UNIVERSITY OF CALIFORNIA,
IRVINE

Spectral Theory of Sample Covariance Matrices from Discretized Lévy Processes

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Mathematics

by

Gregory Zitelli

Dissertation Committee:
Professor Patrick Guidotti, Chair
Professor Knut Solna
Professor Michael Cranston

2019

“An overwhelming abundance of connections, associations . . .
How many sentences can one create out of the twenty-four letters of the alphabet?
How many meanings can one glean from hundreds of weeds, clods of dirt, and other trifles?”
– Witold Gombrowicz
Cosmos

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Algorithms

ACKNOWLEDGMENTS

I would like to thank Dr. Patrick Guidotti and Dr. Knut Solna for their continued support and guidance. I will never forget their kindness.

This work was partially supported by AFOSR grant FA9550-18-1-0217 and NSF grant 1616954.

CURRICULUM VITAE

Gregory Zitelli

EDUCATION

Doctor of Philosophy in Mathematics **2019**
University of California, Irvine *Irvine, CA*

Master of Science in Mathematics **2013**
University of Tennessee, Knoxville *Knoxville, TN*

Bachelor of Science in Mathematics **2011**
University of California, Irvine *Irvine, CA*

ABSTRACT OF THE DISSERTATION

Spectral Theory of Sample Covariance Matrices from Discretized Lévy Processes

By

Gregory Zitelli

Doctor of Philosophy in Mathematics

University of California, Irvine, 2019

Professor Patrick Guidotti, Chair

Asymptotic spectral techniques have become a powerful tool in estimating statistical properties of systems that can be well approximated by rectangular matrices with i.i.d. or highly structured entries. Starting in the mid 2000s, results from Random Matrix Theory were used along these lines to investigate problems related to financial data, particularly the out-of-sample risk underestimation of portfolios constructed through mean-variance optimization. If the returns of assets to be held in a portfolio are assumed independent and stationary, then these results are universal in that they do not depend on the precise distribution of returns. This universality has been somewhat misrepresented in the literature, however, as asymptotic results require that an arbitrarily long time horizon be available before such predictions necessarily become accurate. This makes these methods ill-suited when moving to high frequency data, for example, where the number of data-points increases but the overall time horizon remains the same or even decreases. In order to reconcile these models with the highly non-Gaussian returns observed in financial data, a new ensemble of random rectangular matrices are introduced, modeled on the observations of independent Lévy processes over a fixed time horizon. The key mathematical results describe the eigenvalues of these models' sample covariance matrices, which exhibit remarkably similar scaling behavior to what is seen when working with daily and intraday data on the S&P 500 and Nikkei 225.

Chapter 1

Introduction

Financial datasets have scaled in extreme ways; it is now possible to observe changes in the price of an asset over the period of a minute, a second, a nanosecond. The motivation for the work presented in this manuscript was the failure of previous models to accurately describe the scaling behavior observed in these larger and larger datasets. When considering data of a fixed size, the mathematical field of Random Matrix Theory has found numerous applications in finance (Bai et al., 2009; Burda et al., 2011; El Karoui, 2013; Liu et al., 2016; Bun et al., 2017; Choi et al., 2019, to list only a few), although the results used are usually those based on the asymptotic behavior of eigenvalues for large rectangular matrices. The flavor of these theorems is similar to statistical mechanics: the movement of individual air molecules in a room is chaotic and unpredictable, but the combined motion of all particles results in a steady concept of "temperature" after a certain amount of time has passed. Similarly, the Marčenko–Pastur (M–P) law (1967) and its generalization (Silverstein, 1995) describe the shape of the eigenvalues in a system with large, independent, chaotic fluctuations. In this scenario, each entry in the matrix ensemble is random, as are individual eigenvalues. In spite of this, the mass interactions of independent motion causes the percentage of eigenvalues of a certain size to become predictable as the overall size of the system grows.

The M–P law is used to describe the non-random behavior of eigenvalues in systems or data where no signal is present. By "signal" we mean a strong linear relationship; in financial markets these may be realized as news about changing interest rates, the US equities market as a whole, the behavior of a particular sector, etc. Like most financial modeling, this news itself is thought of as *random*, so it is not a signal in the traditional sense: it is not information sent by an agent with the intention that it will be observed. Nonetheless, the detection of such a signal (that is, the detection of strong linear relationships in the movement of the underlying assets) may help agents in accomplishing specific goals, e.g. portfolio construction via mean-variance optimization.

The law can be paraphrased in the following way. Consider the scenario where N observations about a set of p features are collected, such that the number of observations and the features are both comparably large. The statement of the law is that the sample covariance of pure noise, an estimation of a trivial structure, will exhibit a predictable bulk of eigenvalues, which is a distinctly non-trivial structure. In simpler terms, out of a tremendous amount of noise, any observer will begin to believe they are recognizing meaningful trends. The strength of these phantom trends depends on the ratio of the features to observations, $\lambda = p/N$. This interpretation agrees with the idea of performing portfolio optimization using historical stock data; given noisy measurements, an agent will mistakenly optimize over the noise as well as the signal. The amount by which noise influences the process of mean-variance portfolio optimization specifically was investigated as early as Laloux et al. (1999), with a rigorous proof finally given in Bai et al. (2009).

Like many results in Random Matrix Theory, the M–P law is a statement about the asymptotic behavior of random matrices which grow larger and larger. Despite this, it has remained attractive in applications involving datasets of fixed size due to its universality and speed of convergence. This universality is typically expressed in the standard presentation of the law, where the independent rows of the matrix ensemble have i.i.d. entries. Under these circumstances, the only assumption necessary for the law to hold is that the variance of these entries be finite. At the same time, modestly sized matrices with rows and columns of sizes on the order of only 10^2 have been used as empirical evidence of fast convergence. What is fascinating is that these two properties are never shown together, and the eigenvalues plotted to demonstrate the effectiveness of the law often use matrices with pseudo-random normal or Rademacher entries (see the figures in Baik and Silverstein, 2006; Burda et al., 2011; El Karoui, 2013; Yao et al., 2015). Such distributions exhibit small kurtosis, in stark contrast to the highly leptokurtic distributions which typically appear when observing and modeling asset returns.

Consider the application of M–P-inspired techniques in modeling a simple financial scenario. We imagine a large data matrix of size $N \times p = 2000 \times 500$, representing the daily linear returns on a collection of assets such as the S&P 500 over an 8 year window. Eigenvalue cleaning recipes (see the recent survey of Bun et al., 2017) and mean-variance portfolio bias estimation techniques such as those suggested by Bai et al. (2009) will implicitly rely on the limiting measure for the eigenvalues of a M–P type sample covariance matrix with parameter for $\lambda = 1/4$. Indeed, the histogram of eigenvalues of $\frac{1}{N}\mathbf{Y}^\dagger\mathbf{Y}$, where $\mathbf{Y}$ is an $N \times p = 2000 \times 500$ matrix $\mathbf{Y}$ with $N \cdot p$ i.i.d. standard normal entries, will be strikingly similar to the the limiting M–P distribution with $\lambda = 1/4$ (Figure 1.1a). However, universality of the law is not a statement about the fixed value $N = 2000$, and a similar matrix composed of i.i.d. normalized Lognormal random variables, a distribution with finite moments of arbitrary order, will instead produce eigenvalues outside of the predicted bulk (Figure 1.1b). Such a matrix appears at some point in both models, where it is assumed that N is large enough so that the distribution of the entries is irrelevant.

Methodological problems might also be considered when these models are applied to financial time series. If each row of the matrix ensemble is taken to be derived from daily returns on a collection of assets, then N denotes the number of days for which data has been collected.

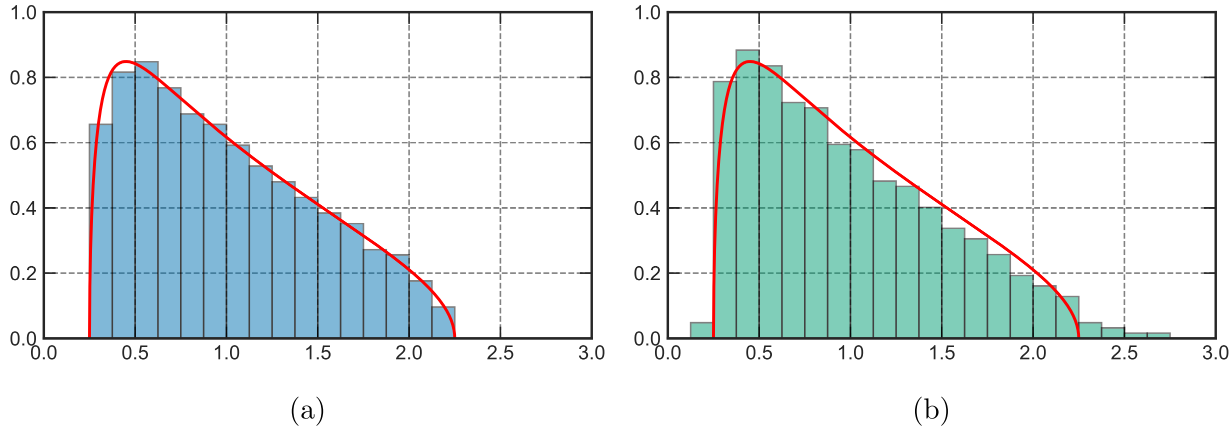


Figure 1.1: Comparison of the Marčenko–Pastur distribution mp_λ (red line) for $\lambda = p/N = 1/4$ to the histogram of eigenvalues of $(1/N)\mathbf{Y}^\dagger\mathbf{Y}$, where $\mathbf{Y}$ is taken to be an $N \times p = 2000 \times 500$ matrix with i.i.d. entries drawn from (a) standard normal and (b) normalized Lognormal distributions.

The asymptotic $N \to \infty$ therefore implies convergence over an arbitrarily long time horizon. In essence, although the individual entries of the matrix (representing daily returns) may be non-Gaussian, the theorem relies on the fact that N can be taken large enough so that the columns of our matrices closely resemble the fluctuations of standard Brownian motion over a finite interval. This is consistent with the popular notion that equity price fluctuations are well modeled by Gaussian only in the long run. On the other hand, the restriction of financial applications to fixed time horizons would suggest that this assumption is unrealistic in many scenarios, as portfolio construction may well be employed over periods of time where the return process is still highly non-Gaussian.

An alternative approach for the purpose of modeling financial data is to design a random matrix ensemble whose entries describe the fluctuations of a collection of p stochastic processes over a fixed horizon $[0, T]$. Rather than the asymptotic $N \to \infty$ signifying additional observations beyond the horizon, we suppose that our observations are occurring at finer and finer discretizations of the interval $0 = t_0 < t_1 < \ldots < t_N = T$. If the rows of the matrix are taken to be i.i.d. (for each fixed N), then this corresponds to an equally spaced net $t_j = jT/N$, and the observations can be thought of as the fluctuations of p independent trajectories of a Lévy process X_t. This coincides with a choice to model the linear or log-price process of asset returns on a Lévy process, which has been a popular choice for many years (see selected chapters and discussions in Carr et al., 2002; Voit, 2005; Jondeau et al., 2007; Jeanblanc et al., 2009; Pascucci, 2011; Fischer, 2014; Maejima, 2015). We propose this as an alternative type of matrix ensemble, whose entries are i.i.d. but vary as $N \to \infty$ such that the sum of the entries in each column matches a fixed, infinitely divisible distribution. We call this new ensemble of random matrices the Sample Lévy Covariance Ensemble (SLCE). The purpose of this work is to introduce this ensemble and prove the existence of a limiting eigenvalue distribution, as well as some broad qualitative properties.

Our departure from previous work in Random Matrix Theory can be understood along a few lines. The classical scenario investigated by Marčenko and Pastur (1967) involves a large random matrix $\mathbf{Y}$ of size $N \times p$, where N and p are both large but comparable, and considers the deterministic limiting histogram of its singular values as $N, p \to \infty$ with $p/N \to \lambda \in (0, 1)$. Initially, the assumptions on the entries of $\mathbf{Y}$ is that they are i.i.d. and follow some fixed, finite variance distribution for all N. This condition can be relaxed somewhat; up to rescaling (and observing instead the eigenvalues of $\mathbf{Y}^\dagger\mathbf{Y}$), we can consider an ensemble of matrices $\mathbf{Y}$ whose entries are i.i.d. following a changing distribution Y_N whose variance is $O(N^{-1})$, with the one additional assumption that the law of the entries satisfies (Bai and Silverstein, 2010, Theorem 3.10)

$$N \cdot \mathbb{E}\Big[\big|Y_N\big|^2 \mathbb{1}_{|Y_N| \geq \eta}\Big] \xrightarrow{N\to\infty} 0$$

for any $\eta > 0$. Any such ensemble falls into the Marčenko–Pastur basin of attraction. Therefore, in order to escape the M–P universe, we are interested in matrices whose entries are i.i.d. random variables drawn from changing distributions which become less Gaussian as N grows. We are strongly motivated by the conclusions in Carr et al. (2002) that the diffusion components of financial data are likely diversifiable, suggesting that noise and small idiosyncratic factors in the market may be more appropriately modeled by a pure point process. The SLCE does precisely this, taking the i.i.d. entries of our matrix to follow the distributions of $X_{T/N}$, where $T > 0$ is a fixed horizon parameter and X_t is a Lévy process. If the right tail of X_t is subexponential, then $\mathbb{P}[X_{T/N} > \eta] \sim \frac{T}{N}\Pi((\eta, \infty))$ as $\eta \to \infty$ (Sato, 2013, Remark 25.14) for the Lévy measure Π (see Chapter 3), and we have that

$$N \cdot \mathbb{E}\left[\left|X_{T/N}\right|^2 \mathbb{1}_{|X_{T/N}| \geq \eta}\right] \sim N \cdot \frac{T}{N} \int_\eta^\infty x^2 d\Pi(x) \sim O(1)$$

This is assuming the variance of X_t even exists. This is significantly different from the Gaussian case, where if $X_{T/N} \in \mathrm{N}(0, T/N)$ then its tails are asymptotically

$$O\left(\sqrt{N} e^{-\eta^2 N/2T}\right), \quad \eta \to \infty$$

Consequently, it is reasonable to expect that for a non-Gaussian Lévy process, the SLCE has the capacity to behave quite differently from the M–P law.

Similarly, one might compare our results to the theory of heavy-tailed random matrices, which was rigorously founded on the work of Ben Arous and Guionnet (2008) and Belinschi et al. (2009a). This began with the study of large square matrices whose entries are heavy-tailed, lying in the domain of attraction of α-stable distributions. Some such banded matrices mimic the M–P case for a particular shape parameter $\lambda \in (0, 1)$, leading to a notion of heavy-tailed (or Lévy) M–P type matrices. Connections with Free Probability were later made in Politi et al. (2010), where they considered "Free Wishart" matrices of the form

$$\left(\frac{1}{M} \sum_{j=1}^{M} \mathbf{U}_j \mathbf{L}_j \mathbf{U}_j^\dagger\right) \mathbf{P}_\lambda$$

Here the $\mathbf{L}_j$ are $N \times N$ matrices with i.i.d. heavy-tailed entries, while the $\mathbf{U}_j$ are independent Haar distributed unitary matrices, and $\mathbf{P}_\lambda$ is a projection onto a subspace of dimension λN.

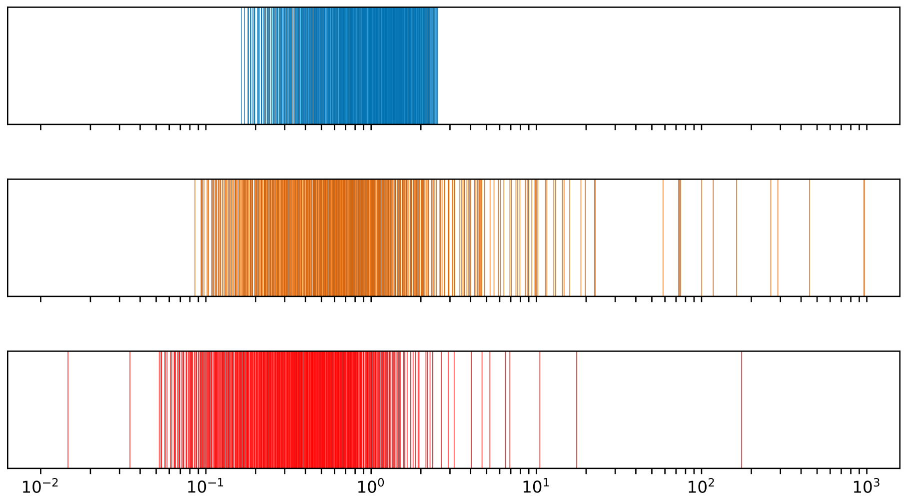


Figure 1.2: Comparison of the eigenvalues of a Marčenko–Pastur matrix, a Lévy matrix, and actual S&P returns. All matrices have size $N \times p = 1258 \times 454$, for a ratio of $\lambda = p/N \approx 0.36$. The Marčenko–Pastur matrix (top) has i.i.d. $\mathrm{N}(0, 1)$ entries. The Lévy matrix (middle) has i.i.d. $\alpha = 3/2$ entries. The S&P data (bottom) is taken from the 462 stocks that appeared in the S&P 500 at any time during the years 2011–2015, and for which there is complete data during that five year period (1258 data points). Neither of the random matrix ensembles appear to adequately capture the bulk of the S&P eigenvalues.

In both cases, the limiting eigenvalue distributions have unbounded right tails which decay like $1/|x|^{\alpha+1}$ as $x \to \infty$. Such a theory is particularly devastating for a model of asset prices, as it implies that noisy eigenvalues of large size may occur with a heavy-tailed frequency. Figure 1.2 contrasts the M–P case and heavy-tailed Lévy case with data from the S&P 500 over the years 2011–2015. Neither matrix model looks similar to the empirical bulk eigenvalues; a very different situation then that implied by Laloux et al. (1999).

In contrast to previous matrix models, the SLCE is parametrized by a Lévy process X_t, which can have a variety of tail behaviors. This follows the general perspective on asset modeling established in Mantegna and Stanley (1994, 1995) and explored in texts like Voit (2005) and Jondeau et al. (2007), where classes of Lévy processes are used in place of Brownian motion. This bridges the gap between the overly conservative M–P setting, which occurs when X_t is Brownian motion, and the wild heavy-tailed setting, which can be captured by taking X_t

as the standard one-dimensional Lévy flight process. Most importantly, however, the SLCE appears to match the scaling behavior observed in the S&P 500 and Nikkei 225 universes when passing from daily to intraday datasets, as discussed in the final sections of this work.

The structure of this work is as follows. Chapter 2 quickly summarizes the necessary mathematical framework for the rest of the document. Chapter 3 introduces and classifies Lévy processes and infinitely divisible distributions. Since our matrix ensembles are intimately tied together with the study of such distributions, many examples are discussed. Chapter 4 defines the Sample Lévy Covariance Ensemble (SLCE), the main object of study. From here, the existence and continuity of limiting distributions for SLCE are established, including our two most significant theorems: Theorem 4.2.1 on the limiting distributions of SLCE's driven by essentially bounded Lévy processes, and Theorem 4.0.2 on the limiting distributions of general SLCE's. The proofs here involved a significant shift in perspective in order to avoid techniques involving Stieltjes transforms, and were guided heavily by Benaych-Georges's work on Rectangular Free Probability as formulated in the seminal articles 2009a; 2009b.

From here, we move to Chapter 5. This interlude covers the topic of Free Probability using simple Complex Analytic techniques. Short introductions to Rectangular Free Probability and its application to sums of rectangular random matrices are discussed. Chapter 6 investigates the gaps in the proof of the M–P law which allow our ensembles to have distinct eigenvalue distributions, and how these gaps might be bridged using Free Probability. An estimation algorithm for the shape of the limiting distributions, whose existence is guaranteed by Theorem 4.0.2, are discussed. The final Section 6.3 approaches problems arising in finance, as established in this introduction.

