)|) - \pi \operatorname{Im}(\frac{\psi'(-i)}{\psi(-i)}). \end{aligned}$$

□

Theorem 5.2.6.

$$\int_{-\infty}^{\infty} \frac{1}{(1+x^2)^3} \log(q(x)) dx = \frac{\pi}{4} (3 \log(|\psi(-i)|) - 3 \operatorname{Im}(\frac{\psi'(-i)}{\psi(-i)}) + \operatorname{Re}\{[\frac{\psi'(-i)}{\psi(-i)}]^2 - \frac{\psi''(-i)}{\psi(-i)}\}).$$

Proof.

$$\int_{-\infty}^{\infty} \frac{1}{(1+x^2)^3} \log(q(x)) dx = 2 \operatorname{Re}\{(-\pi i) \lim_{z \rightarrow -i} \frac{d^2}{dz^2} [\frac{\log(\psi(z))}{(z-i)^3}]\},$$

using Theorem 5.2.2.

$$\begin{aligned} \frac{d}{dz} \left[\frac{\log(\psi(z))}{(z-i)^3} \right] &= -\frac{3 \log(\psi(z))}{(z-i)^4} + \frac{\psi'(z)}{\psi(z)} \frac{1}{(z-i)^3}, \\ \frac{d^2}{dz^2} \left[\frac{\log(\psi(z))}{(z-i)^3} \right] &= \frac{12 \log(\psi(z))}{(z-i)^5} - \frac{6\psi'(z)}{\psi(z)} \frac{1}{(z-i)^4} + \frac{1}{(z-i)^3} \left[\frac{\psi''(z)}{\psi(z)} - \left(\frac{\psi'(z)}{\psi(z)}\right)^2 \right], \\ \lim_{z \rightarrow -i} \frac{d^2}{dz^2} \left[\frac{\log(\psi(z))}{(z-i)^3} \right] &= \frac{12 \log(\psi(-i))}{(-32i)} - \frac{6\psi'(-i)}{\psi(-i)} \frac{1}{16} + \frac{1}{8i} \left[\frac{\psi''(-i)}{\psi(-i)} - \left(\frac{\psi'(-i)}{\psi(-i)}\right)^2 \right]. \end{aligned}$$

Hence,

$$\int_{-\infty}^{\infty} \frac{1}{(1+x^2)^3} \log(q(x)) dx = \frac{\pi}{4} (3 \log(|\psi(-i)|) - 3 \operatorname{Im}(\frac{\psi'(-i)}{\psi(-i)}) + \operatorname{Re}\{[\frac{\psi'(-i)}{\psi(-i)}]^2 - \frac{\psi''(-i)}{\psi(-i)}\}).$$

□

5.3. The Entropy and the Kullback-Leiber (KL) Divergence of Compound Densities

As in chapter two from MC-Tree method, the compound density $r(x) = c \frac{p(x)}{h(x)}$, where $h(x) = (x^2 + a)^k$; the rational function $\bar{r}(x) = d \frac{\bar{p}(x)}{\bar{h}(x)}$, where $\bar{h}(x) = (x^2 + b)^l$, $p(x)$ and $\bar{p}(x)$

are even polynomials with degrees of n and m , respectively. All a, b, c, d are positive, $k \in \mathbb{N}$, and $l \in \mathbb{N}$.

Note that the compound density is a non-negative even rational function, so we can use partial fraction decomposition to obtain

$$r(x) = \frac{c_1}{(x^2 + a)} + \frac{c_2}{(x^2 + a)^2} + \frac{c_3}{(x^2 + a)^3} + \dots + \frac{c_k}{(x^2 + a)^k}.$$

Note that $c_1 = 0$ because the codegree has to be greater than two for a rational function to be a pdf.

The entropy is

$$\begin{aligned} Ent(r(x)) &= - \int_{-\infty}^{\infty} r(x) \log(r(x)) dx \\ &= - \log(c) \left(\int_{-\infty}^{\infty} r(x) dx \right) - \int_{-\infty}^{\infty} r(x) \log(p(x)) dx + k \int_{-\infty}^{\infty} r(x) \log(x^2 + a) dx. \end{aligned}$$

We have

$$\int_{-\infty}^{\infty} r(x) \log(p(x)) dx = c_2 \int_{-\infty}^{\infty} \frac{\log(p(x))}{(x^2 + a)^2} dx + c_3 \int_{-\infty}^{\infty} \frac{\log(p(x))}{(x^2 + a)^3} dx + \dots + c_k \int_{-\infty}^{\infty} \frac{\log(p(x))}{(x^2 + a)^k} dx.$$

Note that $p(z)$ has $\frac{n}{2}$ zeros, $z_1, z_2, \dots, z_{\frac{n}{2}}$ in the upper half plane, $p(z)$ has the spectral factor $\psi(z) = (z - z_1)(z - z_2) \dots (z - z_{\frac{n}{2}})$ and $p(z) = |\psi(z)|^2$.

The Kullback-Leiber (KL) Divergence of $r(x)$ and $\bar{r}(x)$ is

$$\begin{aligned} KL(r(x)|\bar{r}(x)) &= \int_{-\infty}^{\infty} r(x) \log\left\{\frac{r(x)}{\bar{r}(x)}\right\} dx = \int_{-\infty}^{\infty} r(x) \log(r(x)) dx - \int_{-\infty}^{\infty} r(x) \log(\bar{r}(x)) dx \\ &= -Ent(r(x)) - \int_{-\infty}^{\infty} r(x) \log(\bar{r}(x)) dx. \end{aligned}$$

We have

$$\int_{-\infty}^{\infty} r(x) \log(\bar{r}(x)) dx = \log(d) \left(\int_{-\infty}^{\infty} r(x) dx \right) - l \int_{-\infty}^{\infty} r(x) \log(x^2 + b) dx + \int_{-\infty}^{\infty} r(x) \log(\bar{p}(x)) dx.$$

Note that $r(x)$ is a compound density, so $\int_{-\infty}^{\infty} r(x) dx = 1$.

Similarly,

$$\int_{-\infty}^{\infty} r(x) \log(\bar{p}(x)) dx = c_2 \int_{-\infty}^{\infty} \frac{\log(\bar{p}(x))}{(x^2 + a)^2} dx + c_3 \int_{-\infty}^{\infty} \frac{\log(\bar{p}(x))}{(x^2 + a)^3} dx + \dots + c_k \int_{-\infty}^{\infty} \frac{\log(\bar{p}(x))}{(x^2 + a)^k} dx.$$

$\bar{p}(z)$ has $\frac{m}{2}$ zeros, $\bar{z}_1, \bar{z}_2, \dots, \bar{z}_{\frac{m}{2}}$ in the upper half plane, $\bar{p}(z)$ has the spectral factor $\bar{\psi}(z) = (z - \bar{z}_1)(z - \bar{z}_2) \dots (z - \bar{z}_{\frac{m}{2}})$ and $\bar{p}(z) = |\bar{\psi}(z)|^2$. Note that the notation $\bar{\psi}(z)$ is not the complex conjugate.

The entropy and the KL above can be calculated by using Theorem 5.2.1

In general, we can calculate the entropy and the KL for compound densities

$r(x) = c \frac{p_i(x)}{\prod_{i=1}^{j'} (x^2 + a_i)^{k_i}}$ and $\bar{r}(x) = d \frac{\bar{p}_i(x)}{\prod_{i=1}^{j'} (x^2 + b_i)^{l_i}}$, $p_i(x), \bar{p}_i(x)$ are even polynomials, using the product rule for logarithms. We apply partial fraction decomposition to $r(x)$

$$\begin{aligned} r(x) &= \frac{c_1^1}{(x^2 + a_1)} + \frac{c_2^1}{(x^2 + a_1)^2} + \dots + \frac{c_{k_1}^1}{(x^2 + a_1)^{k_1}} \\ &+ \dots \\ &+ \frac{c_1^{j'}}{(x^2 + a_{j'})} + \frac{c_2^{j'}}{(x^2 + a_{j'})^2} + \dots + \frac{c_{k_{j'}}^{j'}}{(x^2 + a_{j'})^{k_{j'}}}. \end{aligned}$$

Consider a convex combination of rational probability density functions

$r(x) = \sum_{i=1}^n \alpha_i r_i(x)$ such that $\alpha_i \geq 0$ and $\sum_{i=1}^n \alpha_i = 1$. Note that $r(x)$ is still a pdf and the entropy of $r(x)$ and the KL Divergence between $r(x)$ and $\bar{r}(x)$ can be calculated, using Theorem 5.2.1 Some examples with specific values of the tree depth N when fixing $m = 1$ are illustrated for analytic computations, as follows:

Example 1: $N = 1$

The compound density is $q(x) = \frac{2}{\pi} \frac{1}{(1+x^2)^2}$. We can prove that

$$\int_{-\infty}^{\infty} q(x)dx = \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{1}{(1+x^2)^2} dx = 1,$$

by using

$$\int_{-\infty}^{\infty} \frac{1}{(1+x^2)^2} dx = (-2\pi i) \lim_{z \rightarrow -i} \frac{d}{dz} \frac{1}{(z-i)^2} = (-2\pi i) \lim_{z \rightarrow -i} \frac{-2}{(z-i)^3} = (-2\pi i) \frac{-2}{8i} = \frac{\pi}{2}.$$

The entropy of this compound density is

$$Ent(q(x)) = - \int_{-\infty}^{\infty} \frac{2}{\pi} \frac{1}{(1+x^2)^2} \log\left[\frac{2}{\pi} \frac{1}{(1+x^2)^2}\right] dx = -\log\left(\frac{2}{\pi}\right) + \frac{4}{\pi} \int_{-\infty}^{\infty} \frac{\log(1+x^2)}{(1+x^2)^2} dx.$$

Apply the formula, we get

$$\int_{-\infty}^{\infty} \frac{\log(1+x^2)}{(1+x^2)^2} dx = \pi \log(2) - \frac{\pi}{2}.$$

$$Ent(q(x)) = -\log\left(\frac{2}{\pi}\right) + \frac{4}{\pi} \pi \left(\log(2) - \frac{1}{2}\right) = -\log\left(\frac{2}{\pi}\right) + 4 \log(2) - 2 \approx 1.2242.$$

Example 2: $N = 2$

The compound density is $q(x) = \frac{\sqrt{2}(x^2+10)}{\pi(x^2+2)^3}$.

First, we can prove that $\int_{-\infty}^{\infty} q(x)dx = 1$.

We have

$$\frac{x^2+10}{(x^2+2)^3} = \frac{1}{(x^2+2)^2} + \frac{8}{(x^2+2)^3}.$$

$$\int_{-\infty}^{\infty} \frac{dx}{(x^2+2)^3} = -2i\pi \frac{1}{2} \lim_{z \rightarrow -\sqrt{2}i} \frac{d^2}{dz^2} \frac{1}{(z-\sqrt{2}i)^3} = -2i\pi \frac{1}{2} \frac{12}{32i\sqrt{2}} = \frac{3\pi}{32\sqrt{2}}.$$

$$\int_{-\infty}^{\infty} \frac{dx}{(x^2+2)^2} = -2i\pi \lim_{z \rightarrow -\sqrt{2}i} \frac{d}{dz} \frac{1}{(z-\sqrt{2}i)^2} = \frac{\pi}{4\sqrt{2}}.$$

Hence,

$$\int_{-\infty}^{\infty} \frac{x^2 + 10}{(x^2 + 2)^3} dx = \frac{\pi}{4\sqrt{2}} + \frac{3\pi}{4\sqrt{2}} = \frac{\pi}{\sqrt{2}},$$

so

$$\int_{-\infty}^{\infty} q(x) dx = 1.$$

Next, calculate the entropy

$$\text{Ent}(q(x)) = - \int_{-\infty}^{\infty} \frac{\sqrt{2}(x^2 + 10)}{\pi(x^2 + 2)^3} \log\left[\frac{\sqrt{2}(x^2 + 10)}{\pi(x^2 + 2)^3}\right] dx.$$

Let $u = \frac{x}{\sqrt{2}}$, so $dx = \sqrt{2} du$.

$$\begin{aligned} \text{Ent}(q(x)) &= - \int_{-\infty}^{\infty} \frac{\sqrt{2}}{4\pi} \frac{u^2 + 5}{(u^2 + 1)^3} \log\left[\frac{\sqrt{2}}{4\pi} \frac{u^2 + 5}{(u^2 + 1)^3}\right] du \\ &= - \log\left(\frac{\sqrt{2}}{4\pi}\right) - \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{u^2 + 5}{(u^2 + 1)^3} \log\left[\frac{u^2 + 5}{(u^2 + 1)^3}\right] du. \end{aligned}$$

We have

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{u^2 + 5}{(u^2 + 1)^3} \log\left[\frac{u^2 + 5}{(u^2 + 1)^3}\right] du &= \int_{-\infty}^{\infty} \frac{u^2 + 5}{(u^2 + 1)^3} \log[u^2 + 5] du \\ &\quad - 3 \int_{-\infty}^{\infty} \frac{u^2 + 5}{(u^2 + 1)^3} \log[u^2 + 1] du. \quad (1) \end{aligned}$$

Observe that

$$\frac{u^2 + 5}{(u^2 + 1)^3} = \frac{1}{(u^2 + 1)^2} + 4 \frac{1}{(u^2 + 1)^3}.$$

Apply the formula, we have

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\log(u^2 + 5)}{(u^2 + 1)^2} du &= 2\operatorname{Re}\{(-2\pi i) \lim_{z \rightarrow -i} \frac{d}{dz} \frac{\log(iz + \sqrt{5})}{(z - i)^2}\} \\ &= 2\operatorname{Re}\{(-2\pi i) \left(-\frac{\log(1 + \sqrt{5})}{4i} + \frac{1}{4i(1 + \sqrt{5})}\right)\} \\ &= \pi(\log(1 + \sqrt{5}) - \frac{1}{1 + \sqrt{5}}) \end{aligned}$$

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\log(u^2 + 5)}{(u^2 + 1)^3} du &= 2\operatorname{Re}\{(-2\pi i) \frac{1}{2} \lim_{z \rightarrow -i} \frac{d^2}{dz^2} \frac{\log(iz + \sqrt{5})}{(z - i)^3}\} \\ &= 2\operatorname{Re}\{(-\pi i) \left(-\frac{3\log(1 + \sqrt{5})}{8i} + \frac{1}{8i(6 + 2\sqrt{5})} + \frac{3}{8(1 + \sqrt{5}i)}\right)\} \\ &= \frac{3}{4}\pi \log(1 + \sqrt{5}) - \frac{3\pi}{4(1 + \sqrt{5})} - \frac{\pi}{4(6 + 2\sqrt{5})}. \end{aligned}$$

Hence,

$$A_1 = \int_{-\infty}^{\infty} \frac{u^2 + 5}{(u^2 + 1)^3} \log[u^2 + 5] du = \pi(4 \log(1 + \sqrt{5}) - \frac{4}{1 + \sqrt{5}} - \frac{1}{6 + 2\sqrt{5}}).$$

Apply the formula, we have

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\log(u^2 + 1)}{(u^2 + 1)^2} du &= 2\operatorname{Re}\{(-2\pi i) \lim_{z \rightarrow -i} \frac{d}{dz} \left[\frac{\log(iz + 1)}{(z - i)^2}\right]\} = \pi \log(2) - \frac{\pi}{2}. \\ \int_{-\infty}^{\infty} \frac{\log(u^2 + 1)}{(u^2 + 1)^3} du &= 2\operatorname{Re}\{(-2\pi i) \frac{1}{2} \lim_{z \rightarrow -i} \frac{d^2}{dz^2} \frac{\log(iz + 1)}{(z - i)^3}\} \\ &= 2\operatorname{Re}\{(-2\pi i) \frac{1}{2} \lim_{z \rightarrow -i} \left[\frac{12\log(1 + iz)}{(z - i)^5} - \frac{7}{(z - i)^5}\right]\} \\ &= \pi \left(\frac{12 \log(2) - 7}{16}\right). \end{aligned}$$

Hence,

$$\begin{aligned} A_2 &= \int_{-\infty}^{\infty} \frac{u^2 + 5}{(u^2 + 1)^3} \log[u^2 + 1] du = (\pi \log(2) - \frac{\pi}{2}) + 4\pi \left(\frac{12 \log(2) - 7}{16} \right) \\ &= \pi \left(4 \log(1 + \sqrt{5}) - \frac{2}{1 + \sqrt{5}} - \frac{1}{6 + 2\sqrt{5}} \right). \end{aligned}$$

Substituting A_1 and A_2 into (1), we obtain

$$\int_{-\infty}^{\infty} \frac{u^2 + 5}{(u^2 + 1)^3} \log \left[\frac{u^2 + 5}{(u^2 + 1)^3} \right] du = \pi \left(4 \log(1 + \sqrt{5}) - 12 \log(2) - \frac{4}{1 + \sqrt{5}} - \frac{1}{6 + 2\sqrt{5}} + \frac{27}{4} \right).$$

Finally,

$$\text{Ent}(q(x)) = -\log\left(\frac{\sqrt{2}}{4\pi}\right) - (2 \log(1 + \sqrt{5}) - 6 \log(2) - \frac{2}{1 + \sqrt{5}} - \frac{1}{12 + 4\sqrt{5}} + \frac{27}{8}) \approx 1.285396.$$

Example 3: $N = 3$

The compound density is

$$q(x) = \frac{2}{\sqrt{3}\pi} \frac{x^4 + 18x^2 + 117}{(x^2 + 3)^4}.$$

Observe that

$$\frac{x^4 + 18x^2 + 117}{(x^2 + 3)^4} = \frac{1}{(x^2 + 3)^2} + \frac{12}{(x^2 + 3)^3} + \frac{72}{(x^2 + 3)^4}.$$

We have

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{dx}{(x^2 + 3)^2} &= (-2\pi i) \lim_{z \rightarrow -\sqrt{3}i} \frac{d}{dz} \left[\frac{1}{(z - \sqrt{3}i)^2} \right] = (-2\pi i) \frac{-2}{24\sqrt{3}i} = \frac{\pi}{6\sqrt{3}}. \\ \int_{-\infty}^{\infty} \frac{dx}{(x^2 + 3)^3} &= (-2\pi i) \frac{1}{2} \lim_{z \rightarrow -\sqrt{3}i} \frac{d^2}{dz^2} \left[\frac{1}{(z - \sqrt{3}i)^3} \right] = (-2\pi i) \frac{1}{2} \frac{-12}{288\sqrt{3}i} = \frac{\pi}{24\sqrt{3}}. \\ \int_{-\infty}^{\infty} \frac{dx}{(x^2 + 3)^4} &= (-2\pi i) \frac{1}{6} \lim_{z \rightarrow -\sqrt{3}i} \frac{d^3}{dz^3} \left[\frac{1}{(z - \sqrt{3}i)^4} \right] = (-\pi i) \frac{1}{3} \frac{-120}{3456\sqrt{3}i} = \frac{5\pi}{432\sqrt{3}}. \end{aligned}$$

Hence,

$$\int_{-\infty}^{\infty} \frac{x^4 + 18x^2 + 117}{(x^2 + 3)^4} dx = \frac{\pi}{6\sqrt{3}} + \frac{\pi}{2\sqrt{3}} + \frac{5\pi}{6\sqrt{3}} = \frac{\pi\sqrt{3}}{2}.$$