This work proposes a new type of random matrix model which is distinct from others popularized over the past few decades in Random Matrix Theory, including those which have seen applications in industry. In spite of this, it is the goal of this work to convince the reader that the design of these ensembles is quite natural. It is a fact that increasing the observations of a Brownian motion process will decrease the error of the sample variance, with no lower bound. This is true regardless of whether additional observations are made by extending the time horizon with discrete points in the future, or by refining the number of points sampled during the current horizon. Such a statement cannot be true for a non-Gaussian Lévy process. Despite the generous sufficient conditions for the M–P law, it must still be understood as a statement about matrices whose asymptotic behavior is Gaussian-like. More recent proofs and generalizations of the law provide support for this viewpoint (Yaskov, 2016a,b), where it is framed as a type of concentration phenomenon much like how large multivariate Gaussian random vectors cluster around the ellipse determined by their covariance matrices. The use of the M–P law to drive financial models may therefore be viewed as a first approximation with strong underlying normality assumptions.

Chapter 2

Preliminary Materials

2.1 Probabilistic Preliminaries

Throughout, we assume the existence of a standard probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a countable collection of random variables, sequences, and cádlág processes, as described in Appendix A.1. The expressions $\mathcal{P}(\mathbb{R})$ and $\mathcal{P}(\mathbb{C})$ denote the collections of probability measures on $\mathbb{R}$ and $\mathbb{C}$, respectively. The class $\mathcal{P}(\mathbb{R}^+)$ indicates those distributions with support contained in $\mathbb{R}^+ = [0, \infty)$. We adopt the typical convention that the point mass at zero is not contained in $\mathcal{P}(\mathbb{R}^+)$, formally: $\mu \in \mathcal{P}(\mathbb{R}^+)$ if and only if $\operatorname{supp}(\mu) \subseteq [0, \infty)$ and $\operatorname{supp}(\mu) \neq \{0\}$. Capital Latin letters such as X and Y will be used to denote random variables, while Greek letters like μ and ν denote measures. The expression $\mathcal{L}(X)$ refers to the law or distribution of a random variable. Whenever it is clear from the context, we use random variables and their distributions interchangeably in the notation, as in φ_X to denote the characteristic function $\varphi_{\mathcal{L}(X)}$ in Section 2.1.2. This include the abuse of notation $X \in$ Class when the distribution of a random variable falls in a particular class of probability measures. The expression $X \overset{d}{=} Y$ means equality in distribution, which is equivalent to $\mathcal{L}(X) = \mathcal{L}(Y)$. Similarly, weak convergence (or convergence in distribution) is indicated by $X_n \xrightarrow{d} X$ for random variables and $\mu_n \xrightarrow{d} \mu$ for probability measures.

A real-valued random variable X is said to be symmetric if $X \stackrel{d}{=} -X$, and the class of symmetric distributions is denoted by $\mathcal{P}_s(\mathbb{R})$. A complex-valued random variable X is said to be circularly symmetric if $X \stackrel{d}{=} e^{i\theta}X$ for any $\theta \in [0, 2\pi)$, and the class of circularly symmetric distributions is denoted by $\mathcal{P}_c(\mathbb{C})$. The notation $\mu * \nu$ indicates the convolution measures, which coincides with the sum of independent random variables. Specifically, if X and Y are independent, then

$$\mathcal{L}(X + Y) = \mathcal{L}(X) * \mathcal{L}(Y)$$

When $\mu \in \mathcal{P}(\mathbb{C})$ is a probability measure, the notation μ^2 refers to the pushforward measure, such that if $\mu = \mathcal{L}(X)$ then $\mu^2 = \mathcal{L}(X^2)$. If $\mu \in \mathcal{P}(\mathbb{R}^+)$ is nonnegative, then the notation $\sqrt{\mu}$ refers explicitly to the symmetric square root measure $\sqrt{\mu} \in \mathcal{P}_s(\mathbb{R})$ such that $\sqrt{\mu}^2 = \mu$.

2.1.1 Chernoff bounds

Theorem 2.1.1 (Markov's Inequality)**.** *Suppose $X \geq 0$ is a random variable, and let $a > 0$. Then*

$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$$

where the right hand side is understood to be infinite if $\mathbb{E}[X]$ is undefined.

Proof. We have an obvious inequality between the three random variables

$$a \cdot \mathbb{1}_{X \geq a} \leq X \cdot \mathbb{1}_{X \geq a} \leq X$$

The expectation of the left is $a \cdot \mathbb{P}[X \geq a]$, while the right is $\mathbb{E}[X]$, and so the result follows.

$\blacksquare$

Theorem 2.1.2 (Chernoff Bound). *Let $X_1, \ldots, X_p$ be i.i.d. random variables, each following a Bernoulli distribution with probability of success q, and let*

$$X = X_1 + \ldots + X_p$$

denote their sum. If $\delta > 0$, then

$$\mathbb{P}[X \geq (1+\delta)pq] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{pq}$$

Proof. We apply Markov's inequality to the random variable e^{tX}, where $t > 0$ is some parameter to be determined later. It follows that

$$\begin{aligned}
\mathbb{P}[X \geq (1+\delta)pq] &= \mathbb{P}\left[e^{tX} \geq e^{t(1+\delta)pq}\right] \leq e^{-t(1+\delta)pq} \prod_{j=1}^{p} \mathbb{E}\left[e^{tX_j}\right] \\
&= e^{-t(1+\delta)pq} \left(1 - q + qe^{t}\right)^{p}
\end{aligned}$$

Substituting $t = \log(1+\delta) > 0$, we get that

$$\mathbb{P}[X \geq (1+\delta)pq] \leq \frac{(1+q\delta)^p}{(1+\delta)^{(1+\delta)pq}} \leq \frac{e^{\delta pq}}{(1+\delta)^{(1+\delta)pq}} = \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{pq}$$

$\blacksquare$

In particular, choosing $\delta = 1$ in Theorem 2.1.2 leads to the fact that for the sum X of p i.i.d. Bernoulli random variables with success rate q, we get

$$\mathbb{P}[X \geq 2pq] \leq \left(\frac{e}{4}\right)^{pq} \tag{2.1}$$

which will be important in the sequel.

2.1.2 Transformations of real-valued random variables

For real-valued probability measures $\mu \in \mathcal{P}(\mathbb{R})$, we let $F_\mu(x) \triangleq \mu((-\infty, x])$ denote the cumulative distribution function (CDF) of μ. If some open $A \subseteq \mathbb{R}$ exists such that $F'_\mu(x)$ is well defined for all $x \in A$, then we use the notation $f_\mu(x) = F'_\mu(x)$ for the probability density function (PDF) where it is defined. The characteristic function of μ, denoted by $\varphi_\mu : \mathbb{R} \to \mathbb{C}$, is defined as

$$\varphi_\mu(z) \triangleq \mathbb{E}[e^{izX}] = \int_{\mathbb{R}} e^{ixz} d\mu(x), \quad z \in \mathbb{R} \tag{2.2}$$

The process of passing from measures to their characteristic functions inherits a number of desirable properties from the theory of commutative Banach algebras (see, for instance, Kaniuth, 2009). For any probability measure, φ_μ is a bounded continuous function, and the map

$$\varphi : \mathcal{P}(\mathbb{R}) \to \mathrm{C}_b(\mathbb{R})$$

into the set of bounded continuous functions $\mathrm{C}_b(\mathbb{R})$ is injective. We say that a distribution is continuous if for any Borel $B \subseteq \mathbb{R}$, we have

$$\mu(B) = \int_{A \cap B} f_\mu(x) dx$$

for the open set $A \subseteq \mathbb{R}$ appearing above. If μ is continuous, then

$$\varphi_\mu(z) = \int_{\mathbb{R}} e^{ixz} f_\mu(x) dx, \quad z \in \mathbb{R}$$

Since f_μ is integrable, the characteristic function φ_μ can be viewed as the Fourier transform of an integrable function.

If X and Y are independent random variables, then we have the classic distributive equation

$$\varphi_{X+Y}(z) = \varphi_X(z)\varphi_Y(z), \quad z \in \mathbb{R}$$

In terms of probability measures, this can be written instead using the convolution of measures,

$$\varphi_{\mu * \nu}(z) = \varphi_\mu(z)\varphi_\nu(z), \quad z \in \mathbb{R}$$

Since φ is continuous and $\varphi_X(0) = 1$ for any real-valued random variable X, it follows that there is a neighborhood of the origin $\mathcal{O}_X \subseteq \mathbb{R}$ on which $\varphi_X(\mathcal{O}_X) \subseteq \mathbb{C} \backslash \{0\}$. If $\varphi_X(z) \neq 0$ for any $z \in \mathbb{R}$, then we simply take $\mathcal{O}_X = \mathbb{R}$. On such a set, the multi-valued complex logarithm of φ_X is well defined and continuous on $\mathcal{O}_X$. We call this function the cumulant generating function (CGF) of X, denoted by

$$\psi_X(z) \triangleq \log \varphi_X(z), \quad z \in \mathcal{O}_X$$

The distribution property of the CGF is now additive, so that the transform can be said to distribute over independent addition of random variables such that

$$\psi_{X+Y}(z) = \psi_X(z) + \psi_Y(z), \quad z \in \mathcal{O}_X \cap \mathcal{O}_Y$$

when X and Y are independent random variables.

If a probability measure $\mu \in \mathcal{P}(\mathbb{R}^+)$, then it is sometimes more convenient to work with the Laplace transform of μ, defined in the right half plane as

$$\phi_\mu(s) \triangleq \mathbb{E}[e^{-sX}] = \int_{\mathbb{R}} e^{-sx} d\mu(x), \quad \mathrm{Re}(s) \geq 0$$

The Laplace transform agrees with the characteristic function for purely imaginary arguments, such that $\phi_\mu(iy) = \varphi_\mu(-y)$ for $y \in \mathbb{R}$. Since the Laplace transform is analytic on the interior of its domain, the distribution μ is uniquely determined by the values of ϕ_μ on a set of analytic capacity; a typical set used is $(0, \infty)$.

2.1.3 Moments and cumulants

The moments of a real-valued random variable X, if they exist and are finite, are defined as

$$m_n[X] = \mathbb{E}[X^n]$$

If a random variable X has finite moments up to order $r \in \mathbb{N}$, then its characteristic function is r-times continuously differentiable and has Taylor expansion at zero given by

$$\varphi_X(z) = \sum_{k=0}^{r} \frac{m_k[X]}{k!}(iz)^k + o(|z|^r)$$

In such a scenario, the CGF of X is also r-times continuously differentiable, with Taylor expansion in $\mathcal{O}_X$ given by

$$\psi_X(z) = \sum_{k=1}^{r} \frac{\kappa_k[X]}{k!}(iz)^k + o(|z|^r) \tag{2.3}$$

The values $\kappa_n[X]$ which appear in this expression are called the cumulants of X. By additivity of the CGF, we have that $\kappa_n[X+Y] = \kappa_n[X]+\kappa_n[Y]$ whenever X and Y are independent and the later two quantities are well defined. By definition of the CGF, the relation between the moments and cumulants can be made explicit:

$$m_n[X] = \sum_{\pi} \prod_{B\in\pi} \kappa_{|B|}[X] \tag{2.4}$$

where the sum runs over all partitions π of the set $\{1, 2, \ldots, n\}$, and the elements $B \in \pi$ are subsets of $\{1, 2, \ldots, n\}$. So, for instance, the third moment can be expressed in terms of the first three cumulants by considering the five partitions of $\{1, 2, 3\}$, and so

$$\begin{aligned} m_3[X] = &\prod_{B\in\{\{1\},\{2\},\{3\}\}} \kappa_{|B|}[X] + \prod_{B\in\{\{1,2\},\{3\}\}} \kappa_{|B|}[X] + \prod_{B\in\{\{1,3\},\{2\}\}} \kappa_{|B|}[X] \\ &+ \prod_{B\in\{\{1\},\{2,3\}\}} \kappa_{|B|}[X] + \prod_{B\in\{\{1,2,3\}\}} \kappa_{|B|}[X] \\ = &\ \kappa_1[X]^3 + \kappa_2[X]\kappa_1[X] + \kappa_2[X]\kappa_1[X] + \kappa_1[X]\kappa_2[X] + \kappa_3[X] \\ = &\ \kappa_1[X]^3 + 3\kappa_2[X]\kappa_1[X] + \kappa_3[X] \end{aligned}$$

Moments and cumulants are similarly well defined for complex-valued random variables, where they are referred to as $*$-moments and $*$-cumulants (Eriksson et al., 2009). These $*$-statistics are computed using conjugate pairs, so that a random variable X has $*$-moments defined as

$$m_{k:l}[X] \triangleq \mathbb{E}\left[X^k \overline{X}^l\right]$$

The $*$-cumulants $\kappa_{k:l}[X]$ are computed using a formula similar to (2.4). The cumulant formula for the variance (the 1 : 1 symmetric cumulant) is what one would expect:

$$\mathrm{var}[X] \triangleq \kappa_{1:1}[X] = m_{1:1}[X] - m_{1:0}[X]m_{0:1}[X] = \mathbb{E}\left[|X|^2\right] - |\mathbb{E}[X]|^2$$

The definition of the kurtosis, the normalized 2 : 2 symmetric cumulant, takes a slightly more complicated form:

$$\begin{aligned}
\mathrm{kurt}[X] &\triangleq \frac{\kappa_{2:2}[X]}{\kappa_{1:1}[X]^2} \\
&= \Big(m_{2:2}[X] - m_{2:0}[X]m_{0:2}[X] \\
&\qquad + 6m_{1:1}[X]m_{1:0}[X]m_{0:1}[X] - 6m_{1:0}[X]^2 m_{0:1}[X]^2 \\
&\qquad + 3m_{1:0}[X]^2 m_{0:2}[X] + 3m_{0:1}[X]^2 m_{2:0}[X] \\
&\qquad - 2m_{2:1}[X]m_{1:0}[X] - 2m_{1:2}m_{1:0}[X] - 2m_{1:1}[X]^2\Big) \Big/ \kappa_{1:1}[X]^2 \\
&= \frac{m_{2:2}[X - m_{1:0}[X]]}{\mathrm{var}[X]^2} - \frac{|m_{2:0}[X - m_{1:0}[X]]|^2}{\mathrm{var}[X]^2} - 2 \qquad (2.5)
\end{aligned}$$

The kurtosis can be seen as a measure of the tail deviation of a random variable from being Gaussian. If X is real-valued, then the middle term in (2.5) is equal to 1, and so the expression reduces to

$$\frac{\mathbb{E}\left[\left|X - \mathbb{E}[X]\right|^4\right]}{\mathrm{var}[X]^2} - 3$$

Since the fourth moment of a standard real-Gaussian $\mathrm{N}(0, 1)$ is equal to 3, this measures how far away such a distribution is from being real-Gaussian. On the other hand, if X is a circularly symmetric complex-valued random variable, then the middle term is equal to 0 and the expression reduces to

$$\frac{\mathbb{E}\left[\left|X - \mathbb{E}[X]\right|^4\right]}{\mathrm{var}[X]^2} - 2$$

This situation is analogous, as the fourth moment of a standard complex-Gaussian $\mathrm{CN}(0,1)$ equals 2.

2.2 Random Matrix Preliminaries

The two central objects we will employ in the study of random matrices and their eigenvalues are the Empirical Spectral Distribution (ESD) of a random matrix and the Stieltjes transform. If $\mathbf{S}$ is a $p \times p$ Hermitian matrix, then its ESD is defined as the probability measure $\mu_{\mathbf{S}}$ given explicitly in terms of point masses by

$$\mu_{\mathbf{S}} \triangleq \frac{1}{p} \sum_{j=1}^{p} \delta_{\lambda_j} \in \mathcal{P}(\mathbb{R})$$

where the λ_j enumerate the eigenvalues of $\mathbf{S}$. This measure can also be described in terms of the Borel sets on $\mathbb{R}$ by considering that

$$\mu_{\mathbf{S}}([a,b]) = \frac{1}{p}\,(\text{Number of eigenvalues of } \mathbf{S} \text{ in } [a,b])$$

For a finite Borel measure μ on $\mathbb{C}$, we can consider the following integral transform

$$S_\mu(z) = \int_{\mathbb{C}} \frac{1}{w-z}\, d\mu(w)$$

This is sometimes called the Cauchy transform or the Cauchy–Stieltjes transform, although the former is often reserved for the case when the support of μ is contained in the unit circle $\mathbb{T}$. The transform $S_\mu(z)$ is finite for almost all $z \in \mathbb{C}$, locally Lebesgue integrable, and analytic outside of the support of μ as the following lemma shows. When μ is compactly supported, $S_\mu(z)$ can be represented as the derivative of the logarithmic potential of μ (Bøgvad and Shapiro, 2016).

Lemma 2.2.1. *The Stieltjes transform S_μ of a finite Radon measure μ on $\mathbb{C}$ is well defined and holomorphic on the domain of its argument $z \in \mathbb{C}\backslash\mathrm{supp}(\mu)$.*

Proof. By the definition we have

$$S_\mu(z) = \int_{\mathbb{C}} \frac{1}{x-z}\, d\mu(x) = \int_{\mathrm{supp}(\mu)} \frac{1}{x-z}\, d\mu(x)$$

Since the map $x \mapsto \frac{1}{x-z}$ is in $\mathrm{C}_0(\mathrm{supp}(\mu))$ for $z \in \mathbb{C}\backslash\mathrm{supp}(\mu)$, it follows that this integral is well defined. To see that it is holomorphic, let Δ be any simple closed piecewise C^1 curve in $\mathbb{C}\backslash\mathrm{supp}(\mu)$, parametrized by arclength using $\gamma : [0,1] \to \Delta$. Since Δ and $\mathrm{supp}(\mu)$ are disjoint closed sets in $\mathbb{C}$, they have some positive distance

$$d(\Delta, \mathrm{supp}(\mu)) = \inf_{x\in\mathrm{supp}(\mu), z\in\Delta} |x-z| = \delta > 0$$

Then it follows that $\left|\frac{1}{x-z}\right| < \frac{1}{\delta}$ for any such $z \in \Delta$ and $t \in \mathrm{supp}(\mu)$. By Fubini's theorem applied to $\frac{1}{x-\gamma(t)}\gamma'(t)$ on the space $\mathbb{C} \times [0,1]$ under the measure $\mu \times m$ where m is Lebesgue measure on $[0,1]$,

$$\begin{aligned}
\oint_\Delta S_\mu(z)\ dz &= \int_0^1 S_\mu(\gamma(t))\gamma'(t)\ dt \\
&= \int_0^1 \left[\int_{\mathbb{C}} \frac{1}{x-\gamma(t)}\ d\mu(x)\right]\gamma'(t)\ dt \\
&= \int_{\mathbb{C}} \left[\int_0^1 \frac{1}{x-\gamma(t)}\gamma'(t)\ dt\right] d\mu(x) \\
&= \int_{\text{supp}(\mu)} \left[\oint_\Delta \frac{1}{x-z}dz\right] d\mu(x) = 0
\end{aligned}$$

Since this can be done for any such curve, by Morera's theorem S_μ is holomorphic on its domain. ■

When $\mu \in \mathcal{P}(\mathbb{R})$, we recover the classical Stieltjes transform

$$S_\mu(z) = \int_{\mathbb{R}} \frac{1}{x-z}\ d\mu(x)$$

which is well defined for z in the upper half plane $\mathbb{C}^+$. The Stieltjes transform is analytic outside of the support of μ and maps $\mathbb{C}^+$ into its closure. Such analytic functions on $\mathbb{C}^+$ with the property that $f : \mathbb{C}^+ \to \overline{\mathbb{C}^+}$ are called Nevanlinna functions $\mathcal{N}$, and will be important in Chapter 5.

We can recover a probability measure through its Stieltjes transform with the following inversion theorem.

Theorem 2.2.2 (Stieltjes Inversion, Yao et al., 2015)**.** *Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous and compactly supported function, then*

$$\int_{\mathbb{R}} g(x)\ d\mu(x) = \lim_{y\to 0^+} \frac{1}{\pi} \int_{\mathbb{R}} g(x)\ \text{Im}\left[S_\mu(x+iy)\right] dx$$

Furthermore, let $a < b$ be continuity points of F_μ. Then it follows that

$$\mu([a,b]) = \lim_{y \to 0^+} \frac{1}{\pi} \int_a^b \mathrm{Im}\left[S_\mu(x+iy)\right] dx$$

This theorem shows that the $\mathbb{C}^+$ pointwise convergence of a sequence of Stieltjes transforms corresponds to the vague convergence of measures. The density of the continuous, compactly supported functions in the space $\mathrm{C}_0(\mathbb{R})$ under the supremum norm guarantees that the first part of the theorem holds for such functions as well. If μ corresponds to a random variable X with density f_X, then

$$f_X(x) = \lim_{y \to 0^+} \frac{1}{\pi} \mathrm{Im}[S_\mu(x+iy)]$$

If the random variable X has finite moments, its Stieltjes transform has a germ at ∞ which expands as the Laurent series

$$S_X(z) = -\frac{1}{z} - \sum_{j=1}^{\infty} \frac{m_j[X]}{z^{j+1}}$$

From this we can see that the (complex) moment generating function for X can be expressed as

$$1 + \sum_{j=1}^{\infty} m_j[X](iz)^j = -\frac{1}{iz} \cdot S_X(1/iz)$$
$$= \int_{\mathbb{R}} \frac{1}{1-izx}\, dF_X(x) = \mathbb{E}\left[\frac{1}{1-izX}\right]$$

where $z \in \mathbb{C}^+$ is taken in the upper half plane.

2.2.1 Marčenko–Pastur Theorem

Consider a collection of p assets whose prices $\{P_n^j\}_{n=0,\ldots,N}^{j=1,\ldots,p}$ are observed over $N+1$ time periods (e.g. days). Models of equity markets often suppose (Bouchaud and Potters, 2000; Voit, 2005; Meucci, 2009) that the collection of compounded returns over a unit time interval

$$\left[\log \frac{P_n^1}{P_{n-1}^1} \;\ldots\; \log \frac{P_n^p}{P_{n-1}^p}\right]$$

can be represented as independent samples of a p-dimensional random vector, possibly after accounting for autocorrelation. Often this random vector is assumed to be normally distributed; at the very least, we would like to suppose that an underlying $p \times p$ covariance matrix $\mathbf{\Sigma}$ exists, whose entries encode a robust set of relationships between the returns on these assets.

Suppose these compounded returns are arranged into an $N \times p$ matrix $\mathbf{X}$, whose columns represent separate assets and whose rows represent dates, such that the entry $[\mathbf{X}]_{ij}$ is the return on the j^{th} asset between day $i-1$ and day i. If the columns of $\mathbf{X}$ have been centered, then the sample covariance matrix is given by

$$\mathbf{S} = \frac{1}{N}\mathbf{X}^\dagger \mathbf{X}$$

There are a number of important questions regarding the properties of this matrix $\mathbf{S}$ and its eigenvalues.

- ▶ The eigenvalues of $\mathbf{S}$ are in correspondence with the singular values of $\mathbf{X}$. Large eigenvalues separated from the bulk of the spectrum imply that there are significant relationships between the columns of $\mathbf{X}$. The factor loadings corresponding to these singular values may be used to identify sectors among the assets, and to measure the extent of their influence. The bulk eigenvalues are often (Bouchaud and Potters, 2011;

Paul and Aue, 2014; Bun et al., 2017) product of noisy measurements in the matrix $\mathbf{X}$. When the large eigenvalues are not clearly separated from the bulk, however, we need more sophisticated techniques to determine which values should be viewed as carrying meaningful information.

- According to mathematical formulation of the optimization problems involved, mean-variance portfolios are constructed using the inverse $\mathbf{\Sigma}^{-1}$ of the underlying covariance matrix. If $\mathbf{S}$ is a poor estimation of $\mathbf{\Sigma}$, this may compromise the construction and out-of-sample performance of these portfolios (Bai et al., 2009). In particular, the estimation of coefficients on the optimal portfolios relies heavily on properly estimating the smallest eigenvalues of $\mathbf{\Sigma}$.

When a large amount of data is available ($N \gg p$), we expect that the entries in $\mathbf{S}$ will converge to the entries of $\mathbf{\Sigma}$, along with all meaningful statistics about $\mathbf{\Sigma}$ and its eigenvalues. On the other hand, what happens when N and p are both large, so that $p/N = \lambda \in (0,1)$? The simplest case is when the p assets are perfectly uncorrelated and each have unit variance. Surprisingly, the eigenvalues of $\mathbf{S}$ will follow a fixed distribution (when N and p are both large) which is parametrized only by λ. This is known as the Marčenko–Pastur (MP) distribution mp_λ with shape parameter $\lambda > 0$, and is given by

$$\mathrm{mp}_\lambda = \max\{0, 1 - 1/\lambda\}\,\delta_0 + \mathrm{mp}_\lambda^{\mathrm{abs}}$$

where δ_0 indicates the point mass at zero and $\mathrm{mp}^{\mathrm{abs}}$ is an absolutely continuous measure with density given by

$$\frac{d\mathrm{mp}_\lambda^{\mathrm{abs}}(x)}{dx} = \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{2\pi\lambda x}, \qquad \begin{array}{l} x \in [\lambda_-, \lambda_+] \\ \lambda_\pm = (1 \pm \sqrt{\lambda})^2 \end{array} \tag{2.6}$$

The precise description of the limiting statistics for matrices of the form $\mathbf{S}$ is described by the following theorem, originally proved by Marčenko and Pastur themselves.

Theorem 2.2.3 (Marčenko and Pastur, 1967; Silverstein, 1995)**.** *Let Y denote a random variable with mean zero and* $\operatorname{var}[Y] = 1$*. Let $\lambda \in (0, \infty)$ be a shape parameter, and suppose $p = p(N)$ is a function $p : \mathbb{N} \to \mathbb{N}$ such that $p(N)/N \to \lambda$ as $N \to \infty$. Let $\mathbf{Y}_N$ be a sequence of $N \times p$ matrices with i.i.d. entries, whose distributions are given by $[\mathbf{Y}_N]_{ij} \overset{d}{=} Y$ for all $N \in \mathbb{N}$. Finally, let*

$$\mathbf{S}_N = \frac{1}{N}\mathbf{Y}_N^\dagger \mathbf{Y}_N$$

denote the sample covariance matrix of $\mathbf{Y}_N$. Then as $N \to \infty$, almost surely we have

$$\mu_{\mathbf{S}_N} \xrightarrow{d} \mathrm{mp}_\lambda$$

where convergence is in the weak sense.

The utility of this theorem is that it describes the distribution of the eigenvalues of a matrix composed entirely of noise. If $\mathbf{X}$ has random entries that are all independent, centered, and have variance equal to some small $\epsilon > 0$, then their true covariance matrix is simply $\epsilon \mathbf{I}_p$, with p eigenvalues all equal to ϵ. The content of this theorem is that the histogram of eigenvalues of the sample covariance matrix $\mathbf{S}$ will be spread out to an interval $[\epsilon(1-\sqrt{\lambda})^2, \epsilon(1+\sqrt{\lambda})^2]$. This is used as evidence in statistical literature (Bickel and Levina, 2008; Paul and Aue, 2014) to support the idea that the estimation of eigenvalues of large sample matrices may be poor when p/N is not close to zero.

2.2.2 Generalized Marčenko–Pastur Theorem

If the p assets in question have a complicated covariance structure $\mathbf{\Sigma}$, but are all related to a family of distributions (e.g. normal), then we might suppose that $\mathbf{X} = \mathbf{Y}\mathbf{\Sigma}^{1/2}$ where the matrix $\mathbf{Y}$ has i.i.d. entries drawn from a fixed distribution as in Theorem 2.2.3. The sample covariance matrix is then of the form

$$\mathbf{S} = \mathbf{\Sigma}^{1/2}\mathbf{S}\mathbf{\Sigma}^{1/2}$$

which has the same eigenvalues as $\mathbf{S\Sigma}$. If the eigenvalues of $\mathbf{\Sigma}$ are deterministic or follow a fixed distribution, it may be possible to understand how the product $\mathbf{S\Sigma}$ distorts them, and a significant amount of literature has been devoted to this topic (Bai et al., 2014; Yao et al., 2015). We mention one key result.

Theorem 2.2.4 (Silverstein, 1995)**.** *Let $\mathbf{Y}_N$ be as in Theorem 2.2.3, and let $\mathbf{T}_N$ be a sequence of $p \times p$ independent Hermitian random matrices which are also independent from $\mathbf{Y}_N$. Suppose that the ESD of the sequence $\mathbf{T}_N$ converges in distribution almost surely to some nonnegative probability measure $\nu \in \mathcal{P}(\mathbb{R}^+)$. Let $\sqrt{\mathbf{T}_N}$ denote the nonnegative Hermitian square root of each $\mathbf{T}_N$. Then the ESD of the product*

$$\frac{1}{N}\sqrt{\mathbf{T}_N}\mathbf{Y}_N^\dagger\mathbf{Y}_N\sqrt{\mathbf{T}_N} \sim \frac{1}{N}\mathbf{Y}_N^\dagger\mathbf{Y}_N\mathbf{T}_N = \mathbf{S}_N\mathbf{T}_N$$

converges in distribution almost surely to a probability measure $\mu_{\lambda,\nu} \in \mathcal{P}(\mathbb{R}^+)$ whose Stieltjes transform $S_{\lambda,\nu} : \mathbb{C}^+ \to \overline{\mathbb{C}^+}$ satisfies the implicit equation

$$S_{\lambda,\nu}(z) = \int_{\mathbb{R}} \frac{1}{x\left(1 - \lambda - \lambda z S_{\lambda,\nu}(z)\right) - z} d\nu(x) \tag{2.7}$$

Chapter 3

Lévy Processes

3.1 Lévy–Khintchine Representation

Every distribution $\mu \in \mathcal{P}(\mathbb{R})$ generates a discrete semigroup of distributions μ^{*n} for $n \in \mathbb{N}$ through the additive convolution operation

$$\mu^{*n} = \underbrace{\mu * \mu * \ldots * \mu}_{n}$$

corresponding to powers of the characteristic function

$$\varphi_{\mu^{*n}}(z) = \varphi^n(z)$$

A natural question to ask is whether this semigroup can be made continuous in its parameter, such that $\varphi_\mu(z)^t$ represents the characteristic function of a distribution for all $t > 0$. This question is equivalent to asking if, for any $n \in \mathbb{N}$, there exists a distribution $\mu^{*1/n} \in \mathcal{P}(\mathbb{R})$

such that

$$\mu = \left(\mu^{*1/n}\right)^{*n} = \underbrace{\mu^{*1/n} * \mu^{*1/n} * \ldots * \mu^{*1/n}}_{n}$$

By appropriate continuity arguments, this ought to extend to all $t > 0$. This is possible in the case of, for instance, the normal or Poisson distributions, as evidenced by their characteristic functions. If the above condition is satisfied, or rather if $\varphi_\mu(z)^t$ represents a characteristic function for all $t > 0$, we say that μ is infinitely divisible, or of class ID($*$). In terms of a random variable X, there exists a random variable $X_{1/n}$ such that

$$X \stackrel{d}{=} Y_1 + Y_2 + \ldots + Y_n$$

where the random variables $Y_1, \ldots, Y_n$ are i.i.d. and $Y_j \stackrel{d}{=} X_{1/n}$. The distribution of $X_{1/n}$ (or $\mu^{*1/n}$) is called the n^{th} convolution root of X (of μ). By the Lévy–Khintchine theorem (see Zolotarev, 1986; Sato, 2013, among many others), there is a bijection between the collection ID($*$) and the collection of distributions of Lévy processes, which are defined as those stochastic cádlág processes with the following properties:

- $X_0 \stackrel{d}{=} 0$
- $X_{t_1} - X_{s_1}$ and $X_{t_2} - X_{s_2}$ are independent for all $0 \leq s_1 \leq t_1 \leq s_2 \leq t_2$.
- $X_t - X_s \stackrel{d}{=} X_{t-s}$ for all $0 \leq s \leq t$.

This correspondence is precisely that a random variable X is infinitely divisible if it follows the distribution $X \stackrel{d}{=} X_1$ of the unit time distribution of a Lévy process X_t. Under such circumstances, we say that the distribution of X drives the process X_t.

Some properties of infinitely divisible distributions are clear. For instance, since $\varphi_\mu(z)^t$ must qualify as a characteristic function for all $t > 0$, it is necessary that $\varphi_\mu(z) \neq 0$ for every

$z \in \mathbb{R}$. The question of a complete classification of infinitely divisible distributions, and thus also Lévy processes, is completely solved in the following theorem.

Theorem 3.1.1 (Lévy–Khintchine Decomposition, Sato, 2013)**.** *If X_t is a real-valued Lévy process, then there exists a unique triplet (μ, σ, Π) consisting of $\mu \in \mathbb{R}$, $\sigma \geq 0$, and a Borel measure Π on $\mathbb{R}$ with the properties*

(1)	$\Pi(\{0\}) = 0$	*No mass at zero.*
(2)	$\Pi((-\infty, -1] \cup [1, \infty)) < \infty$	*Integrable tails.*
(3)	$\int_{-1}^{1} x^2 \; d\Pi(x) < \infty$	*Controlled singularity at the origin.*

such that the CGF $\psi_{X_t}(z)$ *can be expressed as*

$$\frac{1}{t}\psi_{X_t}(z) = i\mu z - \frac{1}{2}\sigma^2 z^2 + \int_{\mathbb{R}} \left[e^{ixz} - 1 - ixz\mathbb{1}_{[-1,1]}(x) \right] d\Pi(x) \tag{3.1}$$

The triplet (μ, σ, Π) is called the Lévy triplet of the process X_t.

The term $ixz\mathbb{1}_{[-1,1]}(x)$ may also be replaced with $\frac{ixz}{1+x^2}$, leading to a different parametrization of the Lévy triplets, but uniqueness still holds. This form of the theorem is sometimes called a soft cutoff of the singular integral.

Corollary 3.1.2. *If X_t is a real-valued Lévy process with finite moments up to some order $n \in \mathbb{N}$, then*

$$\kappa_k[X_t] = t \int_{\mathbb{R}} x^k d\Pi(x) \tag{3.2}$$

for $3 \leq k \leq n$.

The following corollary follows directly from Theorem 3.1.1 and the uniqueness of the characteristic function.

Corollary 3.1.3. *Suppose $X_t^{(n)}$ is a sequence of real-valued Lévy processes which converges in distribution to a Lévy process X_t. This is to say, for any (and all) $t > 0$, we have $X_t^{(n)} \xrightarrow{d} X_t$. Then if $A \subseteq \mathbb{R}$ is any Borel set not containing a neighborhood of the origin,*

$$\Pi^{(n)}(A) \xrightarrow{n\to\infty} \Pi(A)$$

Example 3.1.4. Brownian motion B_t with drift $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ has a characteristic function given by

$$\varphi_{B_t}(z) = e^{i\mu tz - \sigma^2 tz^2/2}$$

This corresponds to the case where $\Pi(\mathbb{R}) = 0$.

The preceding theorem is often used to decompose the Lévy process X_t into the independent sum of other processes with desirable properties. The term $i\mu z$ corresponds to a deterministic process $X_t^{(1)} = \mu t$, while the term $-\frac{1}{2}\sigma^2 z^2$ corresponds to a scaled Brownian motion process $X_t^{(2)} = \sigma B_t$. The interpretation of Π is more nuanced, and is sometimes accomplished by considering the restriction of the measure to the sets $[-1, 1]$ and $\mathbb{R}\backslash[-1, 1]$ and then considering a small activity process $X_t^{(3)}$ and a large jump compound Poisson process $X_t^{(4)}$. Although such a decomposition is possible, Corollary 3.1.7 and Theorem 3.2.5 will provide alternatives in special circumstances that greatly simplify the problem.

Example 3.1.5. The standard Poisson process (with rate 1) is often denoted by N_t, and has characteristic function given by

$$\varphi_{N_t}(z) = e^{t(e^{iz}-1)}$$

From this it is clear that the distribution of N_t is Poisson with rate t. By scaling the parameter t, other Poisson processes can be created. The distribution of N_t at time $t > 0$ is described by a Poisson random variable with rate t, whose support is $\mathbb{N}$ with a discrete mass function given by $e^{-t}\frac{t^k}{k!}$ for $k \in \mathbb{N}$. This immediately demonstrates that a Lévy process can be both discrete and nonnegative.

Example 3.1.6. A compound Poisson process X_t with rate $r > 0$ and jump distribution $\nu \in \mathcal{P}(\mathbb{R})$ is the process given by

$$X_t \stackrel{d}{=} \sum_{j=0}^{N_{rt}} \xi_j$$

where N_t is a standard Poisson process, and ξ_j are i.i.d. random variables (independent from N_t) following the distribution ν. The characteristic function is given by

$$\varphi_{X_t}(z) = e^{rt(\varphi_\nu(z)-1)}$$

If ν has moments $m_n[\nu]$ up to some order, then X_t has cumulants given by $\kappa_n[X_t] = rtm_n[\nu]$. In particular, any sequence of moments of a distribution corresponds to a sequence of cumulants of a compound Poisson distribution.

Corollary 3.1.7. *Let X_t be a real-valued Lévy process, and suppose $\mathbb{E}[|X_1|] < \infty$ for some $t > 0$. Then the CGF $\psi_{X_t}(z)$ can be expressed in a modified form of (3.1) given by*

$$\frac{1}{t}\psi_{X_t}(z) = i\mu z - \frac{1}{2}\sigma^2 z^2 + \int_{\mathbb{R}} \left[e^{ixz} - 1 - ixz\right] d\Pi(x) \tag{3.3}$$

The parameters σ and Π are identical to those found in Theorem 3.1.1, however the value of μ may be different. Furthermore, $\mathbb{E}[X_t] = t\mu$. Under this decomposition, X_t can be realized *as the sum of independent processes*

$$X_t \stackrel{d}{=} \mu t + \sigma B_t + X'_t$$

where B_t is standard Brownian motion, and X'_t is a Lévy process independent from B_t with zero mean.

3.2 Further Classification of Lévy Processes

3.2.1 Existence of Moments and Cumulants

Definition 3.2.1. A function $g : \mathbb{R} \to \mathbb{R}^+$ is called submultiplicative if there exists a constant $a > 0$ such that

$$g(x+y) \leq a \cdot g(x)g(y)$$

for any $x, y \in \mathbb{R}$. We say that g is locally bounded if it is bounded on any compact subset of $\mathbb{R}$.

Examples of submultiplicative functions include

$$\max\{|x|^\alpha, 1\} \qquad e^{c|x|^\beta} \qquad \log(\max\{|x|, e\})$$

for $\alpha, c > 0$ and $0 < \beta \leq 1$. The products of submultiplicative functions remain submultiplicative.

Theorem 3.2.2 (Sato, 2013)**.** *Let g be submultiplicative, locally bounded, and Lebesgue measurable, and let X_t be a real-valued Lévy process. Then $\mathbb{E}[g(X_t)] < \infty$ for any $t > 0$ if* *and only if*

$$\int_{\mathbb{R}\setminus[-1,1]} g(x)\ d\Pi(x) < \infty$$

This is to say, the tail behavior of X_t is determined by the tail behavior of Π.

Corollary 3.2.3. *Let X_t be a real-valued Lévy process. Then X_t has moments and cumulants up to some order n if and only if*

$$\int_{\mathbb{R}\setminus[-1,1]} |x|^n d\Pi(x) < \infty$$

3.2.2 Variation, Activity, and Subordination

Following Sato (2013), we classify Lévy processes in the following way.

Definition 3.2.4. Let X_t be a real-valued Lévy process with generating triple (μ, σ, Π) as in (3.1). Then we say that X_t is of

type (A), of finite activity	if $\sigma = 0$ and $\Pi(\mathbb{R}) < \infty$
type (B), of infinite activity	if $\sigma = 0$, $\Pi(\mathbb{R}) = \infty$, and $\int_{-1}^{1} |x| d\Pi(x) < \infty$
type (C), of infinity variation	if $\sigma > 0$ or $\int_{-1}^{1} |x| d\Pi(x) = \infty$

A process of type (A) is one for which there is a nonzero probability of no activity, or in other words that $\mathbb{P}[X_t = 0] > 0$ for any $t > 0$. As we will see shortly, any such process can be expressed as a compensated compound Poisson process. A process of type (B) has infinite activity, but is of bounded variation almost surely. The class of subordinators, Lévy processes whose distributions are nonnegative for some (equivalently all) $t > 0$ are necessarily of type (B). Finally, processes of type (C) are almost surely of unbounded variation.

Theorem 3.2.5. *Let X_t be a real-valued Lévy process of type* (A) *or* (B). *Then the* CGF

$\psi_{X_t}(z)$ can be expressed in a modified form of (3.1) given by

$$\frac{1}{t}\psi_{X_t}(z) = i\mu z + \int_{\mathbb{R}} \left[e^{ixz} - 1\right] d\Pi(x) \tag{3.4}$$

The Lévy measure Π is identical to that found in Theorem 3.1.1, however the value of μ may be different.