Therefore,

$$\int_{-\infty}^{\infty} q(x) dx = 1.$$

The entropy is

$$\begin{aligned} Ent(q(x)) &= -\frac{2}{\sqrt{3}\pi} \int_{-\infty}^{\infty} \frac{x^4 + 18x^2 + 117}{(x^2 + 3)^4} \log\left[\frac{2}{\sqrt{3}\pi} \frac{x^4 + 18x^2 + 117}{(x^2 + 3)^4}\right] dx \\ &= -\log\left(\frac{2}{\sqrt{3}\pi}\right) - \frac{2}{\sqrt{3}\pi} \int_{-\infty}^{\infty} \frac{x^4 + 18x^2 + 117}{(x^2 + 3)^4} \log\left[\frac{x^4 + 18x^2 + 117}{(x^2 + 3)^4}\right] dx. \end{aligned}$$

Let $u = \frac{x}{\sqrt{3}}$, so $dx = \sqrt{3}du$.

$$\begin{aligned} B &= \int_{-\infty}^{\infty} \frac{x^4 + 18x^2 + 117}{(x^2 + 3)^4} \log\left[\frac{x^4 + 18x^2 + 117}{(x^2 + 3)^4}\right] dx \\ &= \frac{\sqrt{3}}{9} \int_{-\infty}^{\infty} \frac{u^4 + 6u^2 + 13}{(u^2 + 1)^4} \log\left[\frac{u^4 + 6u^2 + 13}{9(u^2 + 1)^4}\right] du \\ &= \frac{\sqrt{3}}{9} \int_{-\infty}^{\infty} \frac{u^4 + 6u^2 + 13}{(u^2 + 1)^4} \log[u^4 + 6u^2 + 13] du \\ &\quad - \frac{4\sqrt{3}}{9} \int_{-\infty}^{\infty} \frac{u^4 + 6u^2 + 13}{(u^2 + 1)^4} \log[u^2 + 1] du \\ &\quad - \frac{4\sqrt{3}}{9} \log(\sqrt{3}) \int_{-\infty}^{\infty} \frac{u^4 + 6u^2 + 13}{(u^2 + 1)^4} du. \end{aligned}$$

We have

$$\int_{-\infty}^{\infty} \frac{u^4 + 6u^2 + 13}{(u^2 + 1)^4} du = \frac{9\pi}{2}.$$

We want to calculate

$$C = \int_{-\infty}^{\infty} \frac{u^4 + 6u^2 + 13}{(u^2 + 1)^4} \log[u^2 + 1] du.$$

Observe that

$$\frac{u^4 + 6u^2 + 13}{(u^2 + 1)^4} = \frac{1}{(u^2 + 1)^2} + 4\frac{1}{(u^2 + 1)^3} + 8\frac{1}{(u^2 + 1)^4}.$$

We have

$$\begin{aligned} C_1 &= \int_{-\infty}^{\infty} \frac{\log(u^2 + 1)}{(u^2 + 1)^2} du = \pi \log(2) - \frac{\pi}{2}. \\ C_2 &= \int_{-\infty}^{\infty} \frac{\log(u^2 + 1)}{(u^2 + 1)^3} du = \pi \frac{12 \log(2) - 7}{16}. \\ C_3 &= \int_{-\infty}^{\infty} \frac{\log(u^2 + 1)}{(u^2 + 1)^4} du = 2\operatorname{Re}\left\{(-2\pi i) \frac{1}{6} \lim_{z \rightarrow -i} \frac{d^3}{dz^3} \left[\frac{\log(1 + iz)}{(z - i)^4} \right]\right\} \\ &= 2\operatorname{Re}\left\{(-2\pi i) \frac{1}{6} \lim_{z \rightarrow -i} \frac{-120 \log(1 + iz) + 74}{(z - i)^7}\right\} = \pi \left(\frac{5}{8} \log(2) - \frac{37}{96} \right). \end{aligned}$$

Hence,

$$C = C_1 + 4C_2 + 8C_3 = \pi \left(9 \log(2) - \frac{16}{3} \right).$$

Next, we want to calculate

$$D = \int_{-\infty}^{\infty} \frac{u^4 + 6u^2 + 13}{(u^2 + 1)^4} \log[u^4 + 6u^2 + 13] du.$$

Firstly, we calculate

$$D_1 = \int_{-\infty}^{\infty} \frac{\log(u^4 + 6u^2 + 13)}{(u^2 + 1)^2} du.$$

We have

$$\psi(z) = (z - z_1)(z - z_2),$$

where

$$z_1 = \frac{\sqrt{2\sqrt{13} - 6}}{2} + i \frac{\sqrt{2\sqrt{13} + 6}}{2}, \quad z_2 = -\frac{\sqrt{2\sqrt{13} - 6}}{2} + i \frac{\sqrt{2\sqrt{13} + 6}}{2}$$

are its zeros in the UHP.

We have

$$\psi(-i) = (-i - z_1)(-i - z_2),$$

so

$$|\psi(-i)| = \left(1 + \frac{\sqrt{2\sqrt{13} + 6}}{2}\right)^2 + \frac{2\sqrt{13} - 6}{4} = \sqrt{13} + 1 + \sqrt{2\sqrt{13} + 6}.$$

We have

$$\begin{aligned} \frac{\psi'(z)}{\psi(z)} &= \frac{2z - (z_1 + z_2)}{(z - z_1)(z - z_2)}, \\ \frac{\psi'(-i)}{\psi(-i)} &= \frac{(2 + \sqrt{2\sqrt{13} + 6})i}{\sqrt{13} + 1 + \sqrt{2\sqrt{13} + 6}}, \\ \operatorname{Im}\left(\frac{\psi'(-i)}{\psi(-i)}\right) &= \frac{2 + \sqrt{2\sqrt{13} + 6}}{\sqrt{13} + 1 + \sqrt{2\sqrt{13} + 6}}. \end{aligned}$$

Hence,

$$\begin{aligned} D_1 &= \int_{-\infty}^{\infty} \frac{\log(u^4 + 6u^2 + 13)}{(u^2 + 1)^2} du = \pi \left\{ -\operatorname{Im}\left(\frac{\psi'(-i)}{\psi(-i)}\right) + \log(|\psi(-i)|) \right\} \\ &= \pi \left[-\frac{(2 + \sqrt{2\sqrt{13} + 6})}{\sqrt{13} + 1 + \sqrt{2\sqrt{13} + 6}} + \log(\sqrt{13} + 1 + \sqrt{2\sqrt{13} + 6}) \right]. \end{aligned}$$

Secondly, we calculate

$$D_2 = \int_{-\infty}^{\infty} \frac{\log(u^4 + 6u^2 + 13)}{(u^2 + 1)^3} du.$$

We have

$$\left[\frac{\psi'(-i)}{\psi(-i)}\right]^2 = -\frac{(2 + \sqrt{2\sqrt{13} + 6})^2}{(\sqrt{13} + 1 + \sqrt{2\sqrt{13} + 6})^2}.$$

We have $\psi''(z) = 2 \forall z$, so

$$\psi''(-i) = 2,$$

and

$$\frac{\psi''(-i)}{\psi(-i)} = -\frac{2}{\sqrt{13+1} + \sqrt{2\sqrt{13+6}}}.$$

Hence,

$$\begin{aligned} D_2 &= \int_{-\infty}^{\infty} \frac{\log(u^4 + 6u^2 + 13)}{(u^2 + 1)^3} du \\ &= \frac{\pi}{4} \left\{ -3 \operatorname{Im} \left(\frac{\psi'(-i)}{\psi(-i)} \right) + 3 \log(|\psi(-i)|) + \operatorname{Re} \left\{ \left[\frac{\psi'(-i)}{\psi(-i)} \right]^2 - \frac{\psi''(-i)}{\psi(-i)} \right\} \right\} \\ &= \frac{\pi}{4} \left\{ 3 \log(\sqrt{13+1} + \sqrt{2\sqrt{13+6}}) - 3 \left(\frac{2 + \sqrt{2\sqrt{13+6}}}{\sqrt{13+1} + \sqrt{2\sqrt{13+6}}} \right) \right. \\ &\quad \left. - \frac{(2 + \sqrt{2\sqrt{13+6}})^2}{(\sqrt{13+1} + \sqrt{2\sqrt{13+6}})^2} + \frac{2}{\sqrt{13+1} + \sqrt{2\sqrt{13+6}}} \right\}. \end{aligned}$$

Thirdly, we calculate

$$D_3 = \int_{-\infty}^{\infty} \frac{\log(u^4 + 6u^2 + 13)}{(u^2 + 1)^4} du.$$

We have

$$\begin{aligned} \left[\frac{\psi'(-i)}{\psi(-i)} \right]^3 &= -\frac{(2 + \sqrt{2\sqrt{13+6}})^3}{(\sqrt{13+1} + \sqrt{2\sqrt{13+6}})^3} i. \\ \psi'''(z) &= 0 \forall z. \\ \frac{\psi'(-i)}{\psi(-i)} \frac{\psi''(-i)}{\psi(-i)} &= -\frac{2(2 + \sqrt{2\sqrt{13+6}})}{(\sqrt{13+1} + \sqrt{2\sqrt{13+6}})^2} i. \\ \operatorname{Im} \left\{ \frac{\psi'(-i)}{\psi(-i)} \frac{\psi''(-i)}{\psi(-i)} \right\} &= -\frac{2(2 + \sqrt{2\sqrt{13+6}})}{(\sqrt{13+1} + \sqrt{2\sqrt{13+6}})^2}. \end{aligned}$$

Hence,

$$\begin{aligned}
 D_3 &= \int_{-\infty}^{\infty} \frac{\log(u^4 + 6u^2 + 13)}{(u^2 + 1)^4} du \\
 &= \pi \left\{ \frac{5}{8} \log(|\psi(-i)|) - \frac{5}{8} \operatorname{Im} \left(\frac{\psi'(-i)}{\psi(-i)} \right) - \frac{1}{4} \operatorname{Re} \left\{ - \left[\frac{\psi'(-i)}{\psi(-i)} \right]^2 + \frac{\psi''(-i)}{\psi(-i)} \right\} \right. \\
 &\quad \left. - \frac{1}{24} \operatorname{Re} \left\{ 2i \left[\frac{\psi'(-i)}{\psi(-i)} \right]^3 - 3i \frac{\psi'(-i)}{\psi(-i)} \frac{\psi''(-i)}{\psi(-i)} + \frac{\psi'''(-i)}{\psi(-i)} i \right\} \right\} \\
 &= \pi \left\{ \frac{5}{8} \log(\sqrt{13} + 1 + \sqrt{2\sqrt{13} + 6}) - \frac{5}{8} \left(\frac{2 + \sqrt{2\sqrt{13} + 6}}{\sqrt{13} + 1 + \sqrt{2\sqrt{13} + 6}} \right) \right. \\
 &\quad \left. - \frac{1}{4} \left[\frac{(2 + \sqrt{2\sqrt{13} + 6})^2}{(\sqrt{13} + 1 + \sqrt{2\sqrt{13} + 6})^2} - \frac{2}{\sqrt{13} + 1 + \sqrt{2\sqrt{13} + 6}} \right] \right. \\
 &\quad \left. - \frac{1}{12} \frac{(2 + \sqrt{2\sqrt{13} + 6})^3}{(\sqrt{13} + 1 + \sqrt{2\sqrt{13} + 6})^3} \right. \\
 &\quad \left. + \frac{1}{8} \frac{2(2 + \sqrt{2\sqrt{13} + 6})}{(\sqrt{13} + 1 + \sqrt{2\sqrt{13} + 6})^2} \right\}.
 \end{aligned}$$

Hence,

$$D = D_1 + 4D_2 + 8D_3.$$

$$B = \frac{\sqrt{3}}{9} D - 2\sqrt{3} \log(\sqrt{3})\pi - \frac{4\sqrt{3}}{9} C.$$

$$\operatorname{Ent}(q(x)) = -\log\left(\frac{2}{\sqrt{3}\pi}\right) - \frac{2}{\sqrt{3}\pi} B \approx 1.3125.$$

Example 4: $N = 4$

The compound density is $q(x) = \frac{1}{\pi} \frac{x^6 + 27x^4 + 312x^2 + 1648}{(x^2 + 4)^5}$.

We can prove that $\int_{-\infty}^{\infty} q(x) dx = 1$.

Let $u = \frac{x}{2}$, so $dx = 2du$.

$$\frac{1}{\pi} \left[\frac{x^6 + 27x^4 + 312x^2 + 1648}{(x^2 + 4)^5} \right] = \frac{1}{64\pi} \left[\frac{4u^6 + 27u^4 + 78u^2 + 103}{(u^2 + 1)^5} \right].$$

The entropy is

$$\begin{aligned} Ent(q(x)) &= - \int_{-\infty}^{\infty} q(x) \log(q(x)) dx \\ &= \log(64\pi) + \frac{5}{32\pi} \int_{-\infty}^{\infty} \frac{4u^6 + 27u^4 + 78u^2 + 103}{(u^2 + 1)^5} \log(1 + u^2) du \\ &\quad - \frac{1}{32\pi} \int_{-\infty}^{\infty} \frac{4u^6 + 27u^4 + 78u^2 + 103}{(u^2 + 1)^5} \log(4u^6 + 27u^4 + 78u^2 + 103) du. \end{aligned}$$

Observe that

$$\frac{4u^6 + 27u^4 + 78u^2 + 103}{(u^2 + 1)^5} = \frac{4}{(1 + u^2)^2} + \frac{15}{(1 + u^2)^3} + \frac{36}{(1 + u^2)^4} + \frac{48}{(1 + u^2)^5}.$$

We want to calculate $E = \int_{-\infty}^{\infty} \frac{4u^6 + 27u^4 + 78u^2 + 103}{(u^2 + 1)^5} \log(1 + u^2) du$.

We have

$$E_1 = \int_{-\infty}^{\infty} \frac{\log(u^2 + 1)}{(u^2 + 1)^2} du = \pi \log(2) - \frac{\pi}{2}.$$

$$E_2 = \int_{-\infty}^{\infty} \frac{\log(u^2 + 1)}{(u^2 + 1)^3} du = 2Re\{(-2\pi i) \frac{1}{2} \lim_{z \rightarrow -i} \frac{d^2}{dz^2} [\frac{\log(1 + iz)}{(z - i)^3}]\} = \pi[\frac{3}{4} \log(2) - \frac{7}{16}].$$

$$\begin{aligned} E_3 &= \int_{-\infty}^{\infty} \frac{\log(u^2 + 1)}{(u^2 + 1)^4} du = 2Re\{(-2\pi i) \frac{1}{6} \lim_{z \rightarrow -i} \frac{d^3}{dz^3} [\frac{\log(1 + iz)}{(z - i)^4}]\} \\ &= 2Re\{(-2\pi i) \frac{1}{6} (\frac{-i}{4} + \frac{3}{32i} - \frac{15i}{32} - \frac{15 \log(2)}{16i})\} = \pi(\frac{60 \log(2) - 37}{96}). \end{aligned}$$

$$\begin{aligned} E_4 &= \int_{-\infty}^{\infty} \frac{\log(u^2 + 1)}{(u^2 + 1)^5} du = 2Re\{(-2\pi i) \frac{1}{24} \lim_{z \rightarrow -i} \frac{d^4}{dz^4} [\frac{\log(1 + iz)}{(z - i)^5}]\} \\ &= \pi[\frac{35 \log(2)}{64} - \frac{533}{1536}]. \end{aligned}$$

We used

$$\begin{aligned} \frac{d^4}{dz^4} \left[\frac{\log(1+iz)}{(z-i)^5} \right] &= -\frac{6i^4}{(z-i)^5(1+iz)^4} - \frac{40i^3}{(z-i)^6(1+iz)^3} - \frac{180i^2}{(z-i)^7(1+iz)^2} \\ &\quad - \frac{840i}{(z-i)^8(1+iz)} + \frac{1680 \log(1+iz)}{(z-i)^9}. \end{aligned}$$

Hence,

$$E = 4E_1 + 15E_2 + 36E_3 + 48E_4 = \pi \left[64 \log(2) - \frac{12251}{32} \right].$$

We want to calculate

$$\begin{aligned} F &= \int_{-\infty}^{\infty} \frac{4u^6 + 27u^4 + 78u^2 + 103}{(u^2 + 1)^5} \log[4u^6 + 27u^4 + 78u^2 + 103] du \\ &= \log(4) \left(\int_{-\infty}^{\infty} \frac{4u^6 + 27u^4 + 78u^2 + 103}{(u^2 + 1)^5} du \right) \\ &\quad + \int_{-\infty}^{\infty} \frac{4u^6 + 27u^4 + 78u^2 + 103}{(u^2 + 1)^5} \log \left[u^6 + \frac{27}{4}u^4 + \frac{78}{4}u^2 + \frac{103}{4} \right] du \\ &= 32\pi \log(4) + \int_{-\infty}^{\infty} \frac{4u^6 + 27u^4 + 78u^2 + 103}{(u^2 + 1)^5} \log \left[u^6 + \frac{27}{4}u^4 + \frac{78}{4}u^2 + \frac{103}{4} \right] du. \end{aligned}$$

Observe that $u^6 + \frac{27}{4}u^4 + \frac{78}{4}u^2 + \frac{103}{4}$ has the spectral factor

$\psi(u) = (u - z_1)(u - z_2)(u - z_3)$, while $\psi(z) = (z - z_1)(z - z_2)(z - z_3)$ has 3 zeros $z_1 = -0.7278 + 1.5252i$, $z_2 = 1.7767i$, and $z_3 = 0.7278 + 1.5252i$ in the UHP.