Example 3.2.6. Consider the Lévy process X_t with $\mu = \sigma = 0$ and Lévy measure $\Pi \in \mathcal{P}(\mathbb{R})$, a probability measure such that $\Pi(\{0\}) = 0$. Then using form (3.4), the process X_t has characteristic function given by

$$\varphi_{X_t}(z) = e^{t(\varphi_\Pi(z)-1)}$$

and is an example of a compound Poisson process

$$X_t \stackrel{d}{=} \sum_{j=0}^{N_t} \xi_j$$

where ξ_j are i.i.d. random variables following the distribution Π, and N_t is an independent Poisson process with rate 1.

Following Examples 3.1.6 and 3.2.6, and Theorem 3.2.5, we have the following result.

Corollary 3.2.7. *If X_t is a real-valued Lévy process of type* (A)*, then it can be expressed as the sum of a deterministic component μt, called the compensation or drift, and a compound Poisson process with rate $r = \Pi(\mathbb{R})$ and jump distribution $r^{-1}\Pi \in \mathcal{P}(\mathbb{R})$.*

We note that the Poisson process N_t described in Example 3.1.5 is nonnegative for all values of $t > 0$. Processes of this type are called subordinators, and have a fairly straightforward classification. First, it is necessary for the Lévy measure Π to be concentrated on $\mathbb{R}^+$, for the drift to be nonnegative, and for the process to have finite variation (of type (A) or (B)).

Theorem 3.2.8 (Sato, 2013)**.** *Let X_t be a real-valued Lévy process. The following are equivalent:*

- $X_t \geq 0$ *for some* $t > 0$.
- $X_t \geq 0$ *for all* $t > 0$.
- X_t *is of type (A) or (B), and in the form (3.4) the measure* Π *is concentrated on* $\mathbb{R}^+$ *and* $\mu \geq 0$.

From here we use the term subordinator to refer to such a nonnegative Lévy process.

Theorem 3.2.9 (Sato, 2013)**.** *Suppose X_t is any Lévy process, and τ_t is a subordinator independent from X_t. Then X_{τ_t} is also a real-valued Lévy process.*

The act of performing the composition X_{τ_t} is called subordinating the process X_t to τ_t. If Brownian motion is subordinated, then the scale invariance of the process implies that

$$B_{\tau_t} \stackrel{d}{=} \sqrt{\tau_t} B_1$$

where B_1 follows the distribution for standard normal distribution. Consequently, the class of subordinated Brownian motion processes corresponds to normal scale mixtures with square roots of subordinators.

Every Lévy process, being a semi-martingale, has a well defined quadratic variation process $[X]_t$ given by the convergence in probability of the sample quadratic variation process, even when the process does not have a well defined variance (Pascucci, 2011). The quadratic variation $[B]_t$ of standard Brownian motion B_t, being a continuous semi-martingale, is equal to its predictable quadratic variation, namely $[B]_t = t$. On the other hand, the quadratic variation of a standard Poisson process N_t is known to be itself, that is $[N]_t = N_t$, owing to the fact that the the jumps of N_t are almost surely separated in time.

The quadratic variation of a compound Poisson process

$$X_t \stackrel{d}{=} \sum_{j=0}^{N_{rt}} \xi_j$$

where $\mathcal{L}(\xi_j) = \nu$ can be computed in a similar manner due to the separation of jumps, and is equal to

$$[X]_t = \sum_{j=0}^{N_{rt}} \xi_j^2$$

which corresponds to a Lévy process with measure $r \cdot \nu^2$. If $d\Pi(x)$ is a continuous density for the Lévy measure of X_t, then it follows that $[X]_t$ will have a Lévy measure with continuous density $d\widetilde{\Pi}(x)$ that is zero for $x \leq 0$ and

$$d\widetilde{\Pi}(x) = \frac{d\Pi(\sqrt{x}) + d\Pi(-\sqrt{x})}{2\sqrt{x}}$$

for $x > 0$. The approximation of Lévy processes by compound Poisson processes implies a bijection between all symmetric Lévy processes and all subordinators given by passing from X_t to $[X]_t$.

3.2.3 Self-Decomposability

A random variables X is said to be self-decomposable (SD) if for any $0 < b < 1$, there exists some random variable Y_b such that, if X and Y_b are independent, then

$$X \stackrel{d}{=} bX + Y_b \tag{3.5}$$

All self-decomposable distributions are infinitely divisible, so we have the nested classes

$$\mathrm{SD} \subset \mathrm{ID}(*)$$

Self-decomposable distributions are precisely those that appear in scaled limits theorems. If X_j are a sequence of independent (but not necessarily identically distribution) random variables and a_j, b_j are normalizing constants, and we have that

$$a_j \left(\sum_{j=1}^{n} X_j \right) + b_j \xrightarrow{d} X$$

then the distribution of X is self-decomposable (Sato, 2013). This makes SD distributions appropriate for situations where a random variable is composed of a linear combination of an unknown number of subcomponents, possibly of various sizes. We note that (3.5) implies that SD distributions are precisely those which appear as marginals in first-order autoregressive equations of the form

$$X_{n+1} = bX_n + \epsilon_n$$

where $0 < b < 1$ and ϵ_n is a sequence of i.i.d. random variables, as observed in Gaver and Lewis (1980). For these reasons, many applications choose models whose distributions are explicitly self-decomposable, as opposed to simply infinitely divisible (Carr et al., 2002).

3.2.4 Generalized Gamma Convolutions and Related Classes

While many distributions, such as Poisson and Gamma, can be seen to be infinitely divisible directly from the form of their characteristic functions, others have been historically much more difficult to prove. In 1977, Thorin introduced a class of Lévy processes for the purpose of proving the infinite divisibility of the Lognormal and other distributions (Thorin, 1977a,b). We introduce this and related classes by means of continuous Lévy measures Π as follows. We say that an infinitely differentiable function $g : (0, \infty) \to \mathbb{R}$ is completely monotone (CM) if $(-1)^n g^{(n)}(x) \geq 0$ for all $x > 0$ and $n \in \mathbb{N}$. Now suppose a Lévy process X_t has Lévy measure Π with density $d\Pi(x) = \mathbb{1}_{(-\infty,0)}(x)\rho_-(-x) + \mathbb{1}_{(0,\infty)}(x)\rho_+(x)$, where $\rho_\pm : \mathbb{R}^+ \to \mathbb{R}^+$ is some pair of functions. Then we say that X_t is in the (Andersen et al., 2015)

- type-G class $G(\mathbb{R})$ when $\rho_\pm(x) = g_\pm(x^2)$ and $g_\pm$ are completely monotone.
- Aoyama class $M(\mathbb{R})$ when $\rho_\pm(x) = |x|^{-1} g_\pm(x^2)$ and $g_\pm$ are completely monotone.
- Thorin class $T(\mathbb{R})$ when $\rho_\pm(x) = |x|^{-1} g_\pm(x)$ and $g_\pm$ are completely monotone.

In each class, if ρ_- is identically zero and the process is a subordinator, then we say that it lies in the subclass $G(\mathbb{R}^+)$, $M(\mathbb{R}^+)$, and $T(\mathbb{R}^+)$, respectively. Elements of the Thorin class $T(\mathbb{R}^+)$ are called generalized gamma convolutions, or GGCs. GGC distributions arise as the weak closure of discrete convolutions of independent Gamma random variables. The larger class $T(\mathbb{R})$ is similarly called extended generalized gamma convolutions, or EGGCs.

The symmetric elements of EGGC and $G(\mathbb{R})$ have particularly convenient interpretations in terms of subordinated Brownian motion processes. A process X_t with a distribution in EGGC$\cap\mathcal{P}_s(\mathbb{R})$ can be realized as B_{τ_t} where τ_t is itself a GGC subordinator (Bondesson, 1992). Similarly, the distributions in $G(\mathbb{R}) \cap \mathcal{P}_s(\mathbb{R})$ can be realized as B_{τ_t} where τ_t is simply any subordinator. In this way, the wider class $G(\mathbb{R})$ can be imagined as "skewed" subordinated Brownian motion processes.

The Aoyama class is of particular interest for the reason that X_t is a process in $M(\mathbb{R})$ if and only if $[X]_t$ is a GGC process. Basic facts about completely monotone functions implies immediately that

$$\text{GGC} = T(\mathbb{R}) \subset M(\mathbb{R}) \subset G(\mathbb{R}) \cap \text{SD} \subset \text{ID}(*)$$

$$\text{EGGC} = T(\mathbb{R}^+) \subset M(\mathbb{R}^+) \subset G(\mathbb{R}^+) \cap \text{SD} \subset \text{ID}(*) \cap \mathcal{P}(\mathbb{R}^+)$$

Gaussian, α-stable, log-normal, Student's-t, Pareto, gamma, χ^2, and generalized inverse Gaussian can all be shown to be GGC or EGGC, thus making them ID($*$). Many popular distributions suggested for the modeling of asset returns are EGGC, including the variance-gamma (VG) model of Madan and Seneta (1990), the normal-inverse Gaussian (NIG) model of Barndorff-Nielsen (1997), and tempered stable distributions and CGMY model of Carr et al. (2002). Other odd distributions lie in the EGGC class, such as the logarithm of a Gamma random variable (which shows that the generalized Logistic distribution is also in the class).

The classes of GGC and EGGC processes were studied extensively in Bondesson (1992), which remains the definitive text on the subject. The most significant contribution to the subject since its publication is Bondesson (2015), in which the startling result in Theorem 3.2.14 was proved. Before stating it in full, we motivate the significance of the result by introducing some additional descriptive characteristics of GGC processes.

Definition 3.2.10. A nonnegative function $f : \mathbb{R}^+ \to \mathbb{R}^+$ is called hyperbolically completely monotone (HCM) if, for each $u > 0$, the function

$$h(w) = f(uv)f(u/v)$$

is a completely monotone function when viewed as a function of the auxiliary variable $w = v+1/v$. A distribution in $\mathcal{P}(\mathbb{R}^+)$ is said to be in the class HCM if it has a continuous density on $(0,\infty)$ which is a HCM function.

Example 3.2.11. The functions x^β, $e^{-\gamma x}$, and $\frac{1}{(1+x)^\gamma}$ for $\beta \in \mathbb{R}$ and $\gamma > 0$ are HCM, as is the density of Lognormal distributions with arbitrary parameters. The class of HCM functions is closed under pointwise multiplication of functions, and with respect to pointwise limits of functions. It is also closed under the operations $f(x^\alpha)$ and $f^p(x)$ for $|\alpha| \leq 1$ and $p > 0$. The following results imply that Lognormal, Beta distributions of the second kind, and powers of Gamma distributions, and generalized inverse Gaussian distributions are all HCM and thus also ID($*$). The positive α-stable distributions (as discussed in Section 3.3.1) were known to be HCM for $\alpha = 1/n$ with $n \in \mathbb{N}$ and $n \geq 2$, and explicitly not HCM for $1/2 < \alpha < 1$. Bondesson (1992) conjectured that this would extend to all $0 < \alpha < 1/2$, but was unable to prove it. No significant progress was made until Fourati (2013) showed that the conjecture is true at least for $\alpha \in (0, 1/4] \cup [1/3, 1/2]$. The issue was finally resolved only a few years ago in Bosch and Simon (2016), where it was shown that the conjecture is true for all $0 < \alpha \leq 1/2$.

Theorem 3.2.12 (Bondesson, 1992)**.** *The class of* GGC *distributions are precisely those nonnegative distributions whose Laplace transforms* $\phi_\mu(s)$ *are* HCM *as functions on* $(0,\infty)$. *On the other hand, the distributions in the class* HCM *are those nonnegative distributions whose support is* $[0,\infty)$ *and whose Laplace transforms* $(-1)^n \phi_\mu^{(n)}(s)$ *are* HCM *functions on* $(0,\infty)$ *for all* $n \in \mathbb{N}$. *Consequently,*

$$\text{HCM} \subset \text{GGC} \subset \text{SD} \subset \text{ID}(*)$$

and the inclusions are strict.

Theorem 3.2.13 (Bondesson, 1992)**.** *Suppose* X *and* Y *are independent random variables with* HCM *distributions. Then* X^q, XY, *and* X/Y *have* HCM *distributions for* $|q| \geq 1$.

Furthermore, if W is an independent GGC *random variable, then WX and W/X are both* GGC*.*

The preceding theorems show that the HCM class has many desirable properties, and interacts nicely with the class GGC. However, GGC is closed under convolution and convolution roots, so that a Lévy process X_t has a GGC density for some $t > 0$ if and only if the property holds for all $t > 0$. In contrast, HCM is not closed under convolution roots, and so the property of being in the class is time dependent. The following recent results were therefore quite shocking when published, as they demonstrate that the GGC class still maintains many of the properties once only known for HCM distributions.

Theorem 3.2.14 (Bondesson, 2015)**.** *Suppose X and Y are independent* GGC *random variables. Then XY is a* GGC *random variables. Consequently,*

$$\mathrm{GGC} \subset \mathrm{ID}(*) \subset \mathcal{P}(\mathbb{R}^+)$$

makes GGC *into a proper subclass of the nonnegative distributions which is closed under independent sums, products, weak limits, convolution roots, and contains the Gamma distributions.*

Corollary 3.2.15 (Bondesson, 2015)**.** *Suppose X and Y are independent symmetric* EGGC *random variables, then XY is a symmetric* EGGC *random variables. Consequently*

$$\mathrm{EGGC} \cap \mathcal{P}_s(\mathbb{R}) \subset \mathrm{ID}(*) \cap \mathcal{P}_s(\mathbb{R}) \subset \mathcal{P}_s(\mathbb{R})$$

is a proper subclass of the symmetric distributions which is closed under independent sums, products, weak limits, convolution roots, and contains the normal distributions.

As noted by Bondesson, these are actually the only known nontrivial subclasses of the two collections $\mathcal{P}(\mathbb{R}^+)$ and $\mathcal{P}_s(\mathbb{R})$ which are closed under independent sums, products, and weak limits.

3.2.5 Mixtures of Exponential Distributions

Mixing models arise frequently in probability, and typically take the form of a independent product XY or quotient X/Y or random variables. Here, X represents the distribution for whom a scale parameter is unknown, while Y contains information about the distribution of this parameter. A natural question to ask is whether or not there exists particular class of distributions for X such that XY or X/Y is guaranteed to be $\mathrm{ID}(*)$? Surprisingly, one such distribution is $\mathrm{Exp}(1)$, the exponential distribution. An analogous result exists in the context of free probability, although the distribution taking on the role of the exponential distribution is quite surprising. This issue is addressed in Example 5.3.6.

Definition 3.2.16. We say that a distribution is a mixture of exponential distributions, or in the class MED, if its law can be expressed in the form E/Y where $E \stackrel{d}{=} \mathrm{Exp}(1)$ follows a standard exponential and $Y \in \mathcal{P}(\mathbb{R}^+)$.

Clearly we can also express such laws as EY rather than E/Y, although classically it has been written in this form so that the distribution follows $\mathrm{Exp}(Y)$, where Y is the parameter which appears in the description of the exponential family.

Theorem 3.2.17 (Bondesson, 1992)**.** *All* MED *distributions are in* $\mathrm{ID}(*)$*, and their convolution roots are also* MED*. Furthermore, if X has an* MED *distribution and $q \geq 1$, then X^q and $e^X - 1$ are both* MED*.*

Some connections exist between the classes MED and GGC, although neither class properly contains the other. One such connection is given below.

Lemma 3.2.18. *Suppose $E \stackrel{d}{=} \mathrm{Exp}(1)$ is exponentially distributed. Then for any $q \geq 2$, there* *exists a* HCM *random variable* F_q *independent from* E *such that the Weibull distribution* $E^q \overset{d}{=} \text{Weib}(1/q)$ *can be written as* $E^q \overset{d}{=} EF_q$. *Consequently,* E^q *is* GGC.

Proof. The decomposition $E^q \overset{d}{=} EF_q$ for some independent random variable $F_q \geq 0$ follows from (Bondesson, 1992, Example 4.3.4), however it is not clear that F_q must be HCM. To show this, we notice that

$$\log(E^q) = q\log(E) \overset{d}{=} \log(E) + \log(F_q)$$

Let these random variables be denoted by $Y_1 \overset{d}{=} Y_2 + Y_3$. Then the Laplace transforms of Y_1 and Y_2 are known explicitly in terms of the gamma function $\phi_1(s) = \Gamma(1 - qs)$, $\phi_2(s) = \Gamma(1 - s)$, when $\text{Im}[s] > 0$. Therefore, the Laplace transform $\phi_3(s)$ of Y_3, if it exists for arguments $\text{Im}[s] > 0$, must be equal to

$$\phi_3(s) = \phi_1(s)/\phi_2(s) = \frac{\Gamma(1 - qs)}{\Gamma(1 - s)}$$

If such a function represents the Laplace transform of a distribution Y_3 such that e^{Y_3} is a GGC function, then the claim is proved. In fact, this is precisely the Laplace transform of $\log(1/X_\alpha)$, where X_α is a nonnegative α-stable random variable where $\alpha = 1/q$ (Bondesson, 1992, Example 7.2.3). Therefore, we can write

$$E^q \overset{d}{=} \frac{E}{X_\alpha}$$

Since the nonnegative α-stable distributions are HCM for $1/\alpha \geq 2$ (Bosch and Simon, 2016), the random variable $F_q = 1/X_\alpha$ is HCM, hence GGC. ■

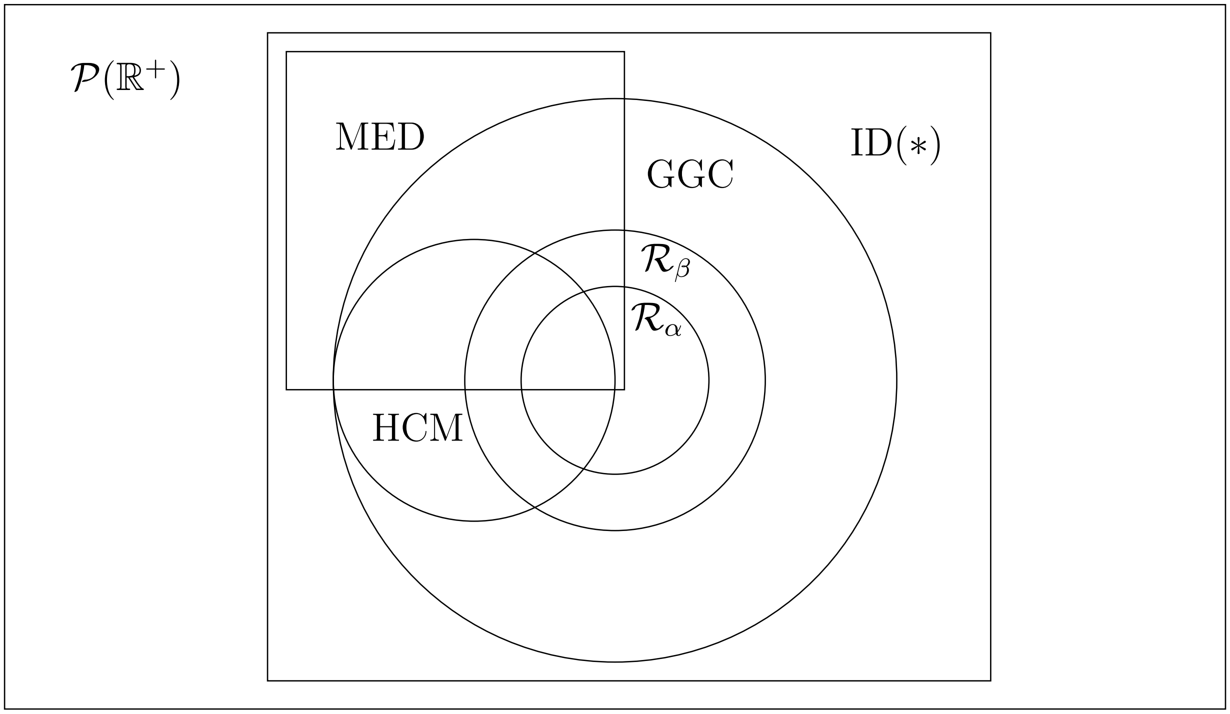


Figure 3.1: Diagram representing relationships between the classes of nonnegative distributions discussed.

3.2.6 Euler Transforms of HCM

Bondesson (1992) also introduced the classes $\mathcal{R}_\beta \subseteq$ GGC for $\beta > 0$ of distributions which are expressible as independent quotients of the form X/Y_β, where $X \in$ GGC and Y_β follows a Gamma distribution Γ_β. It is clear from Theorem 3.2.13 that these are subclasses of GGC. The classes have the following classification.

Theorem 3.2.19 (Bondesson, 1992)**.** *Let $0 < \alpha < \beta$, then $\mathcal{R}_\alpha \subseteq \mathcal{R}_\beta$. The weak closure of*

$$\bigcup_{\beta > 0} \mathcal{R}_\beta$$

is the class GGC*. Furthermore, a distribution is in $\mathcal{R}_\beta$ if and only if it has a continuous density f whose Euler transform $F^{[\beta]}$ of order β, defined as*

$$F^{[\beta]}(x) = \frac{1}{\Gamma(\beta)} \int_0^x (x-y)^{\beta-1} f(y) dy$$

is an HCM *function.*

Consequently, the class $\mathcal{R}_1$ is the class of densities on $[0, \infty)$ whose cumulative distribution functions F are HCM functions. The following result follows directly from the fact that the pointwise product $f_1(x)f_2(x)$ and powers $f_1(x^\alpha)$ for $|\alpha| \leq 1$ of HCM functions are HCM.

Theorem 3.2.20 (Bondesson, 1992)**.** *If $X, Y \in \mathcal{R}_1$ are independent random variables, then X^q for $|q| \geq 1$ and $\max\{X, Y\}$ are also in the class $\mathcal{R}_1$.*

By the form of its cumulative distribution function, it is clear that if X is exponentially distributed then it is not in the class $\mathcal{R}_1$, but $1/X$ is. If X follows a Gamma distribution Γ_t for $0 < t < 1$ (see Section 3.3.2), then X can be written as an MED in the following way

$$X \stackrel{d}{=} E\left(1 + \frac{Y_1}{Y_2}\right)^{-1}$$

where $Y_1 \stackrel{d}{=} \Gamma_{1-t}$ and $Y_2 \stackrel{d}{=} \Gamma_t$ are independent from $E \stackrel{d}{=} \mathrm{Exp}(1)$. Since $1 + \frac{Y_1}{Y_2}$ is GGC, it follows that $1/X$ is in the class $\mathcal{R}_1$. This also follows directly from Theorem 3.2.19.

3.2.7 Complex Generalized Gamma Convolutions

Although not relevant to financial applications, complex-valued random variables X which are circularly symmetric appear in similar applications of the M–P law in signal processing. In particular, Multiple-Input Multiple-Output (MIMO) signal processing techniques employed by Wi-Fi and 4G LTE cellular networks have used underlying random models for cognitive radio networks in order to estimate spectral statistics. The statistics in question, such as the instantaneous mutual information and ergodic capacity, can be analyzed through these models with the help of the M–P law (see Couillet and Debbah, 2011; Bai et al., 2014). Such random variables can be expressed as mixtures of the type $X = |X|e^{i\theta}$, where $|X| \geq 0$

and θ is an independent random variable which is uniformly distributed on the interval $[0, 2\pi]$. Channel models typically construct random variables in this fashion. Unfortunately, these mixtures do not preserve the underlying features of the random variable $|X|$, so X need not be SD or even ID($*$) even when $|X|$ is. A more convenient class of circularly symmetric random variables is that of mixtures of complex Gaussian distributions, which take the form

$$\mathrm{CN}(0, Y) \stackrel{d}{=} \sqrt{Y}Z$$

where Y is a nonnegative random variable (independent from the Gaussian distribution considered). As standard complex Gaussian random variables are already of the form $Z \stackrel{d}{=} \sqrt{E}e^{i\theta}$ for $E \in \mathrm{Exp}(1)$ and θ as above, it follows that all mixtures of complex Gaussian distributions can be written as

$$\mathcal{CN}(0, Y) \stackrel{d}{=} \sqrt{YE}e^{i\theta}$$

where Y, E, and θ are independent. It is clear that all mixtures of complex Gaussian distributions are circularly symmetric. Furthermore, it is easy to see that a circularly symmetric distribution is a mixture of complex-Gaussian distributions if and only if its envelope $|X|$ is an MED.

Definition 3.2.21. We say that a random variable X is a circular generalized gamma convolution (CGGC) if its distribution is of the form

$$X \stackrel{d}{=} \sqrt{\tau}Z \stackrel{d}{=} \mathcal{CN}(0, \tau)$$

where $\tau \in \mathrm{GGC}$ and $Z \in \mathrm{CN}(0, 1)$ are independent.

Alternatively, if B_t is a complex Brownian motion process with $B_t \stackrel{d}{=} \mathrm{CN}(0, t)$ for fixed $t > 0$, then X is a CGGC if it can be written as B_t subordinated by a GGC process τ_t at a fixed time $t = 1$, explicitly given as

$$X \stackrel{d}{=} B_{\tau_1}$$

Consequently, every CGGC distribution X is ID($*$).

Lemma 3.2.22. *A random variable X is a* CGGC *if and only if it is circularly symmetric and $|X|^2$ can be written as the independent product $|X|^2 \stackrel{d}{=} E\tau$ of an exponential random variable E and a* GGC *τ.*

Proof. This follows immediately from the fact that X is CGGC if and only if it takes the form

$$X \stackrel{d}{=} \sqrt{\tau} Z \stackrel{d}{=} \sqrt{\tau E} e^{i\theta} \stackrel{d}{=} |X| e^{i\theta}$$

where $E \stackrel{d}{=} \text{Exp}(1)$, so $|X|^2$ is an independent product of an exponential random variable and a GGC. ■

The class CGGC can be seen as the circularly symmetric analogue to the the (real-valued) EGGC class. As the following proposition shows, it has the significant advantage of preserving self-decomposability under the operations of independent addition and multiplication. Following the recent observations by Bondesson (2015), it is likely the only known nontrivial class of circularly symmetric ID($*$) distributions which is closed under these operations.

Theorem 3.2.23. *Let X and Y be independent* CGGC *random variables, then*

- $X \in \text{SD} \subset \text{ID}(*)$
- $X + Y \in \text{CGGC}$
- $XY \in \text{CGGC}$

Proof. Let $X \overset{d}{=} \sqrt{\tau_1} Z_1$ and $Y \overset{d}{=} \sqrt{\tau_2} Z_2$, where $\tau_i \in \mathrm{GGC}$ and $Z_i \in \mathrm{CN}(0,1)$, all independent. Since $\tau_1 \in \mathrm{GGC}$ we know that it is also SD. Therefore, for any $0 < b < 1$ we can find some independent Y_{b^2} such that

$$\tau_1 \overset{d}{=} b^2 \tau_1 + Y_{b^2}$$

Then it follows that

$$X \overset{d}{=} \sqrt{\tau_1} Z_1 \overset{d}{=} b\sqrt{\tau_1} Z_1 + \sqrt{Y_{b^2}} Z_2 \overset{d}{=} bX + \sqrt{Y_{b^2}} Z_2$$

which shows that $X \in \mathrm{SD}$.

By the marginals on τ_i (Prop. 4.3 and 4.4, Rakvongthai et al., 2010), $X+Y$ follows the distribution $\sqrt{\tau_1 + \tau_2} Z_1$. Since GGC is closed under additive convolutions, $\tau_1 + \tau_2 \in \mathrm{GGC}$, so $X + Y \in \mathrm{CGGC}$.

Now write $Z_1 \overset{d}{=} \sqrt{E} e^{i\theta}$ where $E \in \mathrm{Exp}(1)$ and θ is uniform on $[0, 2\pi]$, with E and θ independent. Then

$$XY \overset{d}{=} \sqrt{\tau_1 \tau_2 E} e^{i\theta} Z_2 \overset{d}{=} \sqrt{\tau_1 \tau_2 E} Z_2$$

By Theorem 3.2.14, $\tau_1 \tau_2 E \in \mathrm{GGC}$, and so $XY \in \mathrm{CGGC}$. ∎

Theorem 3.2.24. *If $X \in \mathrm{CGGC}$, then by Lemma 3.2.22 let $|X|^2 \overset{d}{=} \tau E$ where $\tau \in \mathrm{GGC}$ and $E \in \mathrm{Exp}(1)$. Suppose that we additionally have that either τ is equal to a nonzero constant or $\tau \in \mathrm{HCM}$. Then for any $p, q \in \mathbb{N}$ with $p \neq q$ and any $r \geq 2$, the random variables*

$$X^p (X^*)^q,\ |X|^r e^{i \cdot \arg(X)} \in \mathrm{CGGC}$$

where $\arg(X)$ is the complex angle of X.

Proof. The conditions on p and q ensure that if $X \overset{d}{=} |X|e^{i\theta}$ then

$$X^p (X^*)^q \overset{d}{=} |X|^{p+q} e^{i\theta}$$

with $p + q \geq 1$. We then have $|X|^{p+q} = \sqrt{\tau^{p+q} E^{p+q}}$. If τ is a nonzero constant, then τ^{p+q} is also a nonzero constant and thus GGC. Otherwise $\tau \in \text{HCM}$, so it follows by Theorem 3.2.13 that $\tau^{p+q} \in \text{HCM} \subset \text{GGC}$. By Lemma 3.2.18, $E^{p+q} \overset{d}{=} EF_{p+q}$ for an independent GGC random variable F_{p+q} (if $p + q = 1$, we simply take F_1 as δ_1). We can then write

$$|X|^{p+q} e^{i\theta} \overset{d}{=} \sqrt{\tau^{p+q} F_{p+q} E}\, e^{i\theta}$$

which is CGGC by Lemma 3.2.22.

Since X is circularly symmetric, $\arg(X)$ is uniform on $[0, 2\pi]$ and so $|X|^r e^{i \cdot \arg(X)}$ is circularly symmetric as well. Then we can write

$$|X|^r \overset{d}{=} \sqrt{\tau^r E^r} \overset{d}{=} \sqrt{\tau^r F_r E} \in \text{GGC}$$

by Lemma 3.2.18 once again. ■

Example 3.2.25. The Weibull distribution Weib(β) for $\beta > 0$ is defined as the distribution governing the shape of $E^{1/\beta}$ where $E \in \text{Exp}(1)$ is exponentially distributed. Its density is given by

$$f(x) = \beta x^{\beta-1} e^{-x^\beta}, \quad x > 0 \tag{3.6}$$

For $0 < \beta < 1$, the Weibull has heavier tails than the standard exponential distribution while still having well defined moments of all orders, and has been successful in describing some mobile fading scenarios (Parsons, 2000). The circularly symmetric random variable with Weibull fading envelope $|X| \stackrel{d}{=} \mathrm{Weib}(2/q)$ for $q = 1$ and $q \geq 2$ is CGGC, since it can written in the form

$$X = |X|e^{i\theta} \stackrel{d}{=} \sqrt{E^q} e^{qi\theta} \stackrel{d}{=} Z^q$$

where $Z \in \mathrm{CN}(0,1)$. This includes the case of Rayleigh ($q = 1$) and exponential ($q = 2$) fading. It also includes all powers of normal random variables, since if $Z \stackrel{d}{=} \mathrm{CN}(0,1)$, then $|Z^n| \stackrel{d}{=} \mathrm{Weib}(2/n)$ for $n \in \mathbb{N}$.

Example 3.2.26. The Suzuki (1977) distribution is defined as a mixture of Rayleigh and Lognormal distributions. Its density does not have a closed form, but can be calculated from the expression

$$f(x) = \int_0^\infty \frac{x}{y^2} e^{-x^2/2y^2} \frac{\sqrt{2}}{\sqrt{\pi} y \sigma} e^{-2(\log(y)-\mu/2)^2/\sigma^2} dy$$

Here we use a slightly different parametrization for μ and σ in order to make some computations more convenient. If $E \stackrel{d}{=} \mathrm{Exp}(1)$ and $L \stackrel{d}{=} \mathrm{LogN}(\mu, \sigma^2)$ are independent, then this distribution matches that of $\sqrt{EL}$, where $\sqrt{E}$ is once again understood to be Rayleigh distributed. No scale parameter for the Rayleigh distribution is necessary, as it can be included in the Lognormal parameter μ. A circularly symmetric distribution with Suzuki envelope $|X|$ is easily seen to be CGGC, since Lognormal distributions and their powers (which are also Lognormal) are GGC.

3.2.8 Essentially Bounded Processes

In the categorization of Lévy processes, compensated compound Poisson processes as in Corollary 3.2.7 are described by Lévy measures Π which have a controlled singularity at the origin, namely one which is integrable. This includes those measures Π for which there is some $\epsilon > 0$ such that $\Pi((-\epsilon, \epsilon)) = 0$, so that there are no small jumps appearing along the path. One might be curious what type of distributions can be generated when the opposite restriction is imposed, and we instead consider a process X_t where no large jumps are allowed to occur. This leads to the definition of *essentially bounded* Lévy processes, or those such that the support of the measure Π is contained in some compact interval $[-B, B]$ with $B > 0$. It turns out, as is shown in Lemma 3.2.28 below, that this is equivalent to the condition that the cumulants $\kappa_n[X_1]$ exist and grow no faster than $O(B^n)$. Such a condition guarantees that essentially bounded processes have exponential moments of all orders, that is

$$\mathbb{E}\left[e^{m|X|}\right] < \infty, \qquad m > 0$$

This implies that, although essentially bounded Lévy processes do not have bounded tails (this is impossible for an ID($*$) distribution), their long-term tail behavior is quite dampened.

Essentially bounded processes provide a modeling approach, in the vein of Mantegna and Stanley, to the problem of "ultraslow" convergence of i.i.d. sums of random variables in the central limit theorem. In Mantegna and Stanley (1995), it was famously observed that scaling in the Standard and Poor's 500 index failed to exhibit heavy-tailed behavior for extreme outliers. As a example of a distribution with such properties, they defined their truncated Lévy flight in Mantegna and Stanley (1994) in terms of the α-stable Lévy distribution, whose density is restricted to a large bounded set, although such a distribution fails to represent a Lévy process. The phrase "truncated Lévy flight" was later adopted by Koponen (1995), whose distributions were used as the basis for the CGMY model of Carr et al. (2002). The class of essentially bounded processes follows a similar modification of Lévy flight, as the condition that Π has bounded support still produces a Lévy process that can have arbitrarily large kurtosis, while being more convenient for the purpose of density estimation.

Definition 3.2.27. We say that a real-valued Lévy process X_t is essentially bounded by

$B > 0$ if the support of its Lévy measure Π lies inside $[-B, B]$, such that

$$\Pi((-\infty, -B) \cup (B, \infty)) = 0$$

Lemma 3.2.28. *X_t is essentially bounded by B if and only if then there exists some constant $C > 0$ such that for all $n \in \mathbb{N}$,*

$$|\kappa_n[X_1]| \leq CB^n$$

Proof. For the first direction, suppose X_t is essentially bounded by B. Choose some $0 < \epsilon < B$. By properties of the Lévy measure Π, we have that

$$\int_{[-\epsilon,\epsilon]} x^2 d\Pi(x) = R_1 < \infty$$

Let $R_2 = \Pi([-B, B]\backslash[-\epsilon, \epsilon])$. Since $\text{supp}(\Pi) \subseteq [-B, B]$, we have that for $n \geq 3$,

$$\begin{aligned}
|\kappa_n[X_1]| &= \left| \int_{-B}^{B} x^n d\Pi(x) \right| \leq \int_{-B}^{B} |x|^n d\Pi(x) \\
&\leq \int_{[-\epsilon,\epsilon]} x^2 |x|^{n-2} d\Pi(x) + \int_{[-B,B]\backslash[-\epsilon,\epsilon]} |x|^n d\Pi(x) \\
&\leq R_1 \epsilon^{n-2} + R_2 B^n \leq \left(R_1/B^2 + R_2 \right) B^n
\end{aligned}$$

Now take $C = \max\{\kappa_1[X_1]/B, \kappa_2[X_1]/B^2, R_1/B^2 + R_2\}$, and the result follows.

For the second direction, without loss of generality we will suppose that some of the support of Π lies in a region $[a, b] \subseteq (B, \infty)$, such that $\Pi([a, b]) = R_3 > 0$. Then

$$\kappa_{2n}[X_1] = \int_{\mathbb{R}} x^{2n} d\Pi(x) \geq \int_a^b x^{2n} d\Pi(x) \geq R_3 a^{2n}$$

Then if $C > 0$ is any constant, since $a/B > 1$ choose some $n \in \mathbb{N}$ large enough so that

$(a/B)^{2n} > C/R_3$, then it follows that for such an n,

$$\kappa_{2n}[X_1] \geq R_3 a^{2n} = R_3 \left(\frac{a}{B}\right)^{2n} B^{2n} > CB^{2n}$$

which concludes the proof. ■

Lemma 3.2.29. *If X_t is essentially bounded, then it has finite exponential moments $\mathbb{E}[e^{c|X_t|}] < \infty$ of all orders $c > 0$ for all $t > 0$.*

Proof. This follows directly from the fact that the Lévy measure Π is compactly supported, and from Theorem 3.2.2. ■

Lemma 3.2.30. *If X_t is a real-valued, essentially bounded Lévy process, then the distributions of X_t for all $t > 0$ satisfy the Carleman condition*

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt[2n]{m_{2n}[X_t]}} = +\infty$$

As a result, the distributions of X_t are uniquely defined by their sequence of moments.

Proof. Consider the moments of $m_n[X_t]$, which by (2.4) can be written as sums of terms which, along with coefficients, take the form of products

$$\prod_{j=1}^{n} \kappa_j[X_t]^{k_j} = \prod_{j=1}^{n} t^{k_j} \kappa_j[X_1]^{k_j}$$

with $k_j \in \{0, 1, 2, \ldots, n\}$ such that $\sum_{j=1}^{n} j \cdot k_j = n$. Note that by this condition, $\sum_{j=1}^{n} k_j \leq n$. If $L = \max\{t, 1/t\}$, then we have

$$\left| \prod_{j=1}^{n} \kappa_j[X_t]^{k_j} \right| \leq L^n \prod_{j=1}^{n} |\kappa_j[X_1]|^{k_j} \leq L^n \prod_{j=1}^{n} B^{j \cdot k_j} = L^n B^n$$

Now considering the coefficients in the expansion of the moments $m_n[X_t]$ in terms of lower order cumulants, the sum of the coefficients is the n^{th} Bell number B_n. Following Berend and Tassa (2010), the Bell numbers B_n are known to satisfy the inequality

$$B_n < \left(\frac{0.792n}{\log(n+1)}\right)^n$$

So we have that

$$\sqrt[n]{|m_n[X_t]|} \leq LB\sqrt[n]{B_n} \leq LB \cdot n$$

In particular, $1/\sqrt[2n]{m_{2n}[X_t]} \geq \frac{1}{2LBn}$, which is a divergent series in n, and so the Carleman condition

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt[2n]{m_{2n}[X_t]}} = +\infty$$

is satisfied for all $t > 0$. ■

Lemma 3.2.31. *If X_t is a real-valued, essentially bounded Lévy process, then the* CGF *$\psi_X(z)$ (and consequently, the characteristic function $\varphi_X(z)$) can be extended to an entire function of its argument $z \in \mathbb{C}$.*

Proof. Let $B, C > 0$ be the bounds given by Lemma 3.2.28, such that

$$|\kappa_n[X_1]| \leq CB^n$$

By (2.3), we get that

$$\psi_{X_t}(z) = t\sum_{k=1}^{\infty} \frac{\kappa_k[X_1]}{k!}(iz)^k$$

Since the coefficients in the power series are bounded by

$$\left| t\frac{\kappa_k[X_1]}{k!} \right| \leq tC\frac{B^k}{k!}$$

the radius of convergence is infinite. Since the CGF agrees with its power series expansion within the radius of convergence (see Lukacs, 1970), we can conclude that $\psi_X(z)$ extends to an entire function on $\mathbb{C}$. ■

Example 3.2.32. We provide an example of a process which is essentially bounded, but exhibits infinite variation. We will refer to this process as the Si process, S_t. This is the process with Lévy measure given by

$$\frac{d\Pi(x)}{dx} = \frac{1}{2x^2}\mathbb{1}_{[-1,1]\setminus\{0\}}(x)$$

The process is essentially bounded since the support of the Lévy measure is $[-1, 1]$, but has infinite variation due to the size of the singularity at the origin. The integral (3.3) evaluates to

$$\frac{1}{t}\psi_{S_t}(z) = \int_{-1}^{1} \left[e^{ixz} - 1 - ixz\right] \frac{1}{2x^2}dx = 1 - \cos(z) - z\mathrm{Si}(z)$$

where $\mathrm{Si}(z)$ is the trigonometric sine integral defined by

$$\mathrm{Si}(z) = \int_0^z \frac{\sin(w)}{w}dw$$

The variance and excess kurtosis are easy to calculate given the expression of the Lévy measure, and are $\mathrm{var}[X_t] = t$ and $\mathrm{kurt}[X_t] = 1/3t$. Notice that $1 - \cos(z) - z\mathrm{Si}(z)$ is an entire function, as expected.