Firstly, we calculate

$$F_1 = \int_{-\infty}^{\infty} \frac{\log(u^6 + \frac{27}{4}u^4 + \frac{78}{4}u^2 + \frac{103}{4})}{(u^2 + 1)^2} du.$$

We have

$$\psi(-i) = (-i - z_1)(-i - z_2)(-i - z_3) = 19.1768i,$$

so

$$\begin{aligned} |\psi(-i)| &= 19.1768. \\ \frac{\psi'(z)}{\psi(z)} &= \frac{1}{(z-z_1)} + \frac{1}{(z-z_2)} + \frac{1}{(z-z_3)}, \\ \frac{\psi'(-i)}{\psi(-i)} &= \frac{20.9404}{19.1768}i = 1.0920i, \\ \operatorname{Im}\left(\frac{\psi'(-i)}{\psi(-i)}\right) &= 1.0920. \end{aligned}$$

Hence,

$$F_1 = \pi\{\log(|\psi(-i)|) - \operatorname{Im}\left(\frac{\psi'(-i)}{\psi(-i)}\right)\} = \pi[\log(19.1768) - 1.0920].$$

Secondly, we calculate

$$F_2 = \int_{-\infty}^{\infty} \frac{\log(u^6 + \frac{27}{4}u^4 + \frac{78}{4}u^2 + \frac{103}{4})}{(u^2+1)^3} du.$$

We have

$$\frac{\psi'(-i)}{\psi(-i)} = 1.0920i,$$

so

$$\operatorname{Re}\left\{\left[\frac{\psi'(-i)}{\psi(-i)}\right]^2\right\} = -1.1925.$$

$$\psi''(z) = 2(3z - (z_1 + z_2 + z_3)),$$

so $\psi''(-i) = -15.6542i$, and $\operatorname{Re}\left\{\frac{\psi''(-i)}{\psi(-i)}\right\} = -0.8163$. Hence,

$$\begin{aligned} F_2 &= \frac{\pi}{4}\{3\log(|\psi(-i)|) - 3\operatorname{Im}\left(\frac{\psi'(-i)}{\psi(-i)}\right) + \operatorname{Re}\left\{\left[\frac{\psi'(-i)}{\psi(-i)}\right]^2 - \frac{\psi''(-i)}{\psi(-i)}\right\}\} \\ &= \frac{\pi}{4}\left\{\frac{3}{4}\log(19.1768) - 3(1.0920) - 1.1925 + 0.8163\right\} = \frac{3\pi}{4}\log(19.1768) - 0.91305\pi. \end{aligned}$$

Thirdly, we calculate

$$F_3 = \int_{-\infty}^{\infty} \frac{\log(u^6 + \frac{27}{4}u^4 + \frac{78}{4}u^2 + \frac{103}{4})}{(u^2 + 1)^4} du.$$

We have

$$Im\{[\frac{\psi'(-i)}{\psi(-i)}]^3\} = -1.3022.$$

$$\psi'''(z) = 6 \forall z, \text{ so } \psi'''(-i) = 6.$$

$$\frac{\psi'''(-i)}{\psi(-i)} = -0.3129i, \text{ so } Im\{\frac{\psi'''(-i)}{\psi(-i)}\} = -0.3129.$$

$$\frac{\psi'(-i)\psi''(-i)}{\psi(-i)\psi(-i)} = -0.8914i, \text{ so } Im\{\frac{\psi'(-i)\psi''(-i)}{\psi(-i)\psi(-i)}\} = -0.8914.$$

Hence,

$$\begin{aligned} F_3 &= \pi\{\frac{5}{8}\log(|\psi(-i)|) - \frac{5}{8}Im(\frac{\psi'(-i)}{\psi(-i)}) - \frac{1}{4}Re\{-[\frac{\psi'(-i)}{\psi(-i)}]^2 + \frac{\psi''(-i)}{\psi(-i)}\} \\ &\quad - \frac{1}{24}Re\{2i[\frac{\psi'(-i)}{\psi(-i)}]^3 - 3i\frac{\psi'(-i)\psi''(-i)}{\psi(-i)\psi(-i)} + \frac{\psi'''(-i)}{\psi(-i)}i\}\} \\ &= \pi\{\frac{5}{8}\log(19.1768) - \frac{5}{8}(1.0920) - \frac{1}{4}(-0.8163) + \frac{1}{24}(-0.3129) - \frac{1}{8}(-0.8914) \\ &\quad + \frac{1}{12}(-1.3022)\} = \pi[\frac{5}{8}\log(19.1768) - 0.7867]. \end{aligned}$$

Finally, we calculate

$$F_4 = \int_{-\infty}^{\infty} \frac{\log(u^6 + \frac{27}{4}u^4 + \frac{78}{4}u^2 + \frac{103}{4})}{(u^2 + 1)^5} du.$$

We have

$$\psi''''(z) = 0 \forall z,$$

so

$$\psi''''(-i) = 0.$$

$[\frac{\psi''(-i)}{\psi(-i)}]^2 = 0.6664$, so $Re\{[\frac{\psi''(-i)}{\psi(-i)}]^2\} = 0.6664$.

We have

$$[\frac{\psi''(-i)}{\psi(-i)}]^2 \frac{\psi''(-i)}{\psi(-i)} = (-1.1925)(-0.8163) = 0.9734.$$

$$[\frac{\psi'(-i)}{\psi(-i)}]^4 = (1.0920i)^4 = 1.4220.$$

Hence,

$$\begin{aligned} F_4 &= \pi \left\{ \frac{35}{64} \log(|\psi(-i)|) - \frac{35}{64} Im\left(\frac{\psi'(-i)}{\psi(-i)}\right) - \frac{15}{64} Re\left\{-\left[\frac{\psi'(-i)}{\psi(-i)}\right]^2 + \frac{\psi''(-i)}{\psi(-i)}\right\} \right. \\ &+ \frac{5}{16} Im\left\{2\left[\frac{\psi'(-i)}{\psi(-i)}\right]^3 - 3\frac{\psi'(-i)}{\psi(-i)}\frac{\psi''(-i)}{\psi(-i)} + \frac{\psi'''(-i)}{\psi(-i)}\right\} + \frac{1}{192} Re\left\{-6\left[\frac{\psi'(-i)}{\psi(-i)}\right]^4 \right. \\ &+ \left. 12\left[\frac{\psi'(-i)}{\psi(-i)}\right]^2\frac{\psi''(-i)}{\psi(-i)} - 3\left[\frac{\psi''(-i)}{\psi(-i)}\right]^2 - 4\frac{\psi'(-i)}{\psi(-i)}\frac{\psi'''(-i)}{\psi(-i)} + \frac{\psi''''(-i)}{\psi(-i)}\right\} \\ &= \pi \left\{ \frac{35}{64} \log(19.1768) - \frac{35}{64}(1.0920) - \frac{15}{64}(1.1925 - 0.8163) \right. \\ &+ \frac{5}{96} \{2(-1.3022) - 3(-0.8914) - 0.3129\} \\ &+ \left. \frac{1}{192} \{(-6)1.4220 + 12(0.9734) - 3(0.6664) - 4(0.3417) + 0\} \right\} \\ &= \pi \left[\frac{35}{64} \log(19.1768) - 0.6992 \right]. \end{aligned}$$

Hence,

$$F = 4F_1 + 15F_2 + 36F_3 + 48F_4 = \pi[32 \log(4) + 64 \log(19.1768) - 79.9466].$$

$$Ent(q(x)) = \log(64\pi) - \frac{F}{32\pi} + \frac{5}{32\pi} E \approx 1.33129.$$

In summary, entropy values of compound densities are 1.2242 at $N = 1$, 1.2854 at $N = 2$, 1.3125 at $N = 3$, and 1.33129 at $N = 4$, respectively, when fixing $m = 1$.

5.4. Entropy and the KL Divergence for Cauchy Distributions

Let the variable X follow a Cauchy distribution with the pdf, $p_{l,s}$, on $(-\infty, \infty)$:

$p_{l,s}(x) = \frac{s}{\pi(s^2+(x-l)^2)}$, where l is the location parameter, and $s > 0$ is the scale parameter.

5.4.1. Entropy

Theorem 5.4.1. *The entropy of X is given by $Ent(X) = \log(4\pi s)$.*

Proof.

$$Ent(X) = - \int_{-\infty}^{\infty} p_{l,s}(x) \log[p_{l,s}(x)] dx.$$

Let $u = \frac{x-l}{s}$, so $dx = sdu$.

$$\begin{aligned} Ent(X) &= - \int_{-\infty}^{\infty} \frac{1}{s\pi(u^2+1)} \log\left[\frac{1}{s\pi(u^2+1)}\right] sdu \\ &= \log(\pi s) \left(\int_{-\infty}^{\infty} \frac{1}{\pi(1+u^2)} du \right) + \frac{1}{\pi} \left(\int_{-\infty}^{\infty} \frac{\log(1+u^2)}{1+u^2} du \right) \\ &= \log(\pi s) + \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\log(1+u^2)}{1+u^2} du. \end{aligned}$$

We have

$$\int_{-\infty}^{\infty} \frac{\log(1+u^2)}{1+u^2} du = \pi \log(|\psi(-i)|^2) = \pi \log(4),$$

using Theorem 5.2.3 with $\psi(z) = 1 + iz$. Hence,

$$Ent(X) = \log(\pi s) + \log(4) = \log(4\pi s).$$

□

5.4.2. The KL Divergence

Theorem 5.4.2. *The KL divergence of two Cauchy densities $p_{l,s}$ and $p_{\bar{l},\bar{s}}(x)$ is given by $KL(p_{l,s}|p_{\bar{l},\bar{s}}(x)) = \log\left\{\frac{(s+\bar{s})^2+(l-\bar{l})^2}{4s\bar{s}}\right\}$, where $s, \bar{s} > 0$.*

Proof.

$$KL(p_{l,s}|p_{\bar{l},\bar{s}}(x)) = \int_{-\infty}^{\infty} p_{l,s}(x) \log\left\{\frac{p_{l,s}(x)}{p_{\bar{l},\bar{s}}(x)}\right\} dx.$$

We have

$$\frac{p_{l,s}}{p_{\bar{l},\bar{s}}} = \log\left\{\frac{s}{\bar{s}} \left(\frac{\bar{s}^2 + (x - \bar{l})^2}{s^2 + (x - l)^2}\right)\right\}.$$

so

$$\begin{aligned} KL(p_{l,s}(x)|p_{\bar{l},\bar{s}}(x)) &= \log\left(\frac{s}{\bar{s}}\right) - \int_{-\infty}^{\infty} \frac{s}{\pi(s^2 + (x - l)^2)} \log(s^2 + (x - l)^2) dx \\ &+ \int_{-\infty}^{\infty} \frac{s}{\pi(s^2 + (x - l)^2)} \log(\bar{s}^2 + (x - \bar{l})^2) dx. \end{aligned}$$

Let $u = \frac{x-l}{s}$, so $dx = sdu$.

$$\int_{-\infty}^{\infty} \frac{s}{\pi(s^2 + (x - l)^2)} \log(\bar{s}^2 + (x - \bar{l})^2) dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\log((su + l - \bar{l})^2 + \bar{s}^2)}{1 + u^2} du.$$

And

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\log(su + l - \bar{l})^2 + \bar{s}^2}{1 + u^2} = \frac{1}{\pi} \pi \log(|\psi(-i)|^2) = \log(|\psi(-i)|^2) = \log\{(s + \bar{s})^2 + (l - \bar{l})^2\},$$

using Theorem 5.2.3 with $\psi(z) = (sz + l - \bar{l}) - i\bar{s}$. Hence,

$$\int_{-\infty}^{\infty} \frac{s}{\pi(s^2 + (x - l)^2)} \log(\bar{s}^2 + (x - \bar{l})^2) dx = \log\{(s + \bar{s})^2 + (l - \bar{l})^2\}.$$

And

$$\int_{-\infty}^{\infty} \frac{s}{\pi(s^2 + (x-l)^2)} \log(s^2 + (x-l)^2) dx = \frac{2\log(s)}{\pi} \int_{-\infty}^{\infty} \frac{1}{1+u^2} du + \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\log(1+u^2)}{1+u^2} du.$$

We have

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{1+u^2} du &= \pi, \\ \int_{-\infty}^{\infty} \frac{\log(1+u^2)}{1+u^2} du &= \pi \log(4), \end{aligned}$$

using Theorem 5.2.3 with $\psi(z) = 1 + iz$. Hence,

$$\int_{-\infty}^{\infty} \frac{s}{\pi(s^2 + (x-l)^2)} \log(s^2 + (x-l)^2) dx = 2\log(s) + \log(4).$$

Therefore, $KL(p_{l,s}(x)|p_{\bar{l},\bar{s}}(x)) = \log(\frac{s}{\bar{s}}) - (2\log(s) + \log(4)) + (\log\{(s + \bar{s})^2 + (l - \bar{l})^2\})$.
 $= \log\{\frac{(s+\bar{s})^2+(l-\bar{l})^2}{4s\bar{s}}\}$. □

5.5. Entropy and the KL Divergence Between the Rational Student Densities

5.5.1. Introduction

Let Y denote a univariate random variable that follows a Student t-distribution with odd degree of freedom $\nu \geq 3$. It has the pdf

$$f_{\nu}(y) = \frac{1}{\sqrt{\nu}B(\frac{1}{2}, \frac{\nu}{2})} (1 + \frac{y^2}{\nu})^{-\frac{(\nu+1)}{2}}.$$

Clearly, $f_{\nu}(y)$ is a rational function (only) when ν is odd (see [15]). We will call such a function a rational Student density. In that case an explicit formula for the normalization

factor, $B(\frac{1}{2}, \frac{\nu}{2})$, can be derived as follows. First note that

$$B(\frac{1}{2}, \frac{\nu}{2}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\nu}(1 + \frac{y^2}{\nu})^{(\frac{\nu+1}{2})}} dy.$$

Let $u = \frac{y}{\sqrt{\nu}}$, so $dy = \sqrt{\nu} du$, then we obtain

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{\nu}(1 + \frac{y^2}{\nu})^{(\frac{\nu+1}{2})}} dy = \int_{-\infty}^{\infty} (1 + u^2)^{-(\frac{\nu+1}{2})} du.$$

The rational function $\frac{1}{(1+u^2)^{\frac{\nu+1}{2}}}$ has one pole $u = -i$ with multiplicity $\frac{\nu+1}{2}$ in the lower half plane. Applying our main theorem using $q := e$, we have

$$\int_{-\infty}^{\infty} (1 + u^2)^{-(\frac{\nu+1}{2})} du = (-2\pi i) \frac{1}{(\frac{\nu-1}{2})!} \lim_{z \rightarrow -i} \frac{d^{\frac{\nu-1}{2}}}{dz^{\frac{\nu-1}{2}}} \frac{1}{(z-i)^{\frac{\nu+1}{2}}}.$$

As

$$\frac{d^{\frac{\nu-1}{2}}}{dz^{\frac{\nu-1}{2}}} \left[\frac{1}{(z-i)^{\frac{\nu+1}{2}}} \right] = (-1)^{\frac{\nu-1}{2}} \left(\frac{\nu+1}{2} \right) \left(\frac{\nu+3}{2} \right) \dots \left(\frac{2(\nu-1)}{2} \right) \frac{1}{(z-i)^{\nu}},$$

and

$$(-1)^{\nu} (i)^{\nu} (-1)^{\frac{1-\nu}{2}} = i^{2\nu+1} = (-1)^{\nu} i = -i,$$

it follows that

$$B(\frac{1}{2}, \frac{\nu}{2}) = \frac{\pi (\frac{\nu+1}{2}) (\frac{\nu+3}{2}) \dots (\frac{2(\nu-1)}{2})}{2^{\nu-1} (\frac{\nu-1}{2})!}.$$

5.5.2. Entropy

Theorem 5.5.1. *The entropy of Y is given by*

$$Ent(Y) = \log \left(\sqrt{\nu} B(\frac{1}{2}, \frac{\nu}{2}) \right) + (\nu + 1) \left[\log(2) - \frac{2}{\nu+1} - \frac{2}{\nu+3} - \dots - \frac{2}{2(\nu-1)} \right].$$

Proof.

$$Ent(Y) = - \int_{-\infty}^{\infty} f_{\nu}(y) \log[f_{\nu}(y)] dy = \log(\sqrt{\nu} B(\frac{1}{2}, \frac{\nu}{2})) + \frac{\nu + 1}{2B(\frac{1}{2}, \frac{\nu}{2})} \int_{-\infty}^{\infty} \frac{\log(1 + \frac{y^2}{\nu})}{\sqrt{\nu}(1 + \frac{y^2}{\nu})^{\frac{\nu+1}{2}}} dy.$$

Let $u = \frac{y}{\sqrt{\nu}}$, so $dy = \sqrt{\nu} du$. Hence,

$$\int_{-\infty}^{\infty} \frac{\log(1 + \frac{y^2}{\nu})}{\sqrt{\nu}(1 + \frac{y^2}{\nu})^{\frac{\nu+1}{2}}} dy = \int_{-\infty}^{\infty} \frac{\log(1 + u^2)}{(1 + u^2)^{\frac{\nu+1}{2}}} du.$$

A spectral factor of $1 + u^2$ is $\psi(u) = 1 + iu$, while $\psi(z) = 1 + iz$, $z \in \mathbb{C}$ has one zero, $z = i$, in the upper half plane, and has no zeros in the lower half plane. We have

$$\int_{-\infty}^{\infty} \frac{\log(1 + u^2)}{(1 + u^2)^{\frac{\nu+1}{2}}} du = 2Re\{(-2\pi i) \frac{1}{(\frac{\nu-1}{2})!} \lim_{z \rightarrow -i} \frac{d^{\frac{\nu-1}{2}}}{dz^{\frac{\nu-1}{2}}} [\frac{\log(1 + iz)}{(z - i)^{\frac{\nu+1}{2}}}]\},$$

using Theorem 5.2.2. We have

$$\frac{d}{dz} [\frac{\log(1 + iz)}{(z - i)^{\frac{\nu+1}{2}}}] = -(\frac{\nu + 1}{2}) [\frac{\log(1 + iz) - \frac{2}{\nu+1}}{(z - i)^{\frac{\nu+3}{2}}}] .$$

We have

$$\frac{d^2}{dz^2} [\frac{\log(1 + iz)}{(z - i)^{\frac{\nu+1}{2}}}] = (-\frac{\nu + 1}{2})(-\frac{\nu + 3}{2}) [\frac{\log(1 + iz) - \frac{2}{\nu+1} - \frac{2}{\nu+3}}{(z - i)^{\frac{\nu+5}{2}}}] .$$

In general,

$$\frac{d^{\frac{(\nu-1)}{2}}}{dz^{\frac{(\nu-1)}{2}}} [\frac{\log(1 + iz)}{(z - i)^{\frac{\nu+1}{2}}}] = (-1)^{\frac{\nu-1}{2}} (\frac{\nu + 1}{2}) \dots (\frac{2(\nu - 1)}{2}) [\frac{\log(1 + iz) - \frac{2}{\nu+1} - \dots - \frac{2}{2(\nu-1)}}{(z - i)^{\nu}}] .$$

Taking the limit, we obtain

$$\lim_{z \rightarrow -i} \frac{d^{\frac{(\nu-1)}{2}}}{dz^{\frac{(\nu-1)}{2}}} [\frac{\log(1 + iz)}{(z - i)^{\frac{\nu+1}{2}}}] = \frac{(\frac{\nu+1}{2})(\frac{\nu+3}{2}) \dots (\frac{2(\nu-1)}{2})}{(-i)^{2\nu}} [\log(2) - \frac{2}{\nu + 1} - \frac{2}{\nu + 3} - \dots - \frac{2}{2(\nu - 1)}] .$$

The following expressions are used in our simplification.

$$\begin{aligned} (-2i)^\nu &= (-1)^\nu 2^\nu i^\nu = (-1)^{\frac{3\nu}{2}} 2^\nu. \\ (-1)^{\frac{3\nu}{2}} (-1)^{\frac{1-\nu}{2}} &= (-1)^{\frac{2\nu+1}{2}} = (i)^{2\nu+1} = (-1)^\nu i = -i \end{aligned}$$

for ν odd.

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{\log(1+u^2)}{(1+u^2)^{\frac{\nu+1}{2}}} du &= \frac{\pi(\frac{\nu+1}{2})(\frac{\nu+3}{2})\dots(\nu-1)}{2^{\nu-2}(\frac{\nu-1}{2})!} [\log(2) - \frac{2}{\nu+1} - \frac{2}{\nu+3} - \dots - \frac{2}{2(\nu-1)}] \\ &= 2B\left(\frac{1}{2}, \frac{\nu}{2}\right) [\log(2) - \frac{2}{\nu+1} - \frac{2}{\nu+3} - \dots - \frac{2}{2(\nu-1)}]. \end{aligned}$$

Hence,

$$Ent(Y) = \log(\sqrt{\nu}B\left(\frac{1}{2}, \frac{\nu}{2}\right)) + (\nu+1) [\log(2) - \frac{2}{\nu+1} - \frac{2}{\nu+3} - \dots - \frac{2}{2(\nu-1)}].$$

□

5.5.3. The KL Divergence of Rational Student densities

Theorem 5.5.2. *The KL divergence of two rational student densities $f_\nu(y)$ and $f_{\bar{\nu}}(y)$ is given by*

$$\begin{aligned} KL(f_\nu(y)|f_{\bar{\nu}}(y)) &= - \left\{ \log(\sqrt{\nu}B\left(\frac{1}{2}, \frac{\nu}{2}\right)) + (\nu+1) \left[\log(2) - \frac{2}{\nu+1} - \frac{2}{\nu+3} - \dots - \frac{2}{2(\nu-1)} \right] \right\} \\ &\quad + \log(\sqrt{\bar{\nu}}B\left(\frac{1}{2}, \frac{\bar{\nu}}{2}\right)) - \frac{\bar{\nu}+1}{2B\left(\frac{1}{2}, \frac{\bar{\nu}}{2}\right)} \left[\frac{\pi}{(\frac{\nu-1}{2})2^{\frac{\nu-3}{2}}(1+\sqrt{\frac{\bar{\nu}}{\nu}})^{\frac{\nu-1}{2}}} + J \right] + (\bar{\nu}+1) \log\left(1 + \sqrt{\frac{\bar{\nu}}{\nu}}\right), \end{aligned}$$

where J is given by

$$J = \frac{\pi}{(\frac{\nu-1}{2})!} \sum_{k=1}^{\frac{\nu-3}{2}} \binom{\frac{\nu-1}{2}}{k} \frac{(\frac{\nu-1}{2}-k)!(\frac{\nu+1}{2})(\frac{\nu+3}{2})\dots(\frac{\nu+1}{2}+k-1)}{2^{\frac{\nu-3}{2}+k}(1+\sqrt{\frac{\bar{\nu}}{\nu}})^{\frac{\nu-1}{2}-k}}.$$

Proof.

$$KL(f_\nu(y)|f_{\bar{\nu}}(y)) = \int_{-\infty}^{\infty} f_\nu(y) \log\left(\frac{f_\nu(y)}{f_{\bar{\nu}}(y)}\right) dy =$$

$$-\left\{-\int_{-\infty}^{\infty} f_\nu(y) \log(f_\nu(y)) dy\right\} + \log\left(\sqrt{\bar{\nu}} B\left(\frac{1}{2}, \frac{\bar{\nu}}{2}\right)\right) + \frac{\bar{\nu} + 1}{2B\left(\frac{1}{2}, \frac{\bar{\nu}}{2}\right)} \int_{-\infty}^{\infty} \frac{\log\left(1 + \frac{y^2}{\bar{\nu}}\right)}{\sqrt{\bar{\nu}}\left(1 + \frac{y^2}{\bar{\nu}}\right)^{\frac{\bar{\nu}+1}{2}}} dy. \quad (1)$$

From Theorem 5.5.1, we have

$$Ent(Y) = -\int_{-\infty}^{\infty} f_\nu(y) \log(f_\nu(y)) dy$$

$$= \log\left(\sqrt{\nu} B\left(\frac{1}{2}, \frac{\nu}{2}\right)\right) + (\nu + 1)\left[\log(2) - \frac{2}{\nu + 1} - \frac{2}{\nu + 3} - \dots - \frac{2}{2(\nu - 1)}\right].$$

We need to calculate

$$I = \int_{-\infty}^{\infty} \frac{\log\left(1 + \frac{y^2}{\bar{\nu}}\right)}{\sqrt{\bar{\nu}}\left(1 + \frac{y^2}{\bar{\nu}}\right)^{\frac{\bar{\nu}+1}{2}}} dy.$$

Let $u = \frac{y}{\sqrt{\bar{\nu}}}$, so $dy = \sqrt{\bar{\nu}} du$.

$$I = \int_{-\infty}^{\infty} \frac{\log\left(1 + \frac{\bar{\nu}}{\bar{\nu}} u^2\right)}{(1 + u^2)^{\frac{\bar{\nu}+1}{2}}} du.$$

The rational function $\frac{1}{(1+u^2)^{\frac{\nu+1}{2}}}$ has a pole $u = -i$ in the lower half plane with the multiplicity $\frac{\nu+1}{2}$. The spectral of $1 + \frac{\nu}{\bar{\nu}} u^2$ is $\psi(u) = 1 + \sqrt{\frac{\nu}{\bar{\nu}}} i u$, while $\psi(z) = 1 + \sqrt{\frac{\nu}{\bar{\nu}}} i z$ has the zero in the upper half plane.

Applying Theorem 5.2.2, we have

$$I = 2Re\left\{(-2\pi i) \frac{1}{\left(\frac{\nu-1}{2}\right)!} \lim_{z \rightarrow -i} \frac{d^{\frac{\nu-1}{2}}}{dz^{\frac{\nu-1}{2}}} \left[\frac{\log\left(1 + \sqrt{\frac{\nu}{\bar{\nu}}} i z\right)}{(z - i)^{\frac{\nu+1}{2}}}\right]\right\}$$

Let $f = \log\left(1 + \sqrt{\frac{\nu}{\bar{\nu}}} i z\right)$.

We have

$$\begin{aligned} f^{(1)} &= \frac{1}{z - \sqrt{\frac{\nu}{2}}i}, \\ f^{(2)} &= \frac{(-1)^{2-1}1}{(z - \sqrt{\frac{\nu}{2}}i)^2}, \\ f^{(3)} &= \frac{(-1)^{3-1}1 \cdot 2}{(z - \sqrt{\frac{\nu}{2}}i)^3}. \end{aligned}$$

In general,

$$\begin{aligned} f^{(\frac{\nu-1}{2})} &= \frac{(-1)^{\frac{\nu-3}{2}}(\frac{\nu-3}{2})!}{(z - \sqrt{\frac{\nu}{2}}i)^{\frac{\nu-1}{2}}}, \\ f^{(\frac{\nu-1}{2}-k)} &= \frac{(-1)^{\frac{\nu-1}{2}-k-1}(\frac{\nu-1}{2}-k-1)!}{(z - \sqrt{\frac{\nu}{2}}i)^{\frac{\nu-1}{2}-k}}. \end{aligned}$$

Let $g = \frac{1}{(z-i)^{\frac{\nu+1}{2}}}$, we have

$$\begin{aligned} g^{(1)} &= \left(\frac{-(\nu+1)}{2}\right) \frac{1}{(z-i)^{\frac{\nu+3}{2}}}, \\ g^{(2)} &= \left(\frac{-(\nu+1)}{2}\right) \left(\frac{-(\nu+3)}{2}\right) \frac{1}{(z-i)^{\frac{\nu+5}{2}}}. \end{aligned}$$

In general,

$$\begin{aligned} g^{(\frac{\nu-1}{2})} &= (-1)^{\frac{\nu-1}{2}} \left(\frac{\nu+1}{2}\right) \left(\frac{\nu+3}{2}\right) \dots (\nu-1) \frac{1}{(z-i)^\nu}, \\ g^{(k)} &= (-1)^k \left(\frac{\nu+1}{2}\right) \left(\frac{\nu+3}{2}\right) \dots \left(\frac{\nu+1}{2} + k - 1\right) \frac{1}{(z-i)^{\frac{\nu+1}{2}+k}}. \end{aligned}$$

Apply general Leibniz rule, we obtain

$$(f.g)^{\frac{\nu-1}{2}} = f^{\frac{\nu-1}{2}}g + fg^{\frac{\nu-1}{2}} + \sum_{k=1}^{\frac{\nu-3}{2}} \binom{\frac{\nu-1}{2}}{k} f^{\frac{\nu-1}{2}-k}g^k.$$

Taking the limit, we have

$$\begin{aligned} \lim_{z \rightarrow -i} (fg)^{\frac{\nu-1}{2}} &= \frac{(-1)^{\frac{\nu-3}{2}} (\frac{\nu-3}{2})!}{(-1)^{\frac{\nu-1}{2}} (1 + \sqrt{\frac{\nu}{\nu}})^{\frac{\nu-1}{2}} (i)^{\frac{\nu-1}{2}} (-2i)^{\frac{\nu+1}{2}}} \\ &+ \frac{\log(1 + \sqrt{\frac{\nu}{\nu}}) (-1)^{\frac{\nu-1}{2}} (\frac{\nu+1}{2}) (\frac{\nu+3}{2}) \dots (\nu-1)}{(-2i)^\nu} \\ &+ \sum_{k=1}^{\frac{\nu-3}{2}} \left\{ \binom{\frac{\nu-1}{2}}{k} \left[\frac{(-1)^{\frac{\nu-1}{2}-k-1} (\frac{\nu-1}{2}-k-1)!}{(-1)^{\frac{\nu-1}{2}} (1 + \sqrt{\frac{\nu}{\nu}})^{\frac{\nu-1}{2}} i^{\frac{\nu-1}{2}}} \right] \left[\frac{(-1)^k (\frac{\nu+1}{2}) (\frac{\nu+3}{2}) \dots (\frac{\nu+1}{2} + k - 1)}{(-2i)^{\frac{\nu+1}{2}-k}} \right] \right\}. \end{aligned}$$

We have

$$\begin{aligned} (i)^{\frac{\nu-1}{2}} (i)^{\frac{\nu+1}{2}} &= i^\nu = (-1)^{\frac{\nu}{2}}, \\ (-1)^{\frac{\nu-1}{2}} (-1)^{\frac{\nu+1}{2}} (-1)^{\frac{\nu}{2}} (-1)^{\frac{3-\nu}{2}} &= (i)^{2\nu+3} = i, \\ (-1)^\nu (-1)^{\frac{1-\nu}{2}} (-1)^{\frac{\nu}{2}} &= (-1)^{\nu+\frac{1}{2}} = (i)^{2\nu+1} = -i, \\ (i)^{\frac{\nu-1}{2}-k} (i)^{\frac{\nu+1}{2}+k} &= i^\nu = (-1)^{\frac{\nu}{2}}, \\ (-1)^{\frac{\nu}{2}} (-1)^{\frac{\nu-1}{2}-k} (-1)^{\frac{1-\nu}{2}+k} (-1)^{1+\frac{1-\nu}{2}} &= (i)^{2\nu+3} = (-1)^{\nu+1} i = i. \end{aligned}$$

since ν is odd. Hence,

$$\begin{aligned} \lim_{z \rightarrow -i} (fg)^{\frac{\nu-1}{2}} &= \frac{(\frac{\nu-3}{2})!}{(1 + \sqrt{\frac{\nu}{\nu}})^{\frac{\nu-1}{2}} (2)^{\frac{\nu+1}{2}} i^{2\nu+3}} + \frac{\log(1 + \sqrt{\frac{\nu}{\nu}}) (\frac{\nu+1}{2}) (\frac{\nu+3}{2}) \dots (\nu-1)}{(2)^\nu i^{2\nu+1}} \\ &+ \sum_{k=1}^{\frac{\nu-3}{2}} \binom{\frac{\nu-1}{2}}{k} \frac{(\frac{\nu-1}{2}-k-1)! (\frac{\nu+1}{2}) (\frac{\nu+3}{2}) \dots (\frac{\nu+1}{2} + k - 1)}{(1 + \sqrt{\frac{\nu}{\nu}})^{\frac{\nu-1}{2}-k} (2)^{\frac{\nu+1}{2}+k} i^{2\nu+3}} \\ &= \frac{(\frac{\nu-3}{2})!}{(1 + \sqrt{\frac{\nu}{\nu}})^{\frac{\nu-1}{2}} (2)^{\frac{\nu+1}{2}} i} - \frac{\log(1 + \sqrt{\frac{\nu}{\nu}}) (\frac{\nu+1}{2}) (\frac{\nu+3}{2}) \dots (\nu-1)}{2^\nu i} \\ &+ \sum_{k=1}^{\frac{\nu-3}{2}} \binom{\frac{\nu-1}{2}}{k} \frac{(\frac{\nu-1}{2}-k-1)! (\frac{\nu+1}{2}) (\frac{\nu+3}{2}) \dots (\frac{\nu+1}{2} + k - 1)}{(1 + \sqrt{\frac{\nu}{\nu}})^{\frac{\nu-1}{2}-k} (2)^{\frac{\nu+1}{2}+k} i}. \end{aligned}$$

Hence,

$$I = \frac{\pi}{\left(\frac{\nu-1}{2}\right)!} \left[-\frac{\left(\frac{\nu-3}{2}\right)!}{2^{\frac{\nu-3}{2}} \left(1 + \sqrt{\frac{\bar{\nu}}{\nu}}\right)^{\frac{\nu-1}{2}}} + \frac{\log\left(1 + \sqrt{\frac{\bar{\nu}}{\nu}}\right) \left(\frac{\nu+1}{2}\right) \dots (\nu-1)}{2^{\nu-2}} \right. \\ \left. - \sum_{k=1}^{\frac{\nu-3}{2}} \binom{\frac{\nu-1}{2}}{k} \frac{\left(\frac{\nu-1}{2} - k\right)! \left(\frac{\nu+1}{2}\right) \left(\frac{\nu+3}{2}\right) \dots \left(\frac{\nu+1}{2} + k - 1\right)}{2^{\frac{\nu-3}{2} + k} \left(1 + \sqrt{\frac{\bar{\nu}}{\nu}}\right)^{\frac{\nu-1}{2} - k}} \right].$$

Substituting the expression of I above and $Ent(Y)$ into (1), we are done. \square

5.5.4. Some Specific Cases

1. Case $\nu = 3$

$$B\left(\frac{1}{2}, \frac{3}{2}\right) = \frac{\pi}{2}.$$

$$Ent(Y) = \log(\sqrt{\nu}B\left(\frac{1}{2}, \frac{\nu}{2}\right)) + (\nu + 1)\left[\log(2) - \frac{2}{\nu + 1}\right] \\ = \log(\sqrt{3}B\left(\frac{1}{2}, \frac{3}{2}\right)) + 4\left[\log(2) - \frac{2}{4}\right] = \log(\sqrt{3}\frac{\pi}{2}) + 4\left[\log(2) - \frac{1}{2}\right] \\ = \log(8\sqrt{3}\pi) - 2.$$

$$J = 0.$$

$$KL(f_{\nu}(y)|f_{\bar{\nu}}(y)) = -\left\{\log(\sqrt{\nu}B\left(\frac{1}{2}, \frac{\nu}{2}\right)) + (\nu + 1)\left[\log(2) - \frac{2}{\nu + 1}\right]\right\} \\ + \log(\sqrt{\bar{\nu}}B\left(\frac{1}{2}, \frac{\bar{\nu}}{2}\right)) - \frac{\bar{\nu} + 1}{2B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \left[\frac{\pi}{\left(\frac{\nu-1}{2}\right)2^{\frac{\nu-3}{2}} \left(1 + \sqrt{\frac{\bar{\nu}}{\nu}}\right)^{\frac{\nu-1}{2}}} + J \right] \\ + (\bar{\nu} + 1)\left[\log\left(1 + \sqrt{\frac{\bar{\nu}}{\nu}}\right)\right] \\ = -(\log(8\sqrt{3}\pi) - 2) + \log(\sqrt{\bar{\nu}}B\left(\frac{1}{2}, \frac{\bar{\nu}}{2}\right)) - \frac{\bar{\nu} + 1}{2} \left[\frac{2}{\left(1 + \sqrt{\frac{\bar{\nu}}{\nu}}\right)} \right] \\ + (\bar{\nu} + 1)\left[\log\left(1 + \sqrt{\frac{\bar{\nu}}{\nu}}\right)\right] \\ = (2 - \log(8\sqrt{3}\pi) + \log(\sqrt{\bar{\nu}}B\left(\frac{1}{2}, \frac{\bar{\nu}}{2}\right)) + (\bar{\nu} + 1)\left[\log\left(1 + \sqrt{\frac{\bar{\nu}}{\nu}}\right)\right] - \frac{1}{\left(1 + \sqrt{\frac{\bar{\nu}}{\nu}}\right)} \right].$$

2. Case $\nu = 5$

$$\begin{aligned}
 B\left(\frac{1}{2}, \frac{5}{2}\right) &= \frac{\pi(3.4)}{2^4 2!} = \frac{3\pi}{8}. \\
 Ent(Y) &= \log(\sqrt{\nu}B\left(\frac{1}{2}, \frac{\nu}{2}\right)) + (\nu + 1)\left[\log(2) - \frac{2}{\nu + 1} - \frac{2}{\nu + 3}\right] \\
 &= \log(\sqrt{5}B\left(\frac{1}{2}, \frac{5}{2}\right)) + 6\left[\log(2) - \frac{2}{4} - \frac{2}{8}\right] \\
 &= \log(\sqrt{5}\frac{3\pi}{8}) + 6\left[\log(2) - \frac{1}{2} - \frac{1}{3}\right] = \log(24\sqrt{5}\pi) - 5. \\
 J &= \frac{\pi}{\left(\frac{\nu-1}{2}\right)!} \binom{\frac{\nu-1}{2}}{1} \frac{\left(\frac{\nu-1}{2} - 1\right)!\left(\frac{\nu+1}{2}\right)}{2^{\frac{\nu-3}{2}+1}\left(1 + \sqrt{\frac{\bar{\nu}}{\nu}}\right)^{\frac{\nu-1}{2}-1}} \\
 &= \frac{\pi}{2} \frac{2.3}{4\left(1 + \sqrt{\frac{\bar{\nu}}{5}}\right)} = \frac{3\pi}{4\left(1 + \sqrt{\frac{\bar{\nu}}{5}}\right)}. \\
 KL(f_\nu(y)|f_{\bar{\nu}}(y)) &= -\left\{\log(\sqrt{\nu}B\left(\frac{1}{2}, \frac{\nu}{2}\right)) + (\nu + 1)\left[\log(2) - \frac{2}{\nu + 1} - \frac{2}{\nu + 3}\right]\right\} \\
 &+ \log(\sqrt{\bar{\nu}}B\left(\frac{1}{2}, \frac{\bar{\nu}}{2}\right)) - \frac{\bar{\nu} + 1}{2B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \left[\frac{\pi}{\left(\frac{\nu-1}{2}\right)2^{\frac{\nu-3}{2}}\left(1 + \sqrt{\frac{\bar{\nu}}{\nu}}\right)^{\frac{\nu-1}{2}}} + J \right] \\
 &+ (\bar{\nu} + 1)\left[\log\left(1 + \sqrt{\frac{\bar{\nu}}{\bar{\nu}}}\right)\right] \\
 &= -(\log(24\sqrt{5}\pi) - 5) + \log(\sqrt{\bar{\nu}}B\left(\frac{1}{2}, \frac{\bar{\nu}}{2}\right)) \\
 &- \frac{\bar{\nu} + 1}{2} \left[\frac{1!2^3}{(3)(4)\left(1 + \sqrt{\frac{\bar{\nu}}{5}}\right)^2} + \frac{2}{1 + \sqrt{\frac{\bar{\nu}}{5}}} \right] + (\bar{\nu} + 1)\left[\log\left(1 + \sqrt{\frac{5}{\bar{\nu}}}\right)\right] \\
 &= (5 - \log(24\sqrt{5}\pi) + \log(\sqrt{\bar{\nu}}B\left(\frac{1}{2}, \frac{\bar{\nu}}{2}\right)) + (\bar{\nu} + 1)\left[\log\left(1 + \sqrt{\frac{5}{\bar{\nu}}}\right)\right] \\
 &- \frac{1}{3\left(1 + \sqrt{\frac{\bar{\nu}}{5}}\right)^2} - \frac{1}{1 + \sqrt{\frac{\bar{\nu}}{5}}}\right].
 \end{aligned}$$