The following Lemma is quite powerful, showing that the class of essentially bounded distributions is dense in $\mathrm{ID}(*)$.

Lemma 3.2.33. *Let X_t be a real-valued Lévy process with triplet (μ, σ, Π), and let $X_t^{(B)}$ for $B > 0$ be defined as the Lévy process with triplet $(\mu, \sigma, \Pi^{(B)})$ where $\Pi^{(B)}$ is defined as the Borel measure such that*

$$\Pi^{(B)}(A) \triangleq \Pi(A \cap [-B, B])$$

for every Borel set $A \subseteq \mathbb{R}$. Then for every $t > 0$, $X_t^{(B)} \xrightarrow{d} X_t$ as $B \to \infty$.

Proof. Since $|e^{ixz} - 1| \leq 2$ for $x, z \in \mathbb{R}$, it follows that when $B > 1$,

$$\begin{aligned} \left|\psi_{X_t}(z) - \psi_{X_t^{(B)}}(z)\right| &= t \cdot \left| \int_{-\infty}^{-B} \left[e^{ixz} - 1\right] d\Pi(x) + \int_{B}^{\infty} \left[e^{ixz} - 1\right] d\Pi(x) \right| \\ &\leq 2t \cdot \Pi((-\infty, -B) \cup (B, \infty)) \end{aligned}$$

By condition (2) in Theorem 3.1.1, it follows that this converges to zero for $B \to \infty$. Therefore, $\psi_{X_t^{(B)}}$ converges uniformly to ψ_{X_t}, and the result follows. ∎

Corollary 3.2.34. *Every real-valued Lévy process can be realized as the independent sum of a Brownian motion component, an essentially bounded Lévy process, and a compound Poisson process with arbitrarily small rate $r > 0$.*

Proof. The decomposition is straightforward. Let $r > 0$ be given. Consider that $\Pi((-\infty, -1) \cup (1, \infty)) < \infty$ by condition (2) in Theorem 3.1.1, and so we simply choose $B > 0$ large enough that

$$\Pi((-\infty, -B) \cup (B, \infty)) < r$$

Then Π can be written as $\Pi|_{[-B,B]} + \Pi|_{\mathbb{R}\setminus[-B,B]}$. The first corresponds to an essentially bounded process X_t^{ess}, and the second corresponds to a Lévy process whose Lévy measure has total mass $r > 0$ on the real line. By Corollary 3.2.7, this second process is a compound process P_t with rate r. Therefore, we can write

$$X_t \stackrel{d}{=} X_t^{\text{ess}} + P_t$$

where X_t^{ess} is an essentially bounded Lévy process which also includes the drift and Brownian motion components. ■

3.3 Taxonomy of ID($*$) Distributions

Throughout this section, we consider the so called pure-point processes for which $\mu = \sigma = 0$, so that the Brownian motion component $\mu t + \sigma B_t$ has been factored from the independent Lévy process under investigation.

3.3.1 Lévy α-Stable Process

The Lévy α-stable distributions for $0 < \alpha < 2$ have a convenient representation whose Lévy measure Π is given by

$$\frac{d\Pi(x)}{dx} = \begin{cases} A_+ \frac{1}{|x|^{\alpha+1}}, & x > 0 \\ A_- \frac{1}{|x|^{\alpha+1}}, & x < 0 \end{cases}$$

Here the constants $A_+, A_- \geq 0$ are used to skew the distribution. For our purposes, it will be most convenient to treat the positive and negative parts separately, so that Π can be realized as the Lévy measure of the weighted sum of two independent random variables.

The distribution will not have a well defined expectations for $0 < \alpha < 1$, but can be expressed as in (3.4) and will have paths of bounded variation almost surely. If $1 < \alpha < 2$, then the distribution can be expressed as in (3.3), and has paths of unbounded variation almost surely.

Finally, if $\alpha = 1$ then neither (3.3) nor (3.4) are appropriate, and a general expression such as the soft equation (3.1) must be used. In any case, if $A_+ = A_-$ then the distribution is symmetric.

Each of these integrals can be expressed compactly in terms of analytic functions in the complex plane. We will let $z \mapsto \log z$ denote the principal branch of the complex logarithm, defined on $\mathbb{C}\backslash(-\infty, 0]$. For $0 < \alpha < 2$ with $\alpha \neq 1$, we let $z \mapsto z^\alpha$ denote the function $z \mapsto e^{\alpha \log z}$ on $\mathbb{C}\backslash(-\infty, 0]$, with the addition that $0^\alpha = 0$. The integrals then evaluate as

$$
\begin{aligned}
0 < \alpha < 1 \qquad & \int_0^\infty \left[e^{ixz} - 1\right] \frac{1}{|x|^{\alpha+1}}\, dx = \Gamma(-\alpha) \cdot (-iz)^\alpha \\
& = -\frac{\pi}{\Gamma(\alpha+1)\sin(\alpha\pi)} |z|^\alpha \left[\cos\left(\frac{\alpha\pi}{2}\right) \mp i \sin\left(\frac{\alpha\pi}{2}\right)\right] \\
\alpha = 1 \qquad & \int_0^\infty \left[e^{ixz} - 1 - \frac{ixz}{1+x^2}\right] \frac{1}{|x|^2}\, dx = -i\vartheta\Big(\log(-iz) - (1-\gamma)\Big) \\
& = -\frac{\pi}{2}|z| \left[1 \pm i\frac{2}{\pi}\Big(\log|z| - (1-\gamma)\Big)\right] \\
1 < \alpha < 2 \qquad & \int_0^\infty \left[e^{ixz} - 1 - ixz\right] \frac{1}{|x|^{\alpha+1}}\, dx = \Gamma(-\alpha) \cdot (-iz)^\alpha \\
& = -\frac{\pi}{\Gamma(\alpha+1)\sin(\pi(\alpha+1))} |z|^\alpha \left[\cos\left(\frac{\alpha\pi}{2}\right) \mp i \sin\left(\frac{\alpha\pi}{2}\right)\right]
\end{aligned}
$$

Here $\gamma \approx 0.577$ is the Euler–Mascheroni constant, and is a byproduct of the use of the soft cutoff. The expressions $\pm$ and $\mp$ in the equations represent the functions $\operatorname{sign}(z)$ and $-\operatorname{sign}(z)$, respectively. The integrals evaluated from $(-\infty, 0]$ are simply the complex conjugates of these expressions. When $0 < \alpha < 1$, the function $(-iz)^\alpha$ is holomorphic in the upper half-plane, and so by the Paley–Weiner theorem (Strichartz, 2003) the corresponding distribution will be nonnegative.

Once the general form of the CGF ψ_α for these distributions have been established, they are often modified by introducing constants in the following way. Let

$$h_\alpha(z) = \begin{cases} -(-iz)^\alpha & 0 < \alpha < 1 \\ -iz\Big(\log(-iz) - (1-\gamma)\Big) & \alpha = 1 \\ (-iz)^\alpha & 1 < \alpha < 2 \end{cases}$$

Then the symmetric α-stable distribution with size parameter $c > 0$ has a CGF ψ_α given by

$$\psi_\alpha(c \cdot z) = k \cdot \mathrm{Re}\left[h_\alpha(c \cdot z)\right] = -c^\alpha |z|^\alpha$$

where $k > 0$ is some constant $k = |\sec(\alpha\pi/2)|$ for $\alpha \neq 1$ and $k = \frac{2}{\pi}$ for $\alpha = 1$. The tails of the densities of the symmetric distributions with size c decay like

$$\frac{\Gamma(\alpha+1)\sin(\alpha\pi/2)}{\pi} c^\alpha |x|^{-(\alpha+1)}$$

The non-symmetric α-stable distributions are a mixture of the real and imaginary parts of h_α, so that for the anti-symmetrization parameter $\beta \in [-1, 1]$ we have

$$\psi(z) = k \cdot \Big(\mathrm{Re}[h_\alpha(c \cdot z)] + i\beta \mathrm{Im}[h_\alpha(c \cdot z)]\Big)$$

which leads to the formulas typically used to describe these distributions. The original parameters can be recovered through $\beta = (A_+ - A_-)/(A_+ + A_-)$ and the dilation $t \mapsto \frac{|\Gamma(-\alpha)|}{k} t$. Note that $h_\alpha(0) = 0$, and when $1 < \alpha < 2$ we have $\psi_\alpha'(z) \to 0$ as z approaches 0 in the closed upper half plane. This first condition is necessary in order to make ψ a CGF, while the second guarantees that a mixture of the real and imaginary parts generates a random variable with zero mean.

The Lévy α-stable processes have a long history in financial modeling; see, for instance, Chapter 5 in Voit (2005). The distributions lie in the EGGC class, with the nonnegative versions for $0 < \alpha < 1$ also being GGC. When $\alpha = 1/n$ for $n = 2, 3, 4, \ldots$, the nonnegative distribution is also HCM (Bondesson, 1992, Example 5.6.2).

3.3.2 Gamma Process

The Gamma subordinator process Γ_t is the process given in form (3.4) with $\mu = \sigma = 0$ and Lévy measure Π defined as

$$\frac{d\Pi(x)}{dx} = \frac{e^{-x}}{x}\mathbb{1}_{(0,\infty)}(x)$$

Its characteristic function can be computed explicitly as

$$\varphi_{\Gamma_t}(z) = (1 - iz)^{-t}$$

As its name implies, Γ_t follows a Gamma distribution with unit scale parameter and shape t, such that $\Gamma_1 \sim \text{Exp}(1)$ is exponentially distributed. The Gamma process has finite moments of all orders, and a smooth density given by

$$f(x) = \frac{1}{\Gamma(t)}x^{t-1}e^{-x}, \quad x > 0$$

From this we see that Γ_t is HCM for all $t > 0$, and MED for small times $0 < t \leq 1$.

3.3.3 Variance Gamma Process

The Variance Gamma (VG) process is produced by subordinating a Brownian motion process $at+bB_t$ to an independent scaled Gamma subordinator $v\Gamma_t$, producing a process $av\Gamma_t+bB_{v\Gamma_t}$. The choice of parameters $a \in \mathbb{R}$, $b, v > 0$ can produce a variety of different distributions, however they can all be written in form (3.3) with $\mu = \sigma = 0$ and Lévy measure Π given by

$$\frac{d\Pi(x)}{dx} = a_1 \frac{e^{b_1 x}}{|x|} \mathbb{1}_{(-\infty,0)}(x) + a_2 \frac{e^{-b_2 x}}{x} \mathbb{1}_{(0,\infty)}(x)$$

for some parameters $a_1, a_2, b_1, b_2 > 0$. Under appropriate scaling when the drift $a = 0$, this includes the Laplace distribution (when $t = 1$) with density

$$f(x) = \frac{1}{2} e^{-|x|}$$

In this sense, the VG process is a kind of partial symmetrization of the Gamma process. From the form of the Lévy measure it is clear that the VG process is EGGC.

3.3.4 Inverse Gaussian Process

The Inverse Gaussian (IG) subordinator process T_t is the process given in form (3.4) by taking $\mu = \sigma = 0$ and Lévy measure Π given by

$$\frac{d\Pi(x)}{dx} = \frac{e^{-x}}{x^{3/2}} \mathbb{1}_{(0,\infty)}(x)$$

The CGF can be computed explicitly as

$$\psi_{T_t}(z) = 2\sqrt{\pi} t \left(1 - \sqrt{1 - iz}\right)$$

The Inverse Gaussian process is so named because, for various choices of $t > 0$ and scaling parameters, it occurs as the first passage time for a Brownian motion process with positive drift. The process has a smooth density given by

$$f(x) = \frac{t}{x^{3/2}} e^{-\left(x - \sqrt{\pi} t\right)^2 / x}, \quad x > 0$$

From the Lévy measure it is clear that the IG subordinator is in GGC.

3.3.5 Normal-Inverse Gaussian Process

The Normal Inverse Gaussian (NIG) process is produced by subordinating a Brownian motion process $at + bB_t$ to an independent scaled Inverse Gaussian subordinator vT_t. The Lévy measure and density cannot be expressed without the use of modified Bessel functions, but the CGF is given by

$$\psi(z) = \frac{1 - \sqrt{vb^2z^2 - 2ivaz + 1}}{v}$$

As the process is represented as Brownian motion subordinated by a GGC, it is necessarily EGGC.

3.3.6 Truncated Lévy Process

Truncated Lévy distributions were originally introduced in Mantegna and Stanley (1994) as true truncations of α-stable distributions. Analytic expressions were subsequently derived in Koponen (1995) by using an 'exponential' truncation, in which form they became established in the financial literature (Bouchaud and Potters, 2000; Voit, 2005). Although closed form expressions for their densities are unavailable, they can be understood as modifications of Lévy α-stable distributions which have an additional decay parameter $\lambda > 0$. They can be most easily constructed as a pure point processes whose Lévy measures are given by

$$\frac{d\Pi(x)}{dx} = \begin{cases} A_+ \frac{1}{|x|^{\alpha+1}} e^{-\lambda|x|}, & x > 0 \\ A_- \frac{1}{|x|^{\alpha+1}} e^{-\lambda|x|}, & x < 0 \end{cases}$$

The smooth exponential decay $e^{-\lambda|x|}$ confirms that these distributions have moments of all orders, so they can be expressed in the form of equation (3.3). Splitting the integral according to its positive and negative parts again, we get

$$\begin{aligned} &\int_0^\infty \left[e^{ixz} - 1 - ixz\right] \frac{e^{-\lambda|x|}}{|x|^{\alpha+1}} dx \\ &= \Gamma(-\alpha)\Big[(\lambda - iz)^\alpha - \lambda^\alpha + i\alpha\lambda^{\alpha-1}z\Big] \end{aligned} \tag{3.7}$$

for $\alpha \neq 1$ and

$$\begin{aligned} &\int_0^\infty \left[e^{ixz} - 1 - ixz\right] \frac{e^{-\lambda|x|}}{|x|^2} dx \\ &= (\lambda - iz)\Big(\log(\lambda - iz) - 1\Big) - \lambda\big(\log\lambda - 1\big) + i(\log\lambda)z \end{aligned}$$

for $\alpha = 1$. This mimics the form of the α-stable CGF, but with additional factors which normalize the 0^{th} and 1^{st} cumulants because of the use of $\lambda - iz$ in place of $-iz$. The integrals evaluated from $(-\infty, 0]$ are the complex conjugates of these expressions.

Unlike the α-stable distributions, equation (3.7) remains valid for any $\alpha < 2$ with $\alpha \neq 0, 1$, including negative values of α (Carr et al., 2002). For the special case of $\alpha = 0$, we have

$$\begin{aligned} &\int_0^\infty \left[e^{ixz} - 1 - ixz\right] \frac{e^{-\lambda|x|}}{|x|} dx \\ &= \log(\lambda - iz) - \log\lambda + i\lambda^{-1}z \end{aligned}$$

A similar change of parameters can be used to simplify the expressions for these distributions.

Let H_α be the functions

$$H_\alpha(x) = \begin{cases} \frac{1}{\alpha(\alpha-1)}\left[x^\alpha - \alpha x + \alpha - 1\right], & \alpha \leq 2, \alpha \neq 0, 1 \\ x\left[\log x - 1\right] + 1, & \alpha = 1 \\ x - \log x - 1, & \alpha = 0 \end{cases}$$

which are the solutions to the Cauchy–Euler differential equation

$$x^2 H_\alpha''' - (\alpha - 2) x H_\alpha'' = 0,$$

$$H_\alpha(1) = H_\alpha'(1) = 0,\ H_\alpha''(1) = 1$$

for $\alpha \leq 2$. Now let $C > 0$ be a size parameter, and consider a stochastic process $C \cdot X_t$ with characteristic function

$$\varphi_{C \cdot X_t}(z) = e^{t \psi_{\alpha,\beta}(C \cdot z)}$$

where

$$\psi_{\alpha,\beta}(z) = \mathrm{Re}\left[H_\alpha(1 - iz)\right] + i\beta \mathrm{Im}\left[H_\alpha(1 - iz)\right]$$

and $\beta \in [-1, 1]$ denotes the skewness. This stochastic process is driven by a random variable $C \cdot X$ with characteristic function $\varphi_X(z) = e^{\psi_{\alpha,\beta}(C \cdot z)}$, which follows a truncated Lévy distribution with $\lambda = 1$. Furthermore, any truncated Lévy distributed with a specified λ can be realized this way, since the stochastic process X_t follows a truncated Lévy law with $\lambda \sim t^{1/\alpha}$, up to rescaling.

Let $\mathrm{TL}(\alpha, \beta, C, T)$ denote the law according to $C \cdot X_T$, where $X_1 \sim X$ has CGF $\psi_{\alpha,\beta}$ defined above. From the definition, it is clear that $C \cdot X_T$ has zero mean, variance equal to $C^2 T$, and higher order cumulants (for $j > 2$) given by

$$\kappa_j[C \cdot X_T] = (-i)^j C^j T \cdot \psi^{(j)}_{\alpha,\beta}(0)$$
$$= \begin{cases} C^j T \cdot (2-\alpha) \cdot (3-\alpha) \cdots (j-1-\alpha), & j > 2 \text{ even} \\ \beta C^j T \cdot (2-\alpha) \cdot (3-\alpha) \cdots (j-1-\alpha), & j > 2 \text{ odd} \end{cases}$$

In particular, the kurtosis is given by

$$\mathrm{kurt}[C \cdot X_T] = \frac{\kappa_4[C \cdot X_T]}{\kappa_2[C \cdot X_T]^2} = \frac{(3-\alpha) \cdot (2-\alpha)}{T}$$

which is a strictly decreasing function in α, such that we have $\mathrm{kurt}[C \cdot X_T] \to 0$ as $\alpha \to 2^-$ and $\mathrm{kurt}[C \cdot X_T] \to \infty$ when $\alpha \to -\infty$. When $\beta = 0$, so that the distribution is symmetric, the tails of the density decay like

$$\frac{1}{2\alpha(\alpha-1)\Gamma(-\alpha)} \cdot \frac{C^\alpha T e^{-|x|/C}}{|x|^{\alpha+1}}$$

for $\alpha \neq 0, 1$, and like $C^\alpha T e^{-|x|/C}/2|x|^{\alpha+1}$ for $\alpha = 0, 1$. The expression

$$\frac{1}{2\alpha(\alpha-1)\Gamma(-\alpha)}$$

is positive for $\alpha < 2$ and $\alpha \neq 0, 1$. It approaches $1/2$ as $\alpha \to 1$ and $\alpha \to 0$, is bounded above by 0.6, and approaches zero as $\alpha \to 2^-$ (reflecting the convergence to normal distribution) and as $\alpha \to -\infty$.

Similar to the α-stable distributions, the stochastic process X_t for some fixed α and β has paths of bounded variation almost surely when $\alpha < 1$, and unbounded variation almost surely when $1 \leq \alpha \leq 2$. When $\alpha > 0$ the process X_t has infinite activity, while when $\alpha \leq 0$ it has finite activity and can therefore be realized as a compound Poisson process.