3. Case $\nu = 7$

$$\begin{aligned}
B\left(\frac{1}{2}, \frac{7}{2}\right) &= \frac{\pi(4.5.6)}{2^6 3!} = \frac{5\pi}{16}. \\
Ent(Y) &= \log(\sqrt{\nu}B\left(\frac{1}{2}, \frac{\nu}{2}\right)) + (\nu + 1)\left[\log(2) - \frac{2}{\nu + 1} - \frac{2}{\nu + 3} - \frac{2}{\nu + 5}\right] \\
&= \log(\sqrt{7}B\left(\frac{1}{2}, \frac{7}{2}\right)) + 8\left[\log(2) - \frac{2}{8} - \frac{2}{10} - \frac{2}{12}\right] \\
&= \log(\sqrt{7}\frac{5\pi}{16}) + 8\left[\log(2) - \frac{1}{4} - \frac{1}{5} - \frac{1}{6}\right] = \log(80\sqrt{7}\pi) - \frac{74}{15}. \\
J &= \frac{\pi}{\left(\frac{\nu-1}{2}\right)!} \left[\binom{\frac{\nu-1}{2}}{1} \frac{\left(\frac{\nu-1}{2} - 1\right)! \left(\frac{\nu+1}{2}\right)}{2^{\frac{\nu-3}{2}+1} (1 + \sqrt{\frac{\bar{\nu}}{\nu}})^{\frac{\nu-1}{2}-1}} + \binom{\frac{\nu-1}{2}}{2} \frac{\left(\frac{\nu-1}{2} - 2\right)! \left(\frac{\nu+1}{2}\right) \left(\frac{\nu+3}{2}\right)}{2^{\frac{\nu-3}{2}+2} (1 + \sqrt{\frac{\bar{\nu}}{\nu}})^{\frac{\nu-1}{2}-2}} \right] \\
&= \frac{\pi}{6} \left[3 \frac{(2!) \cdot 4}{8(1 + \sqrt{\frac{\bar{\nu}}{\nu}})^2} + 3 \frac{(1!) \cdot 4 \cdot 5}{16(1 + \sqrt{\frac{\bar{\nu}}{\nu}})} \right] = \frac{\pi}{2(1 + \sqrt{\frac{\bar{\nu}}{7}})^2} + \frac{5\pi}{8(1 + \sqrt{\frac{\bar{\nu}}{7}})}. \\
KL(f_\nu(y)|f_{\bar{\nu}}(y)) &= -\left\{ \log(\sqrt{\nu}B\left(\frac{1}{2}, \frac{\nu}{2}\right)) + (\nu + 1)\left[\log(2) - \frac{2}{\nu + 1} - \frac{2}{\nu + 3} - \frac{2}{\nu + 5}\right] \right\} \\
&+ \log(\sqrt{\bar{\nu}}B\left(\frac{1}{2}, \frac{\bar{\nu}}{2}\right)) - \frac{\bar{\nu} + 1}{2B\left(\frac{1}{2}, \frac{\bar{\nu}}{2}\right)} \left[\frac{\pi}{\left(\frac{\nu-1}{2}\right) 2^{\frac{\nu-3}{2}} (1 + \sqrt{\frac{\bar{\nu}}{\nu}})^{\frac{\nu-1}{2}}} + J \right] \\
&+ (\bar{\nu} + 1)\left[\log\left(1 + \sqrt{\frac{\bar{\nu}}{\nu}}\right)\right] \\
&= -\left(\log(80\sqrt{7}\pi) - \frac{74}{15}\right) + \log(\sqrt{\bar{\nu}}B\left(\frac{1}{2}, \frac{\bar{\nu}}{2}\right)) + (\bar{\nu} + 1)\left[\log\left(1 + \sqrt{\frac{7}{\bar{\nu}}}\right)\right] \\
&- \frac{\bar{\nu} + 1}{2} \left[\frac{(2)! 2^4}{4.5.6(1 + \sqrt{\frac{\bar{\nu}}{\nu}})^3} + \frac{16}{5\pi} \left(\frac{\pi}{2(1 + \sqrt{\frac{\bar{\nu}}{7}})^2} + \frac{5\pi}{8(1 + \sqrt{\frac{\bar{\nu}}{7}})} \right) \right] \\
&= \frac{74}{15} - \log(80\sqrt{7}\pi) + \log(\sqrt{\bar{\nu}}B\left(\frac{1}{2}, \frac{\bar{\nu}}{2}\right)) + (\bar{\nu} + 1)\left[\log\left(1 + \sqrt{\frac{7}{\bar{\nu}}}\right)\right] \\
&- \frac{2}{15(1 + \sqrt{\frac{\bar{\nu}}{7}})^3} - \frac{4}{5(1 + \sqrt{\frac{\bar{\nu}}{7}})^2} - \frac{1}{1 + \sqrt{\frac{\bar{\nu}}{7}}}.
\end{aligned}$$

5.6. Entropy and the KL Divergence of Rational Densities

We use the fact that any rational density $r(x) = \frac{N_1(x)}{D_1(x)}$, where two polynomials $N_1(x)$ and $D_1(x)$ are co-prime, can be decomposed as a unique linear combination of functions of the form $\frac{Ax+B}{((x-u)^2+a^2)^k}$, where k runs from 1 up to the degree of $u+ia$, $a > 0$ as a pole of the rational density.

Consider an integrable and positive rational function $q(x) = \frac{N_2(x)}{D_2(x)}$, where the two polynomials $N_2(x)$ and $D_2(x)$ are co-prime. We have that both $N_2(x)$ and $D_2(x)$ are either positive or negative on the whole real line because $q(x)$ is positive. Without loss of generality, we take $N_2(x)$ and $D_2(x)$ both positive polynomials with degrees of n and m , respectively. Of course, n and m must be even because we consider positive polynomials on the whole real line. Then we use the fact that any positive polynomial $N_2(x)$ of degree n has the spectral factors $\psi(z) = (z-z_1)(z-z_2)\dots(z-z_{\frac{n}{2}})$, $N_2(z) = |\psi(z)|^2$, and $N_2(z)$ has $\frac{n}{2}$ zeros $z_1, z_2, \dots, z_{\frac{n}{2}}$ in the upper half plane. Similarly, any positive polynomial $D_2(x)$ of degree m has the spectral factors $\hat{\psi}(z) = (z-\hat{z}_1)(z-\hat{z}_2)\dots(z-\hat{z}_{\frac{m}{2}})$, $D_2(z) = |\hat{\psi}(z)|^2$, and $D_2(z)$ has $\frac{m}{2}$ zeros $\hat{z}_1, \hat{z}_2, \dots, \hat{z}_{\frac{m}{2}}$ in the upper half plane. Consider

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{Ax+B}{((x-u)^2+a^2)^k} \log(q(x))dx &= \int_{-\infty}^{\infty} \frac{Ax+B}{((x-u)^2+a^2)^k} \log(N_2(x))dx \\ &+ \int_{-\infty}^{\infty} \frac{Ax+B}{((x-u)^2+a^2)^k} \log(D_2(x))dx. \end{aligned}$$

Let $v = \frac{x-u}{a}$, so $x = av + u$, and $dx = adv$.

Let $\hat{N}_2(v) := N_2(av + u)$, a polynomial in v of degree n . Then

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{Ax+B}{((x-u)^2+a^2)^k} \log(N_2(x))dx &= a^{1-2k}(Au+B) \int_{-\infty}^{\infty} \frac{\log(\hat{N}_2(v))}{(v^2+1)^k} dv \\ &+ Aa^{2-2k} \int_{-\infty}^{\infty} \frac{v \log(\hat{N}_2(v))}{(v^2+1)^k} dv. \end{aligned}$$

We can calculate the first term by applying Theorem 5.2.2 directly while we apply the integration by parts to the second term, as follows.

Observe that

$$\int \frac{2v}{(v^2 + 1)^k} dv = \frac{(v^2 + 1)^{(1-k)}}{1-k}, k \neq 1.$$

If $k > 1$, we have

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{2v \log(\hat{N}_2(v))}{(v^2 + 1)^k} dv &= \frac{(v^2 + 1)^{(1-k)}}{1-k} \log(\hat{N}_2(v)) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{\hat{N}'_2(v)}{\hat{N}_2(v)} \frac{(v^2 + 1)^{(1-k)}}{1-k} dv \\ &= 0 - \int_{-\infty}^{\infty} \frac{\hat{N}'_2(v)}{\hat{N}_2(v)} \frac{(v^2 + 1)^{(1-k)}}{1-k} dv = - \int_{-\infty}^{\infty} \frac{\hat{N}'_2(v)}{\hat{N}_2(v)} \frac{(v^2 + 1)^{(1-k)}}{1-k} dv. \end{aligned}$$

Similarly, we can handle

$$\int_{-\infty}^{\infty} \frac{Ax + B}{((x - u)^2 + a^2)^k} \log(D_2(x)) dx$$

when $k > 1$.

If $k = 1$, the numerator must be a constant, $A = 0$, as the codegree has to be greater than two for a rational function to be a pdf as the function will need to be integrable. Consider

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{B}{((x - u)^2 + a^2)^k} \log(q(x)) dx &= \int_{-\infty}^{\infty} \frac{B}{((x - u)^2 + a^2)^k} \log(N_2(x)) dx \\ &\quad - \int_{-\infty}^{\infty} \frac{B}{((x - u)^2 + a^2)^k} \log(D_2(x)) dx. \end{aligned}$$

We have

$$\int_{-\infty}^{\infty} \frac{B}{((x - u)^2 + a^2)^k} \log(N_2(x)) dx = Ba^{1-2k} \int_{-\infty}^{\infty} \frac{\log(\hat{N}_2(v))}{(v^2 + 1)^k} dv,$$

which can be calculated directly by applying theorem 5.2.2. Similarly, we can handle the second term

$$\int_{-\infty}^{\infty} \frac{B}{((x-u)^2 + a^2)^k} \log(D_2(x)) dx.$$

Remark: Any nonnegative function can be viewed as limiting case of positive functions.

Hence, the results found can be extended to non-negative rational densities.

Some examples are illustrated, as follows.

5.6.1. Entropy and the KL Divergence of a Linear Convex Combination of two Cauchy Densities

In this section, we will consider an example of the linear convex combination of two cauchy densities with the equal weights, $p_{l,s}(x) = \frac{1}{2}p_{l_1,s_1}(x) + \frac{1}{2}p_{l_2,s_2}(x)$, where $s_1 = 1, l_1 = 0, l_2 = 1, s_2 = 2$.

$$p_{l,s}(x) = \frac{1}{2} \frac{1}{\pi(1+x^2)} + \frac{1}{2} \frac{2}{\pi(4+(x-1)^2)} = \frac{3x^2 - 2x + 7}{2\pi(x^2+1)((x-1)^2+4)}.$$

The entropy is given by

$$\begin{aligned} Ent(p_{l,s}(x)) &= - \int_{-\infty}^{\infty} p_{l,s}(x) \log(p_{l,s}(x)) dx \\ &= - \frac{1}{2} \int_{-\infty}^{\infty} p_{l_1,s_1}(x) \log(p_{l,s}(x)) dx - \frac{1}{2} \int_{-\infty}^{\infty} p_{l_2,s_2}(x) \log(p_{l,s}(x)) dx. \end{aligned}$$

$$\begin{aligned}
\int_{-\infty}^{\infty} p_{l_1, s_1}(x) \log(p_{l, s}(x)) dx &= \int_{-\infty}^{\infty} p_{l_1, s_1}(x) \log(3x^2 - 2x + 7) dx \\
&- \log(2\pi) \left(\int_{-\infty}^{\infty} p_{l_1, s_1}(x) dx \right) \\
&- \int_{-\infty}^{\infty} p_{l_1, s_1}(x) \log(x^2 + 1) dx \\
&- \int_{-\infty}^{\infty} p_{l_1, s_1}(x) \log((x-1)^2 + 4) dx.
\end{aligned}$$

We have $\int_{-\infty}^{\infty} p_{l_1, s_1}(x) dx = 1$, so

$$\begin{aligned}
\int_{-\infty}^{\infty} p_{l_1, s_1}(x) \log(p_{l, s}(x)) dx &= \int_{-\infty}^{\infty} p_{l_1, s_1}(x) \log(3x^2 - 2x + 7) dx - \log(2\pi) \\
&- \int_{-\infty}^{\infty} p_{l_1, s_1}(x) \log(x^2 + 1) dx \\
&- \int_{-\infty}^{\infty} p_{l_1, s_1}(x) \log((x-1)^2 + 4) dx(1)
\end{aligned}$$

We have

$$\begin{aligned}
I_1 &= \int_{-\infty}^{\infty} p_{l_1, s_1}(x) \log(3x^2 - 2x + 7) dx = \log(3) \left(\int_{-\infty}^{\infty} p_{l_1, s_1}(x) dx \right) \\
&+ \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\log(x^2 - \frac{2}{3}x + \frac{7}{3})}{x^2 + 1} dx = \log(3) + \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\log(x^2 - \frac{2}{3}x + \frac{7}{3})}{x^2 + 1} dx,
\end{aligned}$$

and

$$\int_{-\infty}^{\infty} \frac{\log(x^2 - \frac{2}{3}x + \frac{7}{3})}{x^2 + 1} dx = \pi \log\left(\frac{7}{3} + 1 + \sqrt{\left| \frac{-80}{9} \right|}\right) = \pi \log\left(\frac{10 + \sqrt{80}}{3}\right),$$

using Proposition 3.2.4 (i). Therefore,

$$I_1 = \log(3) + \log\left(\frac{10 + \sqrt{80}}{3}\right) = \log(10 + \sqrt{80}).$$

Next, we calculate

$$I_2 = \int_{-\infty}^{\infty} p_{l_1, s_1}(x) \log(x^2 + 1) dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\log(x^2 + 1)}{x^2 + 1} dx = \left(\frac{1}{\pi}\right)(\pi \log(4)) = \log(4).$$

Finally, we calculate

$$\begin{aligned} I_3 &= \int_{-\infty}^{\infty} p_{l_1, s_1}(x) \log((x-1)^2 + 4) dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\log(x^2 - 2x + 5)}{x^2 + 1} dx \\ &= \left(\frac{1}{\pi}\right)(\pi \log(5 + 1 + \sqrt{|-16|})) = \left(\frac{1}{\pi}\right)(\pi \log(10)) = \log(10), \end{aligned}$$

using Proposition 3.2.4 (i).

Substituting I_1 , I_2 , and I_3 into (1), we have

$$\int_{-\infty}^{\infty} p_{l_1, s_1}(x) \log(p_{l, s}(x)) dx = \log(10 + \sqrt{80}) - \log(2\pi) - \log(4) - \log(10) = \log\left(\frac{10 + \sqrt{80}}{80\pi}\right).$$

Similarly, we can calculate

$$\begin{aligned} \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log(p_{l, s}(x)) dx &= \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log(3x^2 - 2x + 7) dx \\ &\quad - \log(2\pi) \left(\int_{-\infty}^{\infty} p_{l_2, s_2}(x) dx \right) \\ &\quad - \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log(x^2 + 1) dx \\ &\quad - \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log((x-1)^2 + 4) dx. \end{aligned}$$

We have $\int_{-\infty}^{\infty} p_{l_2, s_2}(x) dx = 1$, so

$$\begin{aligned} \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log(p_{l, s}(x)) dx &= \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log(3x^2 - 2x + 7) dx - \log(2\pi) \\ &\quad - \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log(x^2 + 1) dx \\ &\quad - \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log((x-1)^2 + 4) dx. \quad (2) \end{aligned}$$

We have

$$J_1 = \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log(3x^2 - 2x + 7) dx = \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log(3x^2 - 2x + 7)}{(x-1)^2 + 4} dx.$$

Let $u = \frac{x-1}{4}$, so $dx = 4du$ and

$$\begin{aligned} J_1 &= \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log(48u^2 + 16u + 8)}{1 + u^2} du = \frac{2 \log(48)}{\pi} \int_{-\infty}^{\infty} \frac{du}{1 + u^2} + \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log(u^2 + \frac{1}{3}u + \frac{1}{6})}{1 + u^2} du \\ &= 2 \log(48) + \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log(u^2 + \frac{1}{3}u + \frac{1}{6})}{1 + u^2} du = 2 \log(48) + \frac{2}{\pi} \pi \log\left(\frac{7 + 2\sqrt{5}}{6}\right) \\ &= 2 \log(56 + 16\sqrt{5}). \end{aligned}$$

Next, we calculate

$$\begin{aligned} J_2 &= \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log(x^2 + 1) dx = \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log(x^2 + 1)}{(x-1)^2 + 4} dx \\ &= \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log(x^2 + 1)}{x^2 - 2x + 5} dx = \frac{2}{\pi} \pi \log(|\psi(1 - 2i)|^2) = \log(10), \end{aligned}$$

using Proposition 3.2.4 (ii) with $\psi(z) = 1 + iz$.

Finally, we calculate

$$\begin{aligned} J_3 &= \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log((x-1)^2 + 4) dx = \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log((x-1)^2 + 4)}{(x-1)^2 + 4} dx \\ &= \frac{2 \log(4)}{\pi} \int_{-\infty}^{\infty} \frac{du}{u^2 + 1} + \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log(u^2 + 1)}{u^2 + 1} du = \frac{2 \log(4)}{\pi} \pi + \frac{2}{\pi} \pi \log(4) = 2 \log(16). \end{aligned}$$

Substituting J_1 , J_2 , and J_3 into (2), we have

$$\begin{aligned} \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log(p_{l, s}(x)) dx &= 2 \log(56 + 16\sqrt{5}) - \log(2\pi) + \\ &- \log(10) - 2 \log(16) = \log\left(\frac{(56 + 16\sqrt{5})^2}{2\pi 10(16)^2}\right) = \log\left(\frac{(14 + 4\sqrt{5})^2}{320\pi}\right). \end{aligned}$$

Therefore,

$$Ent(p_{l, s}(x)) = -\frac{1}{2} \log\left(\frac{10 + \sqrt{80}}{80\pi}\right) - \frac{1}{2} \log\left(\frac{(14 + 4\sqrt{5})^2}{320\pi}\right) = \log\left(\frac{160\pi}{\sqrt{10 + \sqrt{80}}(14 + 4\sqrt{5})}\right).$$

Now we consider $p_{\bar{l},\bar{s}} = \frac{1}{2}p_{\bar{l}_1,\bar{s}_1} + \frac{1}{2}p_{\bar{l}_2,\bar{s}_2}$, where $\bar{l}_1 = 2$, $\bar{s}_1 = 3$, $\bar{l}_2 = 3$, $\bar{s}_2 = 4$.

$$p_{\bar{l},\bar{s}} = \frac{1}{2} \frac{3}{\pi(9 + (x-2)^2)} + \frac{1}{2} \frac{4}{\pi(16 + (x-3)^2)} = \frac{7x^2 - 34x + 127}{2\pi(9 + (x-2)^2)(16 + (x-3)^2)}.$$

The KL divergence of $p_{l,s}$ and $p_{\bar{l},\bar{s}}$ is given by

$$\begin{aligned} KL(p_{l,s}(x)|p_{\bar{l},\bar{s}}(x)) &= \int_{-\infty}^{\infty} p_{l,s}(x) \log\left[\frac{p_{l,s}(x)}{p_{\bar{l},\bar{s}}(x)}\right] dx \\ &= \int_{-\infty}^{\infty} p_{l,s}(x) \log(p_{l,s}(x)) dx - \int_{-\infty}^{\infty} p_{l,s}(x) \log(p_{\bar{l},\bar{s}}(x)) dx \\ &= -Ent(p_{l,s}(x)) - \int_{-\infty}^{\infty} p_{l,s}(x) \log(p_{\bar{l},\bar{s}}(x)) dx. \end{aligned}$$

We have

$$\int_{-\infty}^{\infty} p_{l,s}(x) \log(p_{\bar{l},\bar{s}}(x)) dx = \frac{1}{2} \int_{-\infty}^{\infty} p_{l_1,s_1}(x) \log(p_{\bar{l},\bar{s}}(x)) dx + \frac{1}{2} \int_{-\infty}^{\infty} p_{l_2,s_2}(x) \log(p_{\bar{l},\bar{s}}(x)) dx. \quad (3)$$

First, we calculate

$$\begin{aligned} K_1 &= \int_{-\infty}^{\infty} p_{l_1,s_1}(x) \log(p_{\bar{l},\bar{s}}(x)) dx = \int_{-\infty}^{\infty} p_{l_1,s_1}(x) \log(7x^2 - 34x + 127) dx \\ &\quad - \log(2\pi) \left(\int_{-\infty}^{\infty} p_{l_1,s_1}(x) dx \right) - \int_{-\infty}^{\infty} p_{l_1,s_1}(x) \log(9 + (x-2)^2) dx \\ &\quad - \int_{-\infty}^{\infty} p_{l_1,s_1}(x) \log(16 + (x-3)^2) dx. \end{aligned}$$

We have

$$\int_{-\infty}^{\infty} p_{l_1,s_1}(x) dx = 1.$$

We calculate

$$\begin{aligned}
\int_{-\infty}^{\infty} p_{l_1, s_1}(x) \log(7x^2 - 34x + 127) dx &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\log(7x^2 - 34x + 127)}{x^2 + 1} dx \\
&= \frac{1}{\pi} \log(7) \left(\int_{-\infty}^{\infty} \frac{dx}{x^2 + 1} \right) \\
&+ \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\log(x^2 - \frac{34}{7}x + \frac{127}{7})}{x^2 + 1} dx \\
&= \log(7) + \log\left(\frac{134 + 20\sqrt{6}}{7}\right) \\
&= \log(134 + 20\sqrt{6}).
\end{aligned}$$

$$\begin{aligned}
\int_{-\infty}^{\infty} p_{l_1, s_1}(x) \log((x-2)^2 + 9) dx &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\log(x^2 - 4x + 13)}{x^2 + 1} dx \\
&= \frac{1}{\pi} (\pi \log(13 + 1 + \sqrt{|-36|})) = \log(20),
\end{aligned}$$

using Proposition 3.2.4 (i). Furthermore, we have

$$\begin{aligned}
\int_{-\infty}^{\infty} p_{l_1, s_1}(x) \log((x-3)^2 + 16) dx &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\log(x^2 - 6x + 25)}{x^2 + 1} dx \\
&= \frac{1}{\pi} (\pi \log(25 + 1 + \sqrt{|-64|})) = \log(34),
\end{aligned}$$

using Proposition 3.2.4 (i). Hence,

$$K_1 = \log(134 + 20\sqrt{6}) - \log(2\pi) - \log(20) - \log(34) = \log\left(\frac{67 + 10\sqrt{6}}{680\pi}\right).$$

Similarly, we calculate

$$\begin{aligned}
K_2 &= \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log(p_{\bar{l}, \bar{s}}(x)) dx = \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log(7x^2 - 34x + 127) dx \\
&- \log(2\pi) \left(\int_{-\infty}^{\infty} p_{l_2, s_2}(x) dx \right) - \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log(9 + (x-2)^2) dx \\
&- \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log(16 + (x-3)^2) dx.
\end{aligned}$$

We have

$$\int_{-\infty}^{\infty} p_{l_2, s_2}(x) dx = 1.$$

We calculate

$$\begin{aligned} \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log(7x^2 - 34x + 127) dx &= \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log(7x^2 - 34x + 127)}{(x-1)^2 + 4} dx \\ &= \frac{2 \log(112)}{\pi} \left(\int_{-\infty}^{\infty} \frac{du}{1+u^2} \right) \\ &+ \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log(u^2 - \frac{5}{7}u + \frac{25}{28})}{u^2 + 1} du \\ &= 2 \log(112) + \frac{2}{\pi} (\pi \log(\frac{25}{28} + 1 + \sqrt{|\frac{-150}{49}|})) \\ &= 2 \log(112) + 2 \log(\frac{53 + 4\sqrt{150}}{28}) \\ &= 2 \log(212 + 16\sqrt{150}). \end{aligned}$$

And we have

$$\begin{aligned} \int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log((x-2)^2 + 9) dx &= \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log(x^2 - 4x + 13)}{(x-1)^2 + 4} dx \\ &= \int_{-\infty}^{\infty} \frac{\log(16u^2 - 8u + 10)}{u^2 + 1} du \\ &= \frac{2 \log(16)}{\pi} \left(\int_{-\infty}^{\infty} \frac{du}{u^2 + 1} \right) \\ &+ \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log(u^2 - \frac{1}{2}u + \frac{5}{8})}{u^2 + 1} du \\ &= 2 \log(16) + 2 \log(\frac{5}{8} + 1 + \sqrt{|\frac{-9}{4}|}) \\ &= 2 \log(16) + 2 \log(\frac{25}{8}) = 2 \log(50). \end{aligned}$$

We have

$$\begin{aligned}
\int_{-\infty}^{\infty} p_{l_2, s_2}(x) \log((x-3)^2 + 16) dx &= \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log((x-3)^2 + 16)}{(x-1)^2 + 4} dx \\
&= \int_{-\infty}^{\infty} \frac{\log(16u^2 - 16u + 20)}{u^2 + 1} \\
&= \frac{2 \log(16)}{\pi} \left(\int_{-\infty}^{\infty} \frac{du}{u^2 + 1} \right) + \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log(u^2 - u + \frac{5}{4})}{u^2 + 1} \\
&= 2 \log(16) + 2 \log\left(\frac{5}{4} + 1 + \sqrt{|-4|}\right) \\
&= 2 \log(16) + 2 \log\left(\frac{17}{4}\right) = 2 \log(68).
\end{aligned}$$

Hence,

$$K_2 = 2 \log(212 + 16\sqrt{150}) - \log(2\pi) - 2 \log(50) - 2 \log(68) = 2 \log\left(\frac{53 + 4\sqrt{150}}{\sqrt{2\pi}850}\right).$$

Substituting K_1 , K_2 into (3), we obtain

$$\begin{aligned}
\int_{-\infty}^{\infty} p_{l, s}(x) \log(p_{\bar{l}, \bar{s}}(x)) dx &= \frac{1}{2} \left(\log\left(\frac{67 + 10\sqrt{6}}{680\pi}\right) + \frac{1}{2} \left(2 \log\left(\frac{53 + 4\sqrt{150}}{\sqrt{2\pi}850}\right) \right) \right) \\
&= \frac{1}{2} \log\left(\frac{67 + 10\sqrt{6}}{680\pi}\right) + \log\left(\frac{53 + 4\sqrt{150}}{\sqrt{2\pi}850}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
KL(p_{l, s}(x) | p_{\bar{l}, \bar{s}}(x)) &= -\log\left(\frac{160\pi}{\sqrt{10 + \sqrt{80}(14 + 4\sqrt{5})}}\right) - \left(\frac{1}{2} \log\left(\frac{67 + 10\sqrt{6}}{680\pi}\right)\right) \\
&+ \log\left(\frac{53 + 4\sqrt{150}}{\sqrt{2\pi}850}\right) = \frac{\sqrt{85(67 + 10\sqrt{6})(10 + \sqrt{80})(14 + 4\sqrt{5})}}{32\pi(53 + 4\sqrt{150})}.
\end{aligned}$$

5.6.2. Entropy and the KL Divergence of a Linear Convex Combination of two Rational Student Densities

We will consider an example of the linear convex combination of two Rational Student t-Densities with the equal weights, $f_\nu(y) = \frac{1}{2}f_{\nu_1}(y) + \frac{1}{2}f_{\nu_2}(y)$, where $\nu_1 = 3$ and $\nu_2 = 5$.

We have

$$\begin{aligned} B\left(\frac{1}{2}, \frac{\nu_1}{2}\right) &= B\left(\frac{1}{2}, \frac{3}{2}\right) = \frac{\pi}{2}. \\ B\left(\frac{1}{2}, \frac{\nu_2}{2}\right) &= B\left(\frac{1}{2}, \frac{5}{2}\right) = \frac{3\pi}{8}. \\ f_\nu(y) &= \frac{1}{\sqrt{3\pi}\left(1 + \frac{y^2}{3}\right)^2} + \frac{4}{3\sqrt{5}\pi\left(1 + \frac{y^2}{5}\right)^3} = \frac{\sqrt{15}\left(1 + \frac{y^2}{5}\right)^3 + 4\left(1 + \frac{y^2}{3}\right)^2}{3\sqrt{5}\left(1 + \frac{y^2}{3}\right)^2\left(1 + \frac{y^2}{5}\right)^3}. \end{aligned}$$

The entropy is given by

$$\begin{aligned} Ent(f_\nu(y)) &= - \int_{-\infty}^{\infty} f_\nu(y) \log(f_\nu(y)) dy \\ &= -\frac{1}{2} \int_{-\infty}^{\infty} f_{\nu_1}(y) \log(f_\nu(y)) dy - \frac{1}{2} \int_{-\infty}^{\infty} f_{\nu_2}(y) \log(f_\nu(y)) dy. \end{aligned}$$

First,

$$\begin{aligned} \int_{-\infty}^{\infty} f_{\nu_1}(y) \log(f_\nu(y)) dy &= \int_{-\infty}^{\infty} f_{\nu_1}(y) \log\left(\sqrt{15}\left(1 + \frac{y^2}{5}\right)^3 + 4\left(1 + \frac{y^2}{3}\right)^2\right) dy \\ &\quad - \log(3\sqrt{5}\pi) \left(\int_{-\infty}^{\infty} f_{\nu_1}(y) dy\right) - 2 \int_{-\infty}^{\infty} f_{\nu_1}(y) \log\left(1 + \frac{y^2}{3}\right) dy \\ &\quad - 3 \int_{-\infty}^{\infty} f_{\nu_1}(y) \log\left(1 + \frac{y^2}{5}\right) dy. \quad (4) \end{aligned}$$

We have

$$\int_{-\infty}^{\infty} f_{\nu_1}(y) dy = 1.$$

Let $u = \frac{y}{\sqrt{3}}$, so $dy = \sqrt{3}du$, and

$$\begin{aligned} \sqrt{15}\left(1 + \frac{y^2}{5}\right)^3 + 4\left(1 + \frac{y^2}{3}\right)^2 &= \sqrt{15}\left(1 + \frac{3u^2}{5}\right)^3 + 4(1 + u^2)^2 = \\ \frac{27\sqrt{15}}{125}\left[u^6 + \frac{135\sqrt{15} + 500}{27\sqrt{15}}u^4 + \frac{225\sqrt{15} + 1000}{27\sqrt{15}}u^2 + \frac{125\sqrt{15} + 500}{27\sqrt{15}}\right]. \end{aligned}$$

Let

$$\begin{aligned} L_1 &= \int_{-\infty}^{\infty} f_{\nu_1}(y) \log\left(\sqrt{15}\left(1 + \frac{y^2}{5}\right)^3 + 4\left(1 + \frac{y^2}{3}\right)^2\right) dy = \\ \log\left(\frac{27\sqrt{15}}{125}\right) + \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log\left(u^6 + \frac{135\sqrt{15} + 500}{27\sqrt{15}}u^4 + \frac{225\sqrt{15} + 1000}{27\sqrt{15}}u^2 + \frac{125\sqrt{15} + 500}{27\sqrt{15}}\right)}{(u^2 + 1)^2} du. \end{aligned}$$

We apply Theorem 5.2.5 with $\psi(z) = (z - z_1)(z - z_2)(z - z_3)$, which has three zeros $z_1 = 2.75424i$, $z_2 = -0.0894889 + 1.05158i$, $z_3 = 0.0894889 + 1.05158i$ in the upper half plane. We have $\psi(-i) = (-i - z_1)(-i - z_2)(-i - z_3) = 15.8316i$, and $|\psi(-i)| = 15.8316$.

We have

$$\begin{aligned} \psi'(z) &= (z - z_2)(z - z_3) + (z - z_1)(z - z_3) + (z - z_1)(z - z_2), \text{ so } \psi'(-i) = (-i - z_2)(-i - z_3) + \\ &(-i - z_1)(-i - z_3) + (-i - z_1)(-i - z_2) = -19.6212, \frac{\psi'(-i)}{\psi(-i)} = 1.2394i \text{ and } \operatorname{Im}\left(\frac{\psi'(-i)}{\psi(-i)}\right) = 1.2394. \end{aligned}$$

Hence,

$$\begin{aligned} &\int_{-\infty}^{\infty} \frac{\log\left(u^6 + \frac{135\sqrt{15} + 500}{27\sqrt{15}}u^4 + \frac{225\sqrt{15} + 1000}{27\sqrt{15}}u^2 + \frac{125\sqrt{15} + 500}{27\sqrt{15}}\right)}{(u^2 + 1)^2} du \\ &= \pi \log(|\psi(-i)|) - \pi \operatorname{Im}\left(\frac{\psi'(-i)}{\psi(-i)}\right) = \pi \log(15.8316) - \pi(1.2394). \end{aligned}$$

$$L_1 = \log\left(\frac{27\sqrt{15}}{125}\right) + 2(\log(15.8316) - 1.2394).$$

Next, we calculate

$$L_2 = \int_{-\infty}^{\infty} f_{\nu_1}(y) \log\left(1 + \frac{y^2}{3}\right) dy = \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log(1 + u^2)}{1 + u^2} du = \frac{2}{\pi}(\pi \log(4)) = \log(16).$$

$$\begin{aligned} L_3 &= \int_{-\infty}^{\infty} f_{\nu_1}(y) \log\left(1 + \frac{y^2}{5}\right) dy = \int_{-\infty}^{\infty} \frac{\log\left(1 + \frac{3}{5}u^2\right)}{(1 + u^2)^2} du \\ &= \log\left(\frac{3}{5}\right) + \frac{2}{\pi} \left[\pi \log(|\psi(-i)|) - \pi \operatorname{Im}\left(\frac{\psi'(-i)}{\psi(-i)}\right) \right], \end{aligned}$$

using Theorem 5.2.5 with $\psi(z) = \sqrt{\frac{5}{3}} + iz$. We have

$$|\psi(-i)| = \sqrt{\frac{5}{3}} + 1$$

and

$$\operatorname{Im}\left(\frac{\psi'(-i)}{\psi(-i)}\right) = \frac{1}{1 + \sqrt{\frac{5}{3}}},$$

so

$$L_3 = \log\left(\frac{3}{5}\right) + 2\left(\log\left(\sqrt{\frac{5}{3}} + 1\right) - \frac{1}{1 + \sqrt{\frac{5}{3}}}\right).$$

Substituting L_1 , L_2 , and L_3 into (4), we obtain

$$\begin{aligned} \int_{-\infty}^{\infty} f_{\nu_1}(y) \log(f_{\nu}(y)) dy &= \log\left(\frac{27\sqrt{15}}{125}\right) + 2(\log(15.8316) - 1.2394) + \\ &- \log(3\sqrt{5}\pi) - 2\log(16) - 3\left[\log\left(\frac{3}{5}\right) + 2\left(\log\left(\sqrt{\frac{5}{3}} + 1\right) - \frac{1}{1 + \sqrt{\frac{5}{3}}}\right)\right] \\ &= \log\left(\frac{(15.8316)^2}{\sqrt{3}\pi 16^2 (\sqrt{\frac{5}{3}} + 1)^6}\right) - 2.4788 + \frac{3}{1 + \sqrt{\frac{5}{3}}}. \end{aligned}$$

Similarly,

$$\begin{aligned} \int_{-\infty}^{\infty} f_{\nu_2}(y) \log(f_{\nu}(y)) dy &= \int_{-\infty}^{\infty} f_{\nu_2}(y) \log\left(\sqrt{15}\left(1 + \frac{y^2}{5}\right)^3 + \right. \\ &\quad \left. 4\left(1 + \frac{y^2}{3}\right)^2\right) dy - \log(3\sqrt{5}\pi) \left(\int_{-\infty}^{\infty} f_{\nu_2}(y) dy\right) - \\ &2 \int_{-\infty}^{\infty} f_{\nu_2}(y) \log\left(1 + \frac{y^2}{3}\right) dy - 3 \int_{-\infty}^{\infty} f_{\nu_2}(y) \log\left(1 + \frac{y^2}{5}\right) dy. \quad (5) \end{aligned}$$

We have

$$\int_{-\infty}^{\infty} f_{\nu_2}(y) dy = 1.$$

Let $v = \frac{y}{\sqrt{5}}$, so $dy = \sqrt{5}dv$, and

$$\begin{aligned} \sqrt{15}\left(1 + \frac{y^2}{5}\right)^3 + 4\left(1 + \frac{y^2}{3}\right)^2 &= \sqrt{15}(1 + u^2)^3 + 4\left(1 + \frac{5}{3}u^2\right)^2 \\ &= \sqrt{15}\left[u^6 + \frac{27\sqrt{15} + 100}{9\sqrt{15}}u^4 + \frac{9\sqrt{15} + 40}{3\sqrt{15}}u^2 + \frac{\sqrt{15} + 4}{\sqrt{15}}\right]. \end{aligned}$$

$$\begin{aligned} H_1 &= \int_{-\infty}^{\infty} f_{\nu_2}(y) \log\left(\sqrt{15}\left(1 + \frac{y^2}{5}\right)^3 + 4\left(1 + \frac{y^2}{3}\right)^2\right) dy \\ &= \log(\sqrt{15}) + \frac{8}{3\pi} \int_{-\infty}^{\infty} \frac{\log\left(u^6 + \frac{27\sqrt{15}+100}{9\sqrt{15}}u^4 + \frac{9\sqrt{15}+40}{3\sqrt{15}}u^2 + \frac{\sqrt{15}+4}{\sqrt{15}}\right)}{(u^2 + 1)^3} du. \end{aligned}$$

We apply Theorem 5.2.6 with $\bar{\psi}(z) = (z - \bar{z}_1)(z - \bar{z}_2)(z - \bar{z}_3)$, which has three zeros $\bar{z}_1 = 2.13343i$, $\bar{z}_2 = -0.0693178 + 0.81455i$, $\bar{z}_3 = 0.0693178 + 0.81455i$ in the upper half plane. We have

$$\bar{\psi}(-i) = (-i - \bar{z}_1)(-i - \bar{z}_2)(-i - \bar{z}_3) = 10.3322i,$$

and

$$|\psi(-i)| = 10.3322.$$

$$\psi'(z) = (z - \bar{z}_2)(z - \bar{z}_3) + (z - \bar{z}_1)(z - \bar{z}_3) + (z - \bar{z}_1)(z - \bar{z}_2),$$

so

$$\bar{\psi}'(-i) = (-i - \bar{z}_2)(-i - \bar{z}_3) + (-i - \bar{z}_1)(-i - \bar{z}_3) + (-i - \bar{z}_1)(-i - \bar{z}_2) = -14.669,$$

$$\frac{\bar{\psi}'(-i)}{\bar{\psi}(-i)} = 1.4197i$$

and

$$\operatorname{Im}\left(\frac{\bar{\psi}'(-i)}{\bar{\psi}(-i)}\right) = 1.4197.$$

We have $\frac{\bar{\psi}'(-i)}{\bar{\psi}(-i)} = -2.0156$, so $\operatorname{Re}\left(\left[\frac{\bar{\psi}'(-i)}{\bar{\psi}(-i)}\right]^2\right) = -2.0156$. We have

$$\bar{\psi}''(z) = 6z - 2(\bar{z}_1 + \bar{z}_2 + \bar{z}_3),$$

so

$$\bar{\psi}''(-i) = -13.5251i$$

and

$$\operatorname{Re}\left(\frac{\bar{\psi}''(-i)}{\bar{\psi}(-i)}\right) = -1.3090.$$