Chapter 4

Sample Lévy Covariance Ensemble

Definition 4.0.1 (SLCE)**.** Let X_t be a fixed Lévy process, $T > 0$ a finite time horizon, and $\lambda \in (0, \infty)$ a rectangular shape parameter. We consider the following random matrix ensemble $\mathbf{S}_N$ parametrized by the triplet (X_t, T, λ), called the Sample Lévy Covariance Ensemble (SLCE), as follows:

- $p = p(N)$ is a function $p : \mathbb{N} \to \mathbb{N}$ such that $p(N)/N \to \lambda$ as $N \to \infty$.
- $\mathbf{X} = \mathbf{X}_N$ is a sequence of $N \times p$ random matrices, such that the entries $[\mathbf{X}_N]_{i,j}$ for $N \in \mathbb{N}$, $1 \le i \le N$, $1 \le j \le p$ are all i.i.d. and follow the fixed distributions

$$[\mathbf{X}_N]_{i,j} \stackrel{d}{=} X_{T/N}$$

- We define the sequence of $p \times p$ matrices $\mathbf{S}_N$ by

$$\mathbf{S}_N = \mathbf{X}_N^\dagger \mathbf{X}_N$$

Theorem 4.0.2. *Let $\mathbf{S}_N$ be an SLCE with parameters (X_t, T, λ), and let μ_N denote the ESD* *of $\mathbf{S}_N$. Then there exists a probability distribution $\mu_{(X_t,T,\lambda)} \in \mathcal{P}(\mathbb{R}^+)$ such that*

$$\mu_N \xrightarrow{d} \mu_{(X_t,T,\lambda)}$$

almost surely. The distribution $\mu_{(X_t,T,\lambda)}$ depends continuously on its parameters $T, \lambda > 0$, and continuously (in the weak sense) on the distribution of X_t. Furthermore, $\mu_{(X_t,T,\lambda)}$ is independent of all higher order skewness in X_t, such that if X_t and X'_t are two Lévy processes with $[X]_t \stackrel{d}{=} [X']_t$, then $\mu_{(X_t,T,\lambda)} \stackrel{d}{=} \mu_{(X'_t,T,\lambda)}$.

4.1 Tools from Random Matrix Theory

Theorem 4.1.1. (Benaych-Georges and Cabanal-Duvillard, 2012, Theorem 3.2) *Let $\mathbf{Y}_N$ be a sequence of $N \times p$ random matrices with i.i.d. centered entries whose distribution might depend on N and p. Suppose $p : \mathbb{N} \to \mathbb{N}$ is a function of N as above, with $p(N)/N \to \lambda > 0$ as $N \to \infty$, and that there is a nonnegative sequence $\mathbf{c} = (c_n)_{n\geq 2}$ such that $c_n^{1/n}$ is bounded, and such that for each fixed $n \geq 2$,*

$$\frac{\mathbb{E}\left[|[\mathbf{Y}]_{1,1}|^n\right]}{N^{n/2-1}} \to c_n, \quad \text{as } N \to \infty$$

Then the ESD of $\frac{1}{N}\mathbf{Y}_N^\dagger \mathbf{Y}_N$ converges, as $N \to \infty$, to a probability measure $\mu_{\lambda,\mathbf{c}} \in \mathcal{P}(\mathbb{R}^+)$ which depends continuously on the pair $(\lambda, \mathbf{c})$. If $c_n = 0$ for all $n \geq 3$, then $\mu_{\lambda,\mathbf{c}}$ is a scaled version of the Marčenko–Pastur law with shape λ. Otherwise, $\mu_{(\lambda,\mathbf{c})}$ has unbounded support but admits exponential moments of all orders.

Lemma 4.1.2. (Benaych-Georges and Cabanal-Duvillard, 2012, Lemma 12.2) *For some increasing sequence of values $p_N \in \mathbb{N}$, let $\mathbf{M}_N$ denote a sequence of $p_N \times p_N$ independent random Hermitian matrices. Suppose that, for any $\epsilon > 0$, there is a sequence $\mathbf{M}_N^\epsilon$ of $p_N \times p_N$ independent random Hermitian matrices whose empirical spectral distributions converge* *almost surely to a probability measure $\mu^\epsilon \in \mathcal{P}(\mathbb{R}^+)$. Furthermore, suppose that*

$$\operatorname{rank}(\mathbf{M}_N - \mathbf{M}_N^\epsilon) \leq \epsilon p_N$$

for N large enough. Then the empirical spectral distribution of $\mathbf{M}_N$ converges in distribution almost surely to a deterministic probability distribution, which coincides with $\lim_{\epsilon\to 0^+} \mu^\epsilon$.

4.2 Limiting Distribution on Essentially Bounded Lévy Processes

Theorem 4.2.1. *Let $\mathbf{S}_N$ be an SLCE with parameters (X_t, T, λ), and let μ_N denote the ESD of $\mathbf{S}_N$. Suppose further that X_t is essentially bounded with zero mean. Then almost surely*

$$\mu_N \xrightarrow{d} \mu_{(X_t,T,\lambda)}$$

where $\mu_{(X_t,T,\lambda)} \in \mathcal{P}(\mathbb{R}^+)$ is a probability measure which depends continuously on the parameters $T, \lambda > 0$, and continuously (in the weak sense) on the distribution of X_t. Furthermore, $\mu_{(X_t,T,\lambda)}$ is independent of the odd cumulants of X_t, such that if X_t and X_t' are essentially bounded Lévy processes and $\kappa_{2n}[X_1] = \kappa_{2n}[X_1']$ for all $n \in \mathbb{N}$, then $\mu_{(X_t,T,\lambda)} \stackrel{d}{=} \mu_{(X_t',T,\lambda)}$.

Proof. Our goal is to show that the SLCE satisfies the conditions of Theorem 4.1.1 in some appropriate sense. We define

$$\mathbf{Y}_N = \sqrt{N}\mathbf{X}_N$$

where $\mathbf{X}_N$ is the $N \times p$ matrix appearing in the SLCE. Then the distribution of the i.i.d. entries in $\mathbf{Y}_N$ follow $\sqrt{N}X_{T/N}$. As the proof of Theorem 4.1.1 relies only on Theorem 2.6

and Proposition 2.7 in Benaych-Georges and Cabanal-Duvillard (2012), which themselves only make assumptions about the even moments, we note that $\mu_{(\lambda,\mathbf{c})}$ only depends on λ and c_{2n} for $n = 1, 2, 3, \ldots$. Our goal is then to show that

$$\frac{\mathbb{E}\left[|[\mathbf{Y}]_{1,1}|^{2n}\right]}{N^{n-1}} = \frac{\mathbb{E}\left[\left(\sqrt{N}X_{T/N}\right)^{2n}\right]}{N^{n-1}} = N \cdot \mathbb{E}[X_{T/N}^{2n}] = N \cdot m_{2n}[X_{T/N}]$$

converges, as $N \to \infty$, to some sequence c_{2n} for which $c_{2n}^{1/2n}$ is bounded.

We observe, as in Lemma 3.2.30, that by (2.4) the moments $m_{2n}[X_{T/N}]$ can be expressed as sums of products of the form

$$\prod_{j=1}^{2n} \kappa_j[X_{T/N}]^{k_j} = \prod_{j=1}^{2n} \left(\frac{T}{N}\right)^{k_j} \kappa_j[X_1]^{k_j}$$

with $k_j \in \{0, 1, 2, \ldots, 2n\}$ such that $\sum_{j=1}^{2n} j \cdot k_j = 2n$, and thus

$$1 \leq \sum_{j=1}^{2n} k_j \leq 2n$$

Consequently, the expression $N \cdot m_{2n}[X_{T/N}]$ can be written as the sum of terms of the form

$$\left(\frac{T}{N}\right)^{\sum_{j=1}^{2n} k_j - 1} \prod_{j=1}^{2n} \kappa_j[X_1]^{k_j}$$

The highest order term in N corresponds to the single choice $k_{2n} = 1$ and $k_j = 0$ for $j = 1, \ldots, 2n-1$, which occurs once in the expansion for $m_{2n}[X_T/N]$ as the term (with unit coefficient) $\kappa_{2n}[X_{T/N}]$. Therefore, we can write

$$N \cdot m_{2n}[X_{T/N}] = T\kappa_{2n}[X_1] + O(N^{-1}) = \kappa_{2n}[X_T] + O(N^{-1})$$

and it follows that

$$\frac{\mathbb{E}\left[|[\mathbf{Y}]_{1,1}|^{2n}\right]}{N^{n-1}} \xrightarrow{N\to\infty} \kappa_{2n}[X_T] = c_{2n}$$

Since X_t is essentially bounded, there exist constants $B, C > 0$ such that $\kappa_{2n}[X_T] \leq \sqrt[2n]{CB^{2n}} \leq (1+C)B$, which shows that $\sqrt[2n]{c_{2n}}$ is bounded. Since

$$\frac{1}{N}\mathbf{Y}_N^\dagger \mathbf{Y}_N = \mathbf{X}_N^\dagger \mathbf{X}_N$$

it follows from Theorem 4.1.1 that the ESD for $\mathbf{S}_N$ has a limiting distribution which depends only on λ and the even moments $\kappa_{2n}[X_T]$. We write this distribution as $\mu_{(X_t,T,\lambda)}$, and note that it depends continuously on λ, and also continuously on T by virtue of the relationship between $\kappa_{2n}[X_T]$ and c_{2n}. Furthermore, if $X_t^{(n)}$ is a sequence of essentially bounded Lévy processes which converges weakly to some process X_t, then convergence of the cumulants implies that $\mu_{(X_t^{(n)},T,\lambda)}$ converges weakly to $\mu_{(X_t,T,\lambda)}$. ■

4.3 Proof of Theorem 4.0.2

Proof. Let X_t be the driving process of the SLCE. By Corollary 3.2.34, for any $r > 0$ we can decompose the process into the sum of two independent processes

$$X_t \stackrel{d}{=} X_t^{\text{ess}} + P_t \tag{4.1}$$

where X_t^{ess} is an essentially bounded process and P_t is a compound Poisson process with rate

r. Then for each $N \in \mathbb{N}$, we can write

$$\mathbf{X}_N = \widehat{\mathbf{X}}_N + \mathbf{P}_N$$

where $\widehat{\mathbf{X}}_N$ and $\mathbf{P}_N$ are two independent $N \times p$ matrices, whose entries are i.i.d. and follow the distributions $X^{\text{ess}}_{T/N}$ and $P_{T/N}$, respectively. Let $\widehat{\mathbf{S}}_N = \widehat{\mathbf{X}}^{\dagger}_N \widehat{\mathbf{X}}_N$. We note that

$$\operatorname{rank}\left(\mathbf{S}_N - \widehat{\mathbf{S}}_N\right) \leq 2 \sum_{j=1}^{p} \mathbf{1}_{\mathbf{X}^j_N \neq \widehat{\mathbf{X}}^j_N} = 2 \sum_{j=1}^{p} \mathbf{1}_{\mathbf{P}^j_N \neq \mathbf{0}} \tag{4.2}$$

where the superscript j indicates the j^{th} column. Since the columns of $\mathbf{P}_N$ are composed of N independent copies of the compound Poisson process $P_{T/N}$, the probability

$$\mathbb{P}\left[\mathbf{1}_{\mathbf{P}^j_N \neq \mathbf{0}} = 0\right] = \mathbb{P}[P_{T/N} = 0]^N = e^{-rT}$$

Therefore, the right-hand side of (4.2) is a multiple of a Bernoulli distribution with $p(N)$ trials and probability of success $q = 1 - e^{-rT}$. Therefore, by (2.1), we have that

$$\mathbb{P}\left[\operatorname{rank}\left(\mathbf{S}_N - \widehat{\mathbf{S}}_N\right) \geq 4p(N)(1 - e^{-rT})\right] \leq \left(\frac{e}{4}\right)^{p(N)(1-e^{-rT})}$$

Since $(e/4)^{1-e^{-rT}} < 1$ and $p(N)/N \to \lambda > 0$ as $N \to \infty$, it follows that

$$\sum_{N=1}^{\infty} \left(\frac{e}{4}\right)^{p(N)(1-e^{-rT})} < \infty$$

By Borel–Cantelli, we have almost surely that

$$\text{rank}\left(\mathbf{S}_N - \widehat{\mathbf{S}}_N\right) \le 4p(N)(1 - e^{-rT}) \tag{4.3}$$

for large enough N. Now if $\epsilon > 0$ is given, we simply choose $r > 0$ small enough so that $4(1 - e^{-rT}) < \epsilon$. By Theorem 4.2.1, the sequence $\widehat{\mathbf{S}}_N$ has a limiting spectral distribution μ^ϵ. By Lemma 4.1.2, so does $\mathbf{S}_N$, given by $\lim_{\epsilon \to 0^+} \mu^\epsilon$. Continuity of μ^ϵ in the parameters $T, \lambda > 0$ implies continuity of μ.

Now suppose that $X_t^{(n)}$ is a sequence of Lévy processes converging in distribution to a Lévy process X_t. Let $\mu^{(n)}$ and μ be the appropriate limiting distributions for the SLCE with parameters $(X_t^{(n)}, T, \lambda)$ and (X_t, T, λ). By Corollary 3.1.3, we have

$$\Pi^{(n)}(\mathbb{R}\backslash[-B, B]) \xrightarrow{n\to\infty} \Pi(\mathbb{R}\backslash[-B, B])$$

Therefore, if $\epsilon > 0$ is given, some $B > 0$ can be chosen in the decompositions of the form (4.1) uniformly for all $X_t^{(n)}$ and X_t. Then the essentially bounded components of each $X_t^{(n)}$ have Lévy measures $\Pi^{(n)}\big|_{[-B,B]}$. Then the ESD of each $\widehat{\mathbf{S}}_N^{(n)}$ converges to some $\mu^{\epsilon,(n)}$, and by the continuity in Theorem 4.2.1 we have that

$$\mu^{\epsilon,(n)} \xrightarrow{n\to\infty} \mu^\epsilon$$

By the uniformity in $\epsilon > 0$ in (4.3), it follows that

$$\mu^{(n)} \xrightarrow{n\to\infty} \mu$$

Therefore, the limiting ESD μ is continuous in its parameters (X_t, T, λ).

Finally, suppose that X_t and X_t' are two Lévy processes with $[X]_t \overset{d}{=} [X']_t$. For any $\epsilon > 0$

given, choose $B > 0$ large enough that the rates $r, r' > 0$ in both decompositions of the type (4.1) are small enough. Now if μ^ϵ and $\mu^{\epsilon'}$ are the limiting distributions of $\widehat{\mathbf{S}}_N$ and $\widehat{\mathbf{S}}'_N$, it follows from Theorem 4.2.1 that $\mu^\epsilon \overset{d}{=} \mu^{\epsilon'}$. Therefore, we get that $\mu \overset{d}{=} \mu'$. ■

Chapter 5

Interlude: Free Probability

5.1 Primer on Free Probability

Free Probability was introduced by Voiculescu in the 1980s as an attempt to solve various problems in the theory of operator algebras (for an introduction, see Voiculescu et al., 1992). In brief, it introduced two commutative operations: Free additive convolution $\mu \boxplus \nu$ defined for any two probability measures μ and ν on $\mathbb{R}$, and Free multiplicative convolution $\mu \boxtimes \nu$ defined for any nonnegative probability measure μ and any general probability measure ν on $\mathbb{R}$. These operations can be defined through somewhat complicated manipulations of the Stieltjes transforms of the measures involved, or by translating the measures onto objects in a large operator algebra.

Connections to Random Matrix Theory were later identified when the following fact was realized. Suppose $\mathbf{A}_N$ and $\mathbf{B}_N$ are independent sequences of Hermitian random matrices, whose ESD's converge empirically to μ and ν, respectively. Let $\mathbf{U}_N$ be a sequence of Haar distributed orthogonal matrices, independent from $\mathbf{A}_N$ and $\mathbf{B}_N$. Then the ESD of the sum

$$\mathbf{U}_N\mathbf{A}_N\mathbf{U}_N^\dagger + \mathbf{B}_N$$

converges empirically to $\mu \boxplus \nu$. Similarly, if μ is nonnegative, then the ESD of the product

$$\sqrt{\mathbf{A}_N}\mathbf{U}_N\mathbf{B}_N\mathbf{U}_N^\dagger\sqrt{\mathbf{A}_N} \sim \mathbf{A}_N\mathbf{U}_N\mathbf{B}_N\mathbf{U}_N^\dagger$$

converges empirically to $\mu \boxtimes \nu$. The inclusion of the matrices $\mathbf{U}_N$ can be omitted if either of the random matrix ensembles $\mathbf{A}_N$ or $\mathbf{B}_N$ are asymptotically unitarily invariant.

Much of the theory of Free Probability mirrors classical probability through the Bercovici–Pata bijection Λ on the space of probability measures, which takes a measure μ in the classical setting and produces its free counterpart $\Lambda(\mu)$. For instance, the Λ-image of normal distribution is Wigner's famous Semicircle distribution, which occurs as the limit in the Free central limit theorem. Similarly, the Marčenko–Pastur distribution mp_λ, after a rescaling, occurs as the Λ-image of Poisson distribution with parameter $1/\lambda$.

The bijection also connects classically infinitely divisible distributions to free infinitely divisible distributions due to the fact that

$$\nu_n^{*n} \to \mu \iff \nu_n^{\boxplus n} \to \Lambda(\mu)$$

This leads naturally to an idea of Free Lévy processes, which occur as the Λ-image of classical Lévy processes. One important distinction is that a classical Lévy process X_t is a subordinator (nonnegative for all $t > 0$) if and only if it is nonnegative for a single $t > 0$. Such a property on the supports of $\Lambda(X_t)$ does not hold, as it is possible to find a non-subordinator X_t while $\Lambda(X_t)$ has strictly positive support for some values of $t \geq t_0$ and not for $0 < t < t_0$. However, if X_t is known to be a subordinator, then it follows that $\Lambda(X_t)$

will have nonnegative support for all $t > 0$. These Λ-images of subordinators are called Free Regular probability measures.

In 2008, Benaych-Georges began publishing on techniques to generalize Free Probability to a setting suitable for the sum of rectangular matrices, with a focus on singular values rather than eigenvalues. For a parameter $\lambda \in (0, 1]$, he defined a commutative binary operation $\boxplus_\lambda$ on the space of probability measures. If $\mathbf{A}_N^{(1)}$ and $\mathbf{A}_N^{(2)}$ are $N \times p$ random matrices with $p/N \to \lambda$ whose empirical singular distributions converge to μ and ν, respecitvely, and if $\mathbf{U}_N, \mathbf{V}_N$ are Haar distributed orthogonal matrices of the appropriate sizes with all matrices independent of one another, then the empirical singular distribution of the sum

$$\mathbf{U}_N \mathbf{A}_N^{(1)} \mathbf{V}_N^\dagger + \mathbf{A}_N^{(2)}$$

converges empirically to $\mu \boxplus_\lambda \nu$.

The approach to the operations in Free Probability presented here follows along the lines of Chistyakov and Götze (2011). This avoids the discussion of operator algebras entirely, and also presents the operations in terms of analytic subordination rather than local inverses. All major lemmas and theorems are proved therein.

5.1.1 Existence of Free Addition and Multiplication

Definition 5.1.1. The nontangential limit $z \xrightarrow{\measuredangle} a$ for $a \in \mathbb{R} \cup \{\infty\}$, is the limit for $z \in \mathbb{C}^+$ or $z \in \mathbb{C}^-$ (depending on the context) to a, with the condition that $|\mathrm{Re}(z)|/|\mathrm{Im}(z)|$ remains bounded.

Definition 5.1.2. An analyic funtion $f : \mathbb{C}^+ \to \overline{\mathbb{C}^+}$ is said to be in the Nevanlinna class $\mathcal{N}$.

If, additionally, we have that

$$\lim_{z \xrightarrow{\measuredangle} \infty} f(z)/z = 1$$

then we say that f is in the reciprocal Cauchy class $\mathcal{F} \subseteq \mathcal{N}$. Alternatively, if f satisfies $\arg(f(z)) \in [\arg z, \pi)$ for all $z \in \mathbb{C}^+$ and extends to an analytic function on $\mathbb{C}\backslash\mathbb{R}^+$ with the properties that $f(\overline{z}) = \overline{f(z)}$, that f is nonpositive on the negative real axis, and that the limits

$$\lim_{z \xrightarrow{\measuredangle} 0} f(z) = 0$$

$$\lim_{x \to 0^-} f(x) = 0$$

hold, then we say that f is in the modified Krein class $\mathcal{K} \subseteq \mathcal{N}$.

Definition 5.1.3. Let $\mu \in \mathcal{P}(\mathbb{R})$ be any probability measure on the real line. We define the following analytic transformations on $\mathbb{C}\backslash\mathbb{R}$:

- ▶ The reciprocal Cauchy (or reciprocal Stieltjes) transform $F_\mu(z) = -1/S_\mu(z)$, where $S_\mu(z)$ is the Stieltjes transform of μ.
- ▶ The η-transform $\eta_\mu(z) = 1 - zF_\mu(1/z)$.