Hence,

$$\begin{aligned} & \int_{-\infty}^{\infty} \frac{\log(u^6 + \frac{27\sqrt{15}+100}{9\sqrt{15}}u^4 + \frac{9\sqrt{15}+40}{3\sqrt{15}}u^2 + \frac{\sqrt{15}+4}{\sqrt{15}})}{(u^2+1)^3} du \\ &= \frac{\pi}{4} \left\{ 3 \log(|\psi(-i)|) - 3 \operatorname{Im}\left(\frac{\bar{\psi}'(-i)}{\bar{\psi}(-i)}\right) + \operatorname{Re}\left(\left[\frac{\bar{\psi}'(-i)}{\bar{\psi}(-i)}\right]^2\right) - \operatorname{Re}\left(\frac{\bar{\psi}''(-i)}{\bar{\psi}(-i)}\right) \right\} \\ &= \frac{\pi}{4} \{ 3 \log(10.3322) - 3(1.4197) - 2.0156 + 1.3090 \} = \pi \left(\frac{3}{4} \log(10.3322) - 1.241425 \right). \end{aligned}$$

$$H_1 = \log(\sqrt{15}) + \frac{8}{3\pi} \left(\pi \left(\frac{3}{4} \log(10.3322) - 1.241425 \right) \right) = \log(\sqrt{15}) + 2 \log(10.3322) - 3.3105.$$

Next, we calculate

$$H_2 = \int_{-\infty}^{\infty} f_{\nu_2}(y) \log\left(1 + \frac{y^2}{3}\right) dy = \frac{8}{3\pi} \int_{-\infty}^{\infty} \frac{\log\left(1 + \frac{5}{3}v^2\right)}{(1+v^2)^3} dv.$$

We apply Theorem 5.2.6 again with $\bar{\psi}(z) = 1 + \sqrt{\frac{5}{3}}iz$, so $|\bar{\psi}(-i)| = 1 + \sqrt{\frac{5}{3}}$. We have

$$\frac{\bar{\psi}'(z)}{\bar{\psi}(z)} = \frac{1}{z - \sqrt{\frac{3}{5}}i},$$

so

$$\frac{\bar{\psi}'(-i)}{\bar{\psi}(-i)} = \frac{1}{1 + \sqrt{\frac{3}{5}}}i$$

and

$$\operatorname{Im}\left(\frac{\bar{\psi}'(-i)}{\bar{\psi}(-i)}\right) = \frac{1}{1 + \sqrt{\frac{3}{5}}}.$$

In addition,

$$\left(\frac{\bar{\psi}'(-i)}{\bar{\psi}(-i)}\right)^2 = -\frac{1}{\left(1 + \sqrt{\frac{3}{5}}\right)^2},$$

so

$$\operatorname{Re}\left[\left(\frac{\bar{\psi}'(-i)}{\bar{\psi}(-i)}\right)^2\right] = -\frac{1}{\left(1 + \sqrt{\frac{3}{5}}\right)^2}.$$

We have $\bar{\psi}''(-i) = 0$ because $\bar{\psi}''(z) = 0$ for all z . Hence,

$$\int_{-\infty}^{\infty} \frac{\log(1 + \frac{5}{3}v^2)}{(1 + v^2)^3} dv = \frac{\pi}{4} \left(3 \log\left(1 + \sqrt{\frac{5}{3}}\right) - \frac{3}{1 + \sqrt{\frac{3}{5}}} - \frac{1}{\left(1 + \sqrt{\frac{3}{5}}\right)^2} \right).$$

Therefore,

$$H_2 = 2 \log\left(1 + \sqrt{\frac{5}{3}}\right) - \frac{2}{1 + \sqrt{\frac{3}{5}}} - \frac{2}{3\left(1 + \sqrt{\frac{3}{5}}\right)^2}.$$

Finally, we calculate

$$H_3 = \int_{-\infty}^{\infty} f_{\nu_2}(y) \log\left(1 + \frac{y^2}{5}\right) dy = \frac{8}{3\pi} \int_{-\infty}^{\infty} \frac{\log(1 + v^2)}{(1 + v^2)^3} dv$$

Applying Theorem 5.2.6 with $\bar{\psi}(z) = 1 + iz$, we obtain

$$\int_{-\infty}^{\infty} \frac{\log(1 + v^2)}{(1 + v^2)^3} dv = \pi \left(\frac{3}{4} \log(2) - \frac{7}{16} \right).$$

Hence,

$$H_3 = \frac{8}{3\pi} \left(\pi \left(\frac{3}{4} \log(2) - \frac{7}{16} \right) \right) = \log(4) - \frac{7}{2}.$$

Substituting H_1 , H_2 , and H_3 into (5), we have

$$\begin{aligned} \int_{-\infty}^{\infty} f_{\nu_2}(y) \log(f_{\nu}(y)) dy &= [\log(\sqrt{15}) + 2 \log(10.3322) - 3.3105] + \\ &- \log(3\sqrt{5}\pi) - 2[2 \log(1 + \sqrt{\frac{5}{3}}) - \frac{2}{1 + \sqrt{\frac{3}{5}}} - \frac{2}{3(1 + \sqrt{\frac{3}{5}})^2}] - 3[\log(4) - \frac{7}{2}] \\ &= \log\left(\frac{(10.3322)^2}{\sqrt{3}\pi 64(1 + \sqrt{\frac{5}{3}})^4}\right) + \frac{4}{1 + \sqrt{\frac{3}{5}}} + \frac{4}{3(1 + \sqrt{\frac{3}{5}})^2} + 7.1895. \end{aligned}$$

Therefore,

$$Ent(f_{\nu}(y)) = -\log\left(\frac{163.575}{128\sqrt{3}\pi(1 + \sqrt{\frac{5}{3}})^5}\right) - \frac{3}{2(1 + \sqrt{\frac{5}{3}})} - \frac{2}{1 + \sqrt{\frac{3}{5}}} - \frac{2}{3(1 + \sqrt{\frac{3}{5}})^2} - 2.35535.$$

Consider $f_{\bar{\nu}}(y) = \frac{1}{2}f_{\bar{\nu}_1}(y) + \frac{1}{2}f_{\bar{\nu}_2}(y)$, where $\bar{\nu}_1 = 7$ and $\bar{\nu}_2 = 3$. Hence

$$f_{\bar{\nu}}(y) = \frac{8}{5\sqrt{7}\pi(1 + \frac{y^2}{7})^4} + \frac{1}{\sqrt{3}\pi(1 + \frac{y^2}{3})^2} = \frac{8\sqrt{3}(1 + \frac{y^2}{3})^2 + 5\sqrt{7}(1 + \frac{y^2}{7})^4}{5\sqrt{21}\pi(1 + \frac{y^2}{7})^4(1 + \frac{y^2}{3})^2}.$$

The KL divergence of $f_{\nu}(y)$ and $f_{\bar{\nu}}(y)$ is given by

$$KL(f_{\nu}(y)|f_{\bar{\nu}}(y)) = \int_{-\infty}^{\infty} f_{\nu}(y) \log\left[\frac{f_{\nu}(y)}{f_{\bar{\nu}}(y)}\right] dy =$$

$$\int_{-\infty}^{\infty} f_{\nu}(y) \log(f_{\nu}(y)) dy - \int_{-\infty}^{\infty} f_{\nu}(y) \log(f_{\bar{\nu}}(y)) dy = -Ent(f_{\nu}(y)) - \int_{-\infty}^{\infty} f_{\nu}(y) \log(f_{\bar{\nu}}(y)) dy.$$

We have

$$\int_{-\infty}^{\infty} f_{\nu}(y) \log(f_{\bar{\nu}}(y)) dy = \frac{1}{2} \int_{-\infty}^{\infty} f_{\nu_1}(y) \log(f_{\bar{\nu}}(y)) dy + \frac{1}{2} \int_{-\infty}^{\infty} f_{\nu_2}(y) \log(f_{\bar{\nu}}(y)) dy. (6)$$

First, we calculate

$$N_1 = \int_{-\infty}^{\infty} f_{\nu_1}(y) \log(f_{\bar{\nu}}(y)) dy = \int_{-\infty}^{\infty} f_{\nu_1}(y) \log(8\sqrt{3}(1 + \frac{y^2}{3})^2 + 5\sqrt{7}(1 + \frac{y^2}{7})^4) dy -$$

$$\log(5\sqrt{21}\pi) \left(\int_{-\infty}^{\infty} f_{\nu_1}(y) dy \right) - 4 \int_{-\infty}^{\infty} f_{\nu_1}(y) \log(1 + \frac{y^2}{7}) dy - 2 \int_{-\infty}^{\infty} f_{\nu_1}(y) \log(1 + \frac{y^2}{3}) dy.$$

Let $u = \frac{y}{\sqrt{3}}$, so $dy = \sqrt{3} du$ and

$$8\sqrt{3}(1 + \frac{y^2}{3})^2 + 5\sqrt{7}(1 + \frac{y^2}{7})^4 = 8\sqrt{3}(1 + u^2)^2 + 5\sqrt{7}(1 + \frac{3u^2}{7})^4$$

$$= \frac{405\sqrt{7}}{2401} [u^8 + \frac{28}{3}u^6 + \frac{49(270\sqrt{7} + 392\sqrt{3})}{405\sqrt{7}}u^4 + \frac{343(60\sqrt{7} + 112\sqrt{3})}{405\sqrt{7}}u^2 + \frac{2401(5\sqrt{7} + 8\sqrt{3})}{405\sqrt{7}}].$$

$$\int_{-\infty}^{\infty} f_{\nu_1}(y) \log(8\sqrt{3}(1 + \frac{y^2}{3})^2 + 5\sqrt{7}(1 + \frac{y^2}{7})^4) dy = \log(\frac{405\sqrt{7}}{2401}) +$$

$$\frac{2}{\pi} (\pi [\log(|\psi(-i)|)] - \text{Im}\{\frac{\psi'(-i)}{\psi(-i)}\}),$$

using Theorem 5.2.5 with $\psi(z) = (z - z_1)(z - z_2)(z - z_3)(z - z_4)$, where

$z_1 = -1.2790 + 2.2783i$, $z_2 = 1.2790 + 2.2783i$, $z_3 = -0.1205 + 1.0615i$ and $z_4 = 0.1205 + 1.0615i$.

We have $\psi(-i) = (12.3831)(4.2643) = 52.8053$, so $|\psi(-i)| = 52.8053$; $\frac{\psi'(z)}{\psi(z)} = \frac{1}{z-z_1} + \frac{1}{z-z_2} + \frac{1}{z-z_3} + \frac{1}{z-z_4}$ so $\frac{\psi'(-i)}{\psi(-i)} = 1.4963i$, and $\text{Im}\{\frac{\psi'(-i)}{\psi(-i)}\} = 1.4963$. Hence,

$$\int_{-\infty}^{\infty} f_{\nu_1}(y) \log(8\sqrt{3}(1 + \frac{y^2}{3})^2 + 5\sqrt{7}(1 + \frac{y^2}{7})^4) dy = \log(\frac{405\sqrt{7}}{2401}) + 2(\log(52.8053) - 1.4963).$$

Next, we calculate

$$\int_{-\infty}^{\infty} f_{\nu_1}(y) \log(1 + \frac{y^2}{7}) dy = \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log(1 + \frac{3}{7}u^2)}{(1 + u^2)^2} du$$

$$= 2 \log(1 + \sqrt{\frac{3}{7}}) - \frac{2}{1 + \sqrt{\frac{7}{3}}},$$

using Theorem 5.2.5 with $\psi(z) = 1 + \sqrt{\frac{3}{7}}iz$. Finally,

$$\int_{-\infty}^{\infty} f_{\nu_1}(y) \log\left(1 + \frac{y^2}{3}\right) dy = \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\log(1+u^2)}{1+u^2} du = \frac{2}{\pi} \pi \log(4) = \log(16).$$

Hence

$$\begin{aligned} N_1 &= \log\left(\frac{405\sqrt{7}}{2401}\right) + 2(\log(52.8053) - 1.4963) - \log(5\sqrt{21}\pi) \\ &- 8\log\left(1 + \sqrt{\frac{3}{7}}\right) + \frac{8}{1 + \sqrt{\frac{7}{3}}} - \log(256) = \log\left(\frac{0.3675}{\sqrt{3}\pi(1 + \sqrt{\frac{3}{7}})^8}\right) + \frac{8}{1 + \sqrt{\frac{7}{3}}} - 2.9926. \end{aligned}$$

Similarly, we calculate

$$\begin{aligned} N_2 &= \int_{-\infty}^{\infty} f_{\nu_2}(y) \log(f_{\nu}(y)) dy = \int_{-\infty}^{\infty} f_{\nu_2}(y) \log\left(8\sqrt{3}\left(1 + \frac{y^2}{3}\right)^2 + 5\sqrt{7}\left(1 + \frac{y^2}{7}\right)^4\right) dy \\ &- \log(5\sqrt{21}\pi) \left(\int_{-\infty}^{\infty} f_{\nu_2}(y) dy\right) - 4 \int_{-\infty}^{\infty} f_{\nu_2}(y) \log\left(1 + \frac{y^2}{7}\right) dy - 2 \int_{-\infty}^{\infty} f_{\nu_2}(y) \log\left(1 + \frac{y^2}{3}\right) dy. \end{aligned}$$

Let $v = \frac{y}{\sqrt{5}}$, so $dy = \sqrt{5}du$ and

$$\begin{aligned} 8\sqrt{3}\left(1 + \frac{y^2}{3}\right)^2 + 5\sqrt{7}\left(1 + \frac{y^2}{7}\right)^4 &= 5\sqrt{7}\left(1 + \frac{5}{7}v^2\right)^4 + 8\sqrt{3}\left(1 + \frac{5v^2}{3}\right)^2 = \\ \frac{3125\sqrt{7}}{2401} [v^8 + \frac{28}{5}v^6 + \frac{(13230\sqrt{7} + 19208\sqrt{3})}{1125\sqrt{7}}v^4 + \frac{102900\sqrt{7} + 192080\sqrt{3}}{9375\sqrt{7}}v^2 + \\ &\frac{2401(5\sqrt{7} + 8\sqrt{3})}{3125\sqrt{7}}]. \end{aligned}$$

$$\begin{aligned} \int_{-\infty}^{\infty} f_{\nu_2}(y) \log\left(8\sqrt{3}\left(1 + \frac{y^2}{3}\right)^2 + 5\sqrt{7}\left(1 + \frac{y^2}{7}\right)^4\right) dy &= \log\left(\frac{3125\sqrt{7}}{2401}\right) + \\ \frac{8}{3\pi} \left(\frac{\pi}{4} [3\log(|\psi(-i)|) - 3\operatorname{Im}\left\{\frac{\psi'(-i)}{\psi(-i)}\right\} + \operatorname{Re}\left\{\left(\frac{\psi'(-i)}{\psi(-i)}\right)^2\right\} - \operatorname{Re}\left\{\frac{\psi'(-i)}{\psi(-i)}\right\}]\right), \end{aligned}$$

using Theorem 5.2.6 with $\psi(z) = (z-z_1)(z-z_2)(z-z_3)(z-z_4)$, where $z_1 = -0.9907 + 1.7647i$, $z_2 = 0.9907 + 1.7647i$, $z_3 = -0.0934 + 0.8222i$ and $z_4 = 0.0934 + 0.8222i$. We have

$\psi(-i) = (8.6251)(3.3291) = 28.7141$, so $|\psi(-i)| = 28.7141$;

$$\frac{\psi'(z)}{\psi(z)} = \frac{1}{z - z_1} + \frac{1}{z - z_2} + \frac{1}{z - z_3} + \frac{1}{z - z_4}$$

so

$$\frac{\psi'(-i)}{\psi(-i)} = 1.7358i,$$

$$\operatorname{Im}\left\{\frac{\psi'(-i)}{\psi(-i)}\right\} = 1.7358$$

and

$$\operatorname{Re}\left\{\left(\frac{\psi'(-i)}{\psi(-i)}\right)^2\right\} = -3.0130.$$

We have $\psi''(z) = 2(6z^2 + z_2z_3 + z_2z_4 + z_3z_4 + z_1(z_2 + z_3 + z_4) - 3z(z_1 + z_2 + z_3 + z_4))$, so $\psi''(-i) = -64.211$ and $\operatorname{Re}\left\{\frac{\psi'(-i)}{\psi(-i)}\right\} = -2.2362$.

Hence,

$$\begin{aligned} \int_{-\infty}^{\infty} f_{\nu_2}(y) \log(8\sqrt{3}(1 + \frac{y^2}{3})^2 + 5\sqrt{7}(1 + \frac{y^2}{4})^4) dy &= \log\left(\frac{3125\sqrt{7}}{2401}\right) + 2(\log(28.7141) - 3.9895) \\ &= \log(1073.12\sqrt{7}) - 3.9895. \end{aligned}$$

Next, we calculate

$$\begin{aligned} \int_{-\infty}^{\infty} f_{\nu_2}(y) \log\left(1 + \frac{y^2}{7}\right) dy &= \frac{8}{3\pi} \int_{-\infty}^{\infty} \frac{\log(1 + \frac{5}{7}v^2)}{(1 + v^2)^3} dv \\ &= 2\log\left(1 + \sqrt{\frac{5}{7}}\right) - \frac{2}{1 + \sqrt{\frac{7}{5}}} - \frac{2}{3(1 + \sqrt{\frac{7}{5}})^2}, \end{aligned}$$

using Theorem 5.2.6 with $\psi(z) = 1 + \sqrt{\frac{5}{7}}iz$. Finally,

$$\int_{-\infty}^{\infty} f_{\nu_2}(y) \log\left(1 + \frac{y^2}{3}\right) dy = \frac{8}{3\pi} \int_{-\infty}^{\infty} \frac{\log(1 + \frac{5}{3}v^2)}{(1 + v^2)^3} dv$$

5.7 On Approximate Entropy Calculation for Finite Gaussian Mixtures 183

$$= 2 \log(1 + \sqrt{\frac{5}{3}}) - \frac{2}{1 + \sqrt{\frac{3}{5}}} - \frac{2}{3(1 + \sqrt{\frac{3}{5}})^2},$$

using Theorem 5.2.6 with $\psi(z) = 1 + \sqrt{\frac{5}{3}}iz$. Hence

$$\begin{aligned} N_2 &= \log\left(\frac{1073.12}{5\sqrt{3}\pi(1 + \sqrt{\frac{5}{7}})^8(1 + \sqrt{\frac{5}{3}})^4}\right) - 3.9895 \\ &+ \frac{8}{1 + \sqrt{\frac{7}{5}}} + \frac{8}{3(1 + \sqrt{\frac{7}{5}})^2} + \frac{4}{1 + \sqrt{\frac{3}{5}}} + \frac{4}{3(1 + \sqrt{\frac{3}{5}})^2}. \end{aligned}$$

Combining N_1 and N_2 , we obtain

$$\begin{aligned} f_\nu(y) \log(f_\nu(y)) &= \frac{1}{2}N_1 + \frac{1}{2}N_2 = \log\left(\frac{19.8588}{\sqrt{15}\pi(1 + \sqrt{\frac{3}{7}})^4(1 + \sqrt{\frac{5}{7}})^4(1 + \sqrt{\frac{5}{3}})^2}\right) - 3.49105 \\ &+ \frac{4}{1 + \sqrt{\frac{7}{3}}} + \frac{4}{1 + \sqrt{\frac{7}{5}}} + \frac{4}{3(1 + \sqrt{\frac{7}{5}})^2} + \frac{2}{1 + \sqrt{\frac{3}{5}}} + \frac{2}{3(1 + \sqrt{\frac{3}{5}})^2}. \end{aligned}$$

Therefore,

$$\begin{aligned} KL(f_\nu(y)|f_\nu(y)) &= \log\left(\frac{0.0644\sqrt{5}(1 + \sqrt{\frac{3}{7}})^4(1 + \sqrt{\frac{5}{7}})^4}{1 + \sqrt{\frac{5}{3}}}\right) + 5.8464 \\ &+ \frac{3}{2(1 + \sqrt{\frac{5}{3}})} - \frac{4}{1 + \sqrt{\frac{7}{3}}} - \frac{4}{1 + \sqrt{\frac{7}{5}}} - \frac{2}{3(\frac{6}{5} + \sqrt{\frac{7}{5}})}. \end{aligned}$$

5.7. On Approximate Entropy Calculation for Finite Gaussian Mixtures

We consider the linear convex combination $f_\nu(t) = \sum_{i=1}^n \alpha_i f_{\nu_i}(t)$, where $f_{\nu_i}(t)$ converges to $f_i(t)$ when $\nu_i \rightarrow \infty$, where $f_i(t)$ denotes the PDF of $\mathcal{N}(\mu_i, \frac{1}{\sigma_i})$. Applying Theorem 5.2.2, we can obtain entropy of $f_\nu(t)$, which converges to entropy of the finite Gaussian mixture.

It is illustrated by the following example. We consider

$$f_{\nu_1}(t) = \frac{\Gamma(\frac{\nu_1+1}{2})}{\Gamma(\frac{\nu_1}{2})} \left(\frac{\sigma_1}{\nu_1\pi}\right)^{(1/2)} \left(1 + \frac{\sigma_1}{\nu_1}(t - \mu_1)^2\right)^{-\frac{\nu_1+1}{2}}.$$

$$f_{\nu_2}(t) = \frac{\Gamma(\frac{\nu_2+1}{2})}{\Gamma(\frac{\nu_2}{2})} \left(\frac{\sigma_2}{\nu_2\pi}\right)^{(1/2)} \left(1 + \frac{\sigma_2}{\nu_2}(t - \mu_2)^2\right)^{-\frac{\nu_2+1}{2}}.$$

We have $f_{\nu_1}(t) \rightarrow f_1(t)$ when $\nu_1 \rightarrow \infty$, where $f_1(t)$ is the PDF of $\mathcal{N}(\mu_1, \frac{1}{\sigma_1})$. And $f_{\nu_2}(t) \rightarrow f_2(t)$ when $\nu_2 \rightarrow \infty$, where $f_2(t)$ is the PDF of $\mathcal{N}(\mu_2, \frac{1}{\sigma_2})$. When $\nu_1 \rightarrow \infty, \nu_2 \rightarrow \infty$, entropy of the combination $\alpha_1 f_{\nu_1}(t) + \alpha_2 f_{\nu_2}(t)$, $\alpha_1 + \alpha_2 = 1$, converges to entropy of the finite Gaussian mixture $\alpha_1 f_1(t) + \alpha_2 f_2(t)$. If we take $\nu_1 = \nu_2 = \nu$, $\mu_1 = 0$, $\sigma_1 = 2$, $\mu_2 = 1$, $\sigma_2 = 1$, entropy of the finite Gaussian mixture is 1.4048. The following table shows the convergence when increasing the degree of freedom (Df).

Table 5.1: Convergence of entropy of linear combination of student distributions to entropy of finite Gaussian mixture when increasing the degree of freedom.

Df	Entropy
31	1.4310
33	1.4294
99	1.4129
119	1.4115
399	1.4068
999	1.4056
1399	1.4048

Chapter 6

Conclusion

6.1. Concluding Remarks

This research introduces a new method, the MC-Tree method, for pricing options with the requirement of high accuracy and efficiency, both of which are important in applications. It employs a mixing distribution on the tree parameters to give the prescribed mean and variance. Given the family of mixing densities, we obtain the corresponding compound densities of the tree outcomes at the final time step. We look for mixing densities for which the corresponding compound density has high entropy level. The compound density (after a logarithmic transformation of the asset prices) is not the ideal Gaussian distribution but has entropy values close to the maximum possible Gaussian entropy. This thesis introduces two techniques to correct for the deviation from the ideal Gaussian pricing measure. The distribution correction technique ensures that expectations calculated with the method are taken with respect to the desired Gaussian measure. The bias-correction technique ensures that the probability distributions used are risk-neutral in each of the trees.

We write codes to implement all algorithms and the theory of the MC-Tree method in

this thesis. We illustrate that the MC-Tree method is straightforward to implement in practice, demonstrated by numerical results in this thesis. Besides option pricing, we apply correction techniques to develop an algorithm for calculations of the credit valuation adjustment (CVA) on an American option. Numerical examples of the workings of the MC-Tree approach are provided, which show good performance in terms of accuracy and computational speed.

The theory of MC-Tree method in the thesis can be an efficient method to price options on single or multiple assets. The model can be applied to practical development in financial industry due to its high accuracy.

The method shows how close our results are to the "true" prices and shows the tiny confidence interval containing the "true" one with a given (high) probability at 95% when limiting the tree depth.

The recombining multinomial tree serves as the generalization, so keep most aspects and properties of recombining binomial tree, except the difference in the number of underlying assets. The benefit of the recombining binomial/multinomial tree is the reduction of the number of nodes in the tree, so leads to reduce computation time. The completeness of the model allows to provide hedging strategies.

6.2. Further Research

Pricing options and CVA calculations using MC-Tree are still in their infancy, and there is potential research in the future, based on work completed to date.

Apart from the high accuracy requirement, a fast computation speed is crucial in practical applications at the trading desk. Hence, it is necessary to improve the efficiency of the pricing method. We plan to experiment with using a special chip (a field-programmable gate array (FPGA)) to do fast tree computations. The FPGA implementation will likely be

much faster than a CPU implementation. The computation time of the MC-Tree method might be faster in higher-dimensional cases with the FPGA implementation. Mahony et al. [36] developed the novel FPGA architecture on recombining multinomial trees based on Pascal's simplex and showed the fast computation speed. The novelty in option pricing acceleration of the paper is based on Pascal's simplex. Except for recomputing the input parameters to the FPGA, we will use this exact FPGA architecture on the MC-Tree method. Applying FPGA hardware to the MC-Tree methods should allow to show that the MC-Tree methods are not only of theoretical interest but can be made sufficiently fast to become interesting for practitioners in the financial industry. This could lead to interesting new computational approaches to option pricing, but also, as yet unexplored, speed-ups in expectation calculations for higher dimensional Gaussians and Gaussian mixtures. Here contact could be made with developments in Machine Learning or data science in which Gaussian mixtures play an important role. But the same can be said for applications in Econometrics and other applications of mathematical statistics.

Our contribution is to introduce a new a computational approach for fast and accurate pricing, but we still use constant model parameters. The future research direction is to generalize the model to a real market with stochastic parameters.

Another promising research direction is to develop MC-Tree in the financial industry, especially pricing and risk management industry. We can produce a software with the Graphical user interface (GUI) or an App from our codes, which allow any users to change the model inputs to obtain the price.

Besides, the usage of the correction factor brings very high accuracy in European option pricing using MC-Tree method. We would exploit how to use the correction factor in pricing the American option.

The analytic compound density is used in option pricing with the distribution correction technique. A natural question is how to apply the distribution correction factor to pricing

multi-asset options.

Further research is required to get an analytical expression for the compound density in case of three or more assets. The conditioning method applied to the two-asset case can be generalized to higher dimensions but it remains to be seen whether again, using Euler substitutions or otherwise, rational density contributions are obtained for which we have analytical technique, to compute the necessary integrals.

Alternatively, the compound density on the radius can be approximately computed using a MC technique. It remains to be seen how this can be implemented to obtain a workable distribution correction factor.

Appendix A

Choices of Mixing Density

A.1. Choice of Mixing Density in One Asset Case

We claim that $cf(p_1, p_2)dp_1$ will describe the same density as $cf(p_1, p_2)dp_2$.

Proof.

Given $p_1 = \cos^2(\theta)$, we have $dp_1 = -2\cos(\theta)\sin(\theta)d\theta$.

Given $p_2 = \sin^2(\theta)$, we have $dp_2 = 2\cos(\theta)\sin(\theta)d\theta$.

Hence, $|dp_1| = |dp_2|$. By the construction of our tree, we have the symmetry w.r.t permutations of p_1 and p_2 . Hence, $f(p_1, p_2)dp_1$ will describe the same density as $f(p_1, p_2)dp_2$.

A.2. Choice of Mixing Density in Two Asset Case

We claim that the choice of mixing density $cf(p_1, p_2, p_3)dp_1dp_2$, $cf(p_1, p_2, p_3)dp_1dp_3$, and $cf(p_1, p_2, p_3)dp_2dp_3$ produces the same density.

Proof.

We define a density on the triangle formed by all convex combinations of $e_1, e_2, e_3 \in \mathbb{R}^3$.

The set is also described by

$$\{(p_1, p_2, p_3) : p_1 + p_2 + p_3 = 1, p_1 \geq 0, p_2 \geq 0, p_3 \geq 0.\}$$

We can describe each point by two of the three coordinates as the third is one minus the sum of others. We will find the size of the volume element when taking p_1 and p_2 as variables. The area of the triangle itself is obtained as follows:

The sides of triangle has equal length

$$\|e_1 - e_2\| = \|e_2 - e_3\| = \|e_1 - e_3\| = \sqrt{2}.$$

If we start from a point (p_1, p_2, p_3) and change p_1 to p'_1 , p_2 to p'_2 , and p_3 to $p'_3 = p_3 - \Delta p_1 - \Delta p_2$, where $\Delta p_1 = p'_1 - p_1$, $\Delta p_2 = p'_2 - p_2$. So a change of Δp_1 from p_1 to $p_1 + \Delta p_1$ in this case corresponds to $\Delta p_1 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ and a change of Δp_2 from p_2 to $p_2 + \Delta p_2$ corresponds to $\Delta p_2 \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$. Hence, we consider the triangle with vertices $p = (p_1, p_2, p_3), p + \Delta p_1(e_1 - e_3) = (p_1 + \Delta p_1, p_2, p_3 - \Delta p_1), p + \Delta p_2(e_2 - e_3) = (p_1, p_2 + \Delta p_2, p_3 - \Delta p_2)$. The area element is twice the area of triangle with these vertices. We can translate to the origin $(0,0,0)$ without changing the area of triangle.

The angle θ between $e_1 - e_3$ and $e_2 - e_3$ is $\frac{\pi}{3}$ because

$$\cos(\theta) = \frac{\langle e_1 - e_3, e_2 - e_3 \rangle}{\|e_1 - e_3\| \|e_2 - e_3\|} = \frac{1}{2}.$$

The area of the triangle is

$$\frac{1}{2} \sqrt{2} \sqrt{2} \sin\left(\frac{\pi}{3}\right) = \frac{1}{2} \sqrt{3}.$$

We get the area of an equilateral parallelogram is $\sqrt{3}$. The density on the triangle formed by all convex combinations of $e_i, i = 1, 2, 3$ is $cf(p_1, p_2, p_3)\sqrt{3}dp_1dp_2$. Taking $f = 1$, solve for c

$$\int_0^1 \int_0^{1-p_2} c\sqrt{3}dp_1dp_2 = c\sqrt{3} \int_0^1 (1-p_2)dp_2 = c\frac{1}{2}\sqrt{3}.$$

We have

$$\int_0^1 \int_0^{1-p_2} c\sqrt{3}dp_1dp_2 = 1 \Leftrightarrow c = \frac{2}{3}\sqrt{3}.$$

Hence, $c\sqrt{3} = 2$ because the area of the triangle is half the area of the equilateral parallelogram which is used in computation of the area element of dp_1dp_2 .

If we use $cf(p_1, p_2, p_3)\sqrt{3}dp_2dp_3$, this will describe the same density on the triangle by the

symmetry w.r.t permutation of p_1 , p_2 , and p_3 .

This is similar to the choice of the mixing density $cf(p_1, p_2, p_3)\sqrt{3}dp_1dp_3$.

References

- [1] Kaushik Amin and Ajay Khanna. Convergence of American option values from discrete-to continuous-time financial models 1. *Mathematical Finance*, 4(4):289–304, 1994.
- [2] BCBS. International convergence of capital measurement and capital standards: A revised framework - comprehensive version. Technical report, Bank for International Settlements. URL: <https://www.bis.org/publ/bcbs128.htm>, 2006.
- [3] BCBS. A global regulatory framework for more resilient banks and banking systems. Technical report, Bank for International Settlements. URL: <https://www.bis.org/publ/bcbs189.htm>, 2011.
- [4] Dario Andrea Bini, Giuseppe Fiorentino, Luca Gemignani, and Beatrice Meini. Effective fast algorithms for polynomial spectral factorization. *Numerical Algorithms*, 34(2):217–227, 2003.
- [5] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.
- [6] Phelim P Boyle. Options: A Monte Carlo approach. *Journal of Financial Economics*, 4(3):323–338, 1977.

-
- [7] Phelim P Boyle. A lattice framework for option pricing with two state variables. *Journal of Financial and Quantitative Analysis*, pages 1–12, 1988.
- [8] Phelim P Boyle, Jeremy Evnine, and Stephen Gibbs. Numerical evaluation of multivariate contingent claims. *The Review of Financial Studies*, 2(2):241–250, 1989.
- [9] Paolo Brandimarte. *Numerical methods in finance and economics: a MATLAB-based introduction*. John Wiley & Sons, 2013.
- [10] Paolo Brandimarte. *Handbook in Monte Carlo simulation: applications in financial engineering, risk management, and economics*. John Wiley & Sons, 2014.
- [11] Damiano Brigo, Agostino Capponi, and Andrea Pallavicini. Arbitrage-free bilateral counterparty risk valuation under collateralization and application to credit default swaps. *Mathematical Finance*, 24(1):125–146, 2014.
- [12] Damiano Brigo and Fabio Mercurio. *Interest rate models-theory and practice: with smile, inflation and credit*. Springer Science & Business Media, 2007.
- [13] Mark Broadie and Paul Glasserman. Pricing American-style securities using simulation. *Journal of Economic Dynamics and Control*, 21(8-9):1323–1352, 1997.
- [14] Mark Broadie, Paul Glasserman, et al. A stochastic mesh method for pricing high-dimensional American options. *Journal of Computational Finance*, 7:35–72, 2004.
- [15] George Casella and Roger L Berger. *Statistical inference*. Cengage Learning, 2021.
- [16] Giovanni Cesari, John Aquilina, Niels Charpillon, Zlatko Filipovic, Gordon Lee, and Ion Manda. *Modelling, pricing, and hedging counterparty credit exposure: A technical guide*. Springer Science & Business Media, 2009.
- [17] Frédéric Chyzak and Frank Nielsen. A closed-form formula for the Kullback-Leibler divergence between Cauchy distributions. *arXiv preprint arXiv:1905.10965*, 2019.

-
- [18] John C Cox, Jonathan E Ingersoll Jr, and Stephen A Ross. An intertemporal general equilibrium model of asset prices. *Econometrica*, pages 363–384, 1985.
- [19] John C Cox, Stephen A Ross, and Mark Rubinstein. Option pricing: A simplified approach. *Journal of Financial Economics*, 7(3):229–263, 1979.
- [20] Darrell Duffie, Jun Pan, and Kenneth Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376, 2000.
- [21] Darrell Duffie and Kenneth J Singleton. Modeling term structures of defaultable bonds. *The Review of Financial Studies*, 12(4):687–720, 1999.
- [22] Dan Franzén and Otto Sjöholm. Credit valuation adjustment: In theory and practice, 2014. URL: <https://www.diva-portal.org/smash/get/diva2:692743/FULLTEXT01.pdf>.
- [23] Tryphon T Georgiou and Anders Lindquist. Kullback-Leibler approximation of spectral density functions. *IEEE Transactions on Information Theory*, 49(11):2910–2917, 2003.
- [24] Tim NT Goodman, Charles A Micchelli, Giuseppe Rodriguez, and Sebastiano Seatzu. Spectral factorization of Laurent polynomials. *Advances in Computational Mathematics*, 7(4):429–454, 1997.
- [25] Bernard Hanzon and Raimund J Ober. A state-space calculus for rational probability density functions and applications to non-Gaussian filtering. *SIAM Journal on Control and Optimization*, 40(3):724–740, 2001.
- [26] Bernard Hanzon and Wolfgang Scherrer. Filtering and estimation in stochastic volatility models with rationally distributed disturbances. *arXiv preprint arXiv:0706.3335*, 2007.

-
- [27] John Hull and Alan White. CVA and wrong-way risk. *Financial Analysts Journal*, 68(5):58–69, 2012.
- [28] Robert A Jarrow. *Continuous-time asset pricing theory*. Springer, 2018.
- [29] Robert A Jarrow and Andrew Rudd. *Option Pricing*. Homewood, Illinois: Richard D. Irwin, 1983.
- [30] Lambert Pierrat Samuel Kotz. A simple analytic approximation for entropy of Student-t distribution and its relation with normal distribution. *International Journal of Performability Engineering*, 11(5):509, 2015.
- [31] Dietmar PJ Leisen. Pricing the American put option: A detailed convergence analysis for binomial models. *Journal of Economic Dynamics and Control*, 22(8-9):1419–1444, 1998.
- [32] Dietmar PJ Leisen and Matthias Reimer. Binomial models for option valuation—examining and improving convergence. *Applied Mathematical Finance*, 3(4):319–346, 1996.
- [33] Qian Liu. Calculation of credit valuation adjustment based on least square Monte Carlo methods. *Mathematical Problems in Engineering*, 2015, 2015.
- [34] Gane Samb Lo, Modou Ngom, and Tchilabalo Atozou Kpanzou. *Weak Convergence (IA). Sequences of random vectors*. *arXiv preprint arXiv:1610.05415*, 2016.
- [35] Francis A Longstaff and Eduardo S Schwartz. Valuing American options by simulation: a simple least-squares approach. *The Review of Financial Studies*, 14(1):113–147, 2001.
- [36] Aidan O Mahony, Gil Zeidan, Bernard Hanzon, and Emanuel Popovici. A parallel and pipelined implementation of a Pascal-simplex based multi-asset option pricer

- on FPGA using OpenCL. *Microprocessors and Microsystems*, 2022. doi:10.1016/j.micpro.2022.104508.
- [37] Frank Nielsen and Richard Nock. Entropies and cross-entropies of exponential families. In *2010 IEEE International Conference on Image Processing*, pages 3621–3624. IEEE, 2010.
- [38] F. E. N. G. Qian. *Advanced estimation of credit valuation adjustment*. PhD thesis, Technische Universiteit Delft, April 2017.
- [39] Y. Shen. *Credit value adjustment for multi-asset options*. PhD thesis, Technische Universiteit Delft, 2014.
- [40] Dirk Sierag and Bernard Hanzon. Pricing derivatives on multiple assets: recombining multinomial trees based on Pascal’s simplex. *Annals of Operations Research*, 266(1-2):101–127, 2018.
- [41] Daniel E Steffy. Exact solutions to linear systems of equations using output sensitive lifting. *ACM Communications in Computer Algebra*, 44(3/4):160–182, 2011.
- [42] René M Stulz. Options on the minimum or the maximum of two risky assets: analysis and applications. *Journal of Financial Economics*, 10(2):161–185, 1982.
- [43] Yisong Tian. A modified lattice approach to option pricing. *The Journal of Futures Markets (1986-1998)*, 13(5):563, 1993.
- [44] JA Tilley. Valuing American options in a path simulation model. *Insurance Mathematics and Economics*, 2(16):169, 1995.
- [45] Steven H Zhu and Michael Pykhtin. A guide to modeling counterparty credit risk. *GARP Risk Review*, July/August, 2007.