We note that the reciprocal Cauchy transform can be extended to all $z \in \mathbb{C}\backslash\text{supp}(\mu) \subseteq \mathbb{C}\backslash\mathbb{R}$.

The following lemma states that the class of reciprocal Cauchy transforms behaves as its name implies: each arises precisely from a probability measure on $\mathbb{R}$, and any two such measures have a particular subordination property that allows us to define an associative, symmetric binary operation $\boxplus$ on $\mathcal{P}(\mathbb{R})$.

Lemma 5.1.4. *If $\mu \in \mathcal{P}(\mathbb{R})$, then $S_\mu(z)$ is a Nevanlinna function. As a result, $F_\mu(z)$ is a Nevanlinna function as well, with $F_\mu \in \mathcal{F}$. Furthermore, any functions in $\mathcal{F}$ can be realized* *as the reciprocal Cauchy transform of some probability measures. Furthermore, if $\nu \in \mathcal{P}(\mathbb{R})$, then there exist unique Nevanlinna functions $Z_1, Z_2 \in \mathcal{F}$ such that*

$$F_\mu(Z_1(z)) = F_\nu(Z_2(z)) = Z_1(z) + Z_2(z) - z, \qquad z \in \mathbb{C}^+$$

By the definition of $\mathcal{F}$, the function $F_\mu(Z_1(z)) = F_\nu(Z_2(z))$ is in $\mathcal{F}$ as well. As a result, there is a unique probability measure, denoted by $\mu \boxplus \nu \in \mathcal{P}(\mathbb{R})$, such that

$$F_{\mu\boxplus\nu}(z) = F_\mu(Z_1(z)) = F_\nu(Z_2(z)), \qquad z \in \mathbb{C}^+$$

This defines the operation $\boxplus$ as an associative, symmetric binary operation on the space of probability measures $\mathcal{P}(\mathbb{R})$, called Free additive convolution.

As was discussed in Section 3.1, distributions $\mu \in \mathcal{P}(\mathbb{R})$ generate a discrete semigroup μ^{*n} for $n \in \mathbb{N}$ by taking successive classical additive convolutions. The extension of the parameter to continuous values is the focus in the theory of classical infinite divisibility. In contrast, the following lemma shows that the Free additive case is completely different: a probability measure generates a continuous semigroup $\mu^{\boxplus t}$, where t is allowed to take values $\{0\} \cup [1, \infty)$ without any restrictions on the initial distribution μ. The theory of Free infinite divisibility will therefore be concerned primarily with the extension of t to the range $(0, 1)$.

Lemma 5.1.5. *If we consider the n-fold convolution $\mu^{\boxplus n}$, there exists a unique $Z \in \mathcal{F}$ such that*

$$z = nZ(z) - (n-1)F_\mu(Z(z))$$

This Z is such that $Z(z) = F_{\mu^{\boxplus n}}(z)$. This relation can be relaxed by replacing $n \in \mathbb{N}$ with $t \geq 1$, so that there is a $\boxplus$-semigroup of probability measures $\mu^{\boxplus t}$ for $t \geq 1$ with $\mu^{\boxplus 1} = \mu$, *such that for any $t_1, t_2 \geq 1$,*

$$\mu^{\boxplus t_1} \boxplus \mu^{\boxplus t_2} = \mu^{\boxplus (t_1+t_2)}$$

Finally, a similar subordination property can be applied to the η-functions to produce a multiplicative Free operation on nonnegative probability measures.

Lemma 5.1.6 (Arizmendi and Hasebe, 2013)**.** *The modified Krein class $\mathcal{K}$ is precisely the class of η-functions of nonnegative measures $\mu \in \mathcal{P}(\mathbb{R}^+)$. Furthermore, if $\mu, \nu \in \mathcal{P}(\mathbb{R}^+)$, then there exist two unique members of the modified Krein class $K_1, K_2 \in \mathcal{K}$ such that*

$$\eta_\mu(K_1(z)) = \eta_\nu(K_2(z)) = \frac{K_1(z)K_2(z)}{z}, \qquad z \in \mathbb{C}^+$$

As a result, there exists a unique probability measure, denoted by $\mu \boxtimes \nu \in \mathcal{P}(\mathbb{R}^+)$, such that

$$\eta_{\mu\boxtimes\nu}(z) = \eta_\mu(K_1(z)) = \eta_\nu(K_2(z)), \qquad z \in \mathbb{C}^+$$

This makes $\boxtimes$ into an associative symmetric binary operation on the space of nonnegative probability measures $\mathcal{P}(\mathbb{R}^+)$, called the Free multiplicative convolution.

5.1.2 Free Lévy Processes and Infinite Divisibility

Similar to classical infinite divisibility, there exists complete characterizations of infinite divisibility for Free additive convolution. We denote this class by $\mathrm{ID}(\boxplus)$.

Theorem 5.1.7 (Voiculescu et al., 1992; Pérez-Abreu and Sakuma, 2012)**.** *A measure $\mu \in \mathrm{ID}(\boxplus)$ if and only if it admits a right inverse F_μ^{-1} on a region of the shape*

$$\Gamma_{\eta,M} = \{z \in \mathbb{C}^+ : |\mathrm{Re}(z)| < \eta|\mathrm{Im}(z)|,\ \mathrm{Im}(z) > M\}$$

where F_μ^{-1} is univalent, and furthermore the function $R_\mu(z) \triangleq zF_\mu^{-1}(1/z) - 1$ for $1/z \in \Gamma_{\eta,M}$ can be expressed as

$$R_\mu(z) = \gamma z + \sigma^2 z^2 + \int_{\mathbb{R}} \left(\frac{1}{1 - xz} - 1 - xz\mathbb{1}_{[-1,1]}(x) \right) d\Pi(x)$$

for a triplet (γ, σ, Π) where $\gamma \in \mathbb{R}$, $\sigma \geq 0$, and Π is a Lévy measure. When this exists, the triplet uniquely determines the distribution of μ.

Corollary 5.1.8 (Bercovici and Pata, 1999)**.** *There exists a well defined bijection between the space of classically infinitely divisible distributions* $\mathrm{ID}(*) \subset \mathcal{P}(\mathbb{R})$ *and the space of Freely infinitely divisible distributions* $\mathrm{ID}(\boxplus) \subset \mathcal{P}(\mathbb{R})$*, called the Bercovici–Pata bijection* $\Lambda : \mathrm{ID}(*) \to \mathrm{ID}(\boxplus)$*, which maps the Lévy triplet of one to the other.*

Example 5.1.9. The Bercovici–Pata bijection implies that all classical distributions have some sort of Free analog. The most immediate question is, which distribution is the Free Gaussian, or rather what is $\Lambda(\boldsymbol{g})$ where $\boldsymbol{g} \stackrel{d}{=} \mathrm{N}(0,1)$? The answer turns out to be the standard Semicircle distribution $\boldsymbol{s}$, which has a continuous, compactly supported density given by

$$\frac{d\boldsymbol{s}(x)}{dx} = \frac{1}{2\pi}\sqrt{4 - x^2}\mathbb{1}_{[-2,2]}(x)$$

Theorem 5.1.10 below shows that limit theorems from classical probability are guaranteed to have free analogues. For instance, a free central limit theorem must hold, where the limiting distribution must be $\boldsymbol{s}$.

Theorem 5.1.10 (Bercovici and Pata, 1999)**.** *Let $\mu_j \in \mathcal{P}(\mathbb{R})$ be a sequence of distributions, and $a_j \in \mathbb{R}$ and $b_j > 0$ some sequences of real numbers. Then the sequence of probability measures converges in distribution*

$$D_{1/b_n}(\mu_1 * \mu_2 * \cdots * \mu_n) * \delta_{a_n} \xrightarrow{d} \mu$$

if and only if the corresponding free sequence also converges in distribution

$$D_{1/b_n}(\mu_1 \boxplus \mu_2 \boxplus \cdots \boxplus \mu_n) \boxplus \delta_{a_n} \xrightarrow{d} \nu$$

Furthermore, in this case $\mu \in \mathrm{ID}(*)$ *and* $\nu \in \mathrm{ID}(\boxplus)$, *and* $\nu = \Lambda(\mu)$.

Example 5.1.11. The free analogue of the Poisson distribution with rate $t > 0$ can be calculated by considering that

$$\left(\left(1 - \frac{t}{n}\right)\delta_0 + \frac{t}{n}\delta_1\right)^{\boxplus n} \xrightarrow{d} \Lambda(N_t)$$

as $n \to \infty$, where N_t is the classical Poisson process with rate t. The corresponding distribution is called the Free Poisson $\boldsymbol{\pi}^{\boxplus t}$, which can be decomposed as

$$\boldsymbol{\pi}^{\boxplus t} = \max\{0, 1-t\}\delta_0 + \boldsymbol{\pi}_t^{\mathrm{abs}}$$

where the absolutely continuous part $\boldsymbol{\pi}_t^{\mathrm{abs}}$ is given by

$$\frac{d\boldsymbol{\pi}_t^{\mathrm{abs}}(x)}{dx} = \frac{\sqrt{(t_+ - x)(x - t_-)}}{2\pi x}\mathbb{1}_{[t_-, t_+]}(x) \tag{5.1}$$

where $t_\pm = (1 \pm \sqrt{t})^2$. The distribution consists of a point mass at zero for small times $0 < t < 1$, and a bulk with support $[t_-, t_+]$ of width $4\sqrt{t}$ and center $1 + t$. As $t \to 1^-$, the bulk approaches the origin, finally colliding with it and "absorbing" the point mass when $t = 1$, after which it moves away from the origin as $t > 1$ grows. We note that this behavior is consistent with Theorem 5.1.16 below.

Lemma 5.1.12 (Hasebe, 2012)**.** *Suppose* $\mu \in \mathrm{ID}(\boxplus)$. *Then* $\mu^{\boxplus t} \in \mathcal{P}(\mathbb{R}^+)$ *for all* $t \geq 0$ *if and only if* $\mu = \Lambda(\nu)$ *for some* $\nu \in \mathrm{ID}(*)$ *with* $\nu \in \mathcal{P}(\mathbb{R}^+)$, *which is to say that* ν *is* *the distribution of a subordinator process. We say that such a distribution is Free Regular, written* $\mu \in \mathrm{FR}$.

5.1.3 Free Regular Distributions

A distribution $\mu^{\boxplus t}$ whose distribution $\mu = \mu^{\boxplus 1}$ is in $\mathcal{P}(\mathbb{R}^+) \cap \mathrm{ID}(\boxplus)$ may not necessarily have nonnegative support for all small times $\mu^{\boxplus t}$, $0 < t < 1$. Those distributions which have this property, satisfying the conditions of Lemma 5.1.12, are called Free Regular and denoted by FR. The Free Regular distributions are in correspondence with classical nonnegative infinitely divisible distributions, the subordinators, as being nonnegative is a path property in the classical setting.

Example 5.1.13. Early on in the theory of Free Probability, it was suspected that the Free multiplication

$$\mu \boxtimes \nu, \quad \mu, \nu \in \mathrm{ID}(\boxplus)$$

would remain in $\mathrm{ID}(\boxplus)$. Unfortunately, this is not the case, as the shifted semicircle distribution $\mu = \boldsymbol{s} \boxplus \delta_2$ with density

$$\frac{d\mu(x)}{dx} = \frac{1}{2\pi}\sqrt{4 - (x-2)^2}\mathbb{1}_{[0,4]}(x)$$

is positive and $\mu \in \mathrm{ID}(\boxplus)$, however $\mu \boxtimes \mu \notin \mathrm{ID}(\boxplus)$. We note that, in spite of the fact that $\mu \in \mathcal{P}(\mathbb{R}^+)$, the distributions of $\mu^{\boxplus t}$ for $0 < t < 1$ are no longer nonnegative. In fact, the support of $\mu^{\boxplus t}$ is $[2t - 2\sqrt{t}, 2t + 2\sqrt{t}]$ for all $t > 0$, and so $\mu^{\boxplus t} \notin \mathcal{P}(\mathbb{R}^+)$ for $0 < t < 1$. As a result, $\mu \notin \mathrm{FR}$. The following theorem shows that this issue is resolved when considering products $\mu \boxtimes \nu$ and one of the two distributions is Free Regular.

Lemma 5.1.14 (Arizmendi et al., 2013)**.** *Let* $\mu, \nu \in \mathrm{FR}$ *and* $\sigma \in \mathrm{ID}(\boxplus)$*. Then*

- $\mu \boxtimes \nu \in \mathrm{FR}$
- $\mu \boxtimes \sigma \in \mathrm{ID}(\boxplus)$
- $\sqrt{\boldsymbol{\pi} \boxtimes \mu} \in \mathrm{ID}(\boxplus)$
- $\sigma^2 \in \mathrm{FR}$, *and there exists some unique* $\upsilon \in \mathrm{FR}$ *such that* $\sigma^2 = \boldsymbol{\pi} \boxtimes \upsilon$

5.1.4 Regularization of Free Addition

Theorem 5.1.15 (Belinschi, 2008)**.** *Let* $\mu, \nu \in \mathcal{P}(\mathbb{R})$*, neither of which are point masses. Then* $\mu \boxplus \nu$ *has no singular continuous part, and we have the following properties:*

- *The discrete part of* $\mu \boxplus \nu$ *has an atom located at* $a \in \mathbb{R}$ *if and only if* a *can be written as* $a = b+c$ *for some (necessarily unique) values* $b, c \in \mathbb{R}$ *such that* $\mu(\{b\})+\nu(\{c\}) > 1$*. Furthermore,*

  $$(\mu \boxplus \nu)(\{a\}) = \mu(\{b\}) + \nu(\{c\}) - 1$$

- *The absolutely continuous part of* $\mu \boxplus \nu$ *can be written with a density* $f : U \to [0, \infty)$ *where* $U \subseteq \mathbb{R}$ *is open and* f *is analytic on* U*, such that*

  $$(\mu \boxplus \nu)^{\mathrm{abs}}(A) = \int_A f(x)dx$$

  for any Borel set $A \subseteq \mathbb{R}$*.*

Theorem 5.1.16 (Belinschi and Bercovici, 2004)**.** *Let* $\mu \in \mathcal{P}(\mathbb{R})$*, and let* $t > 1$*. Then the measure* $\mu^{\boxplus t}$ *has no singular continuous part, and we have the following properties:*

- *If* $a \in \mathbb{R}$*, the discrete part of* $\mu^{\boxplus t}$ *includes an atom located at* $a \cdot t$ *if and only if* a *is an* *atom of μ with $\mu(\{a\}) > 1 - 1/t$, in which case*

$$\mu^{\boxplus t}(\{a \cdot t\}) = 1 - t(1 - \mu(\{a\}))$$

Note that if N_t denotes the total number of atoms of $\mu^{\boxplus t}$, then this leads to the inequality $N_t < \frac{t}{t-1}$.

- *If $\mu^{\boxplus t}$ has two distinct atoms $a < b$, then $\mu^{\boxplus t}((a,b)) > 0$.*
- *The absolutely continuous part of $\mu^{\boxplus t}$ can be written with a density $f_t : U_t \to [0, \infty)$ where $U_t \subseteq \mathbb{R}$ is open and each f_t is analytic on each U_t, such that*

$$(\mu^{\boxplus t})^{\mathrm{abs}}(A) = \int_{A \cap U_t} f_t(x)dx$$

for any Borel set $A \subseteq \mathbb{R}$.

5.2 Rectangular Free Probability

Definition 5.2.1. For $\lambda \in [0,1]$, we define the U and T function as

$$U(z) \triangleq \frac{(\lambda+1) + \sqrt{(\lambda+1)^2 + 4\lambda z}}{2\lambda}$$
$$T(z) \triangleq (\lambda z + 1)(z + 1)$$

the former of which is analytic for $|z| < (\lambda+1)^2/4\lambda$. When $\lambda = 0$ we simply take $U(z) = z$. We note that $T(U(z-1)) = z$ where U is analytic. If $\mu \in \mathcal{P}(\mathbb{R})$, we define the rectangular Cauchy transform with ratio λ of μ as the analytic function

$$H_\mu(z) \triangleq z \cdot T\left(\frac{1}{z}G_{\mu^2}\left(\frac{1}{z}\right) - 1\right) = \frac{\lambda}{z}G_{\mu^2}\left(\frac{1}{z}\right)^2 + (1-\lambda)G_{\mu^2}\left(\frac{1}{z}\right)$$

which is well defined on $\mathbb{C}\backslash\mathbb{R}^+$.

Lemma 5.2.2 (Benaych-Georges, 2009a)**.** *Let $\lambda \in [0,1]$, and let $\mu \in \mathcal{P}(\mathbb{R})$ be a probability measure. Then H_μ has a well defined functional inverse on an interval $(-\epsilon, 0)$ for some $\epsilon > 0$, and the expression*

$$C_\mu(z) \triangleq U\left(\frac{z}{H_\mu^{-1}(z)} - 1\right)$$

is well defined on such an interval.

Theorem 5.2.3 (Benaych-Georges, 2009a)**.** *Let $\lambda \in [0,1]$, and let $\mu, \nu \in \mathcal{P}_s(\mathbb{R})$ be symmetric probability measures. Then there exists a unique symmetric probability measure, denoted by $\mu \boxplus_\lambda \nu$, such that*

$$C_{\mu \boxplus_\lambda \nu}(z) = C_\mu(z) + C_\nu(z)$$

on the intersection of the intervals where the latter two are defined. This introduces a well defined symmetric binary operation $\boxplus_\lambda$ on the subset of symmetric probability measures in $\mathcal{P}(\mathbb{R})$, called the λ-shaped Rectangular Free additive convolution.

Corollary 5.2.4. *If $\mu, \nu \in \mathcal{P}_s(\mathbb{R})$ are symmetric probability measures, then*

$$\mu \boxplus_0 \nu = \sqrt{\mu^2 \boxplus \nu^2}$$

$$\mu \boxplus_1 \nu = \mu \boxplus \nu$$

Theorem 5.2.5 (Benaych-Georges, 2010)**.** *If $\mu, \nu \in \mathcal{P}(\mathbb{R}^+)$ are nonnegative probability measures and $\lambda \in (0,1)$, then*

$$\sqrt{\mu \boxtimes \mathrm{mp}_\lambda} \boxplus_\lambda \sqrt{\nu \boxtimes \mathrm{mp}_\lambda} = \sqrt{(\mu \boxplus \nu) \boxtimes \mathrm{mp}_\lambda}$$

Theorem 5.2.6 (Belinschi et al., 2009b)**.** *If $\mu, \nu \in \mathcal{P}_s(\mathbb{R})$ are symmetric probability measures* *and* $\lambda \in (0,1)$*, then*

$$(\mu \boxplus_\lambda \nu)(\{0\}) \geq \mu(\{0\}) + \nu(\{0\}) - 1$$

On the other hand, if $\mu(\{0\}) + \nu(\{0\}) < 1$ *then there is some* $\epsilon > 0$ *such that*

$$(\mu \boxplus_\lambda \nu)((-\epsilon, \epsilon)) = 0$$

The following two theorems summarize the theory of Rectangular Free infinite divisibility.

Theorem 5.2.7 (Benaych-Georges, 2010)**.** *A measure*

$$\mu \in \mathcal{P}_s(\mathbb{R})$$

belongs to $\mathrm{ID}(\boxplus_\lambda)$ *if and only if there exists a Freely regular probability measure* $\nu \in \mathrm{FR}$ *such that*

$$\mu = \sqrt{\nu \boxtimes \mathrm{mp}_\lambda}$$

Furthermore, for any $t > 0$ *we have*

$$\mu^{\boxplus_\lambda t} = \sqrt{\nu^{\boxplus t} \boxtimes \mathrm{mp}_\lambda}$$

Theorem 5.2.8 (Belinschi et al., 2009b)**.** *Suppose* $\mu, \nu \in \mathcal{P}_s(\mathbb{R})$ *are symmetric probability measures with* $\lambda \in (0,1)$*, and* $\mu \in \mathrm{ID}(\boxplus_\lambda)$*. If* $(\mu \boxplus_\lambda \nu)(\{0\}) > 0$ *then* $\mu(\{0\}) + \nu(\{0\}) > 1$ *and*

$$(\mu \boxplus_\lambda \nu)(\{0\}) = \mu(\{0\}) + \nu(\{0\}) - 1$$

5.3 Asymptotics of Large Random Matrices

Theorem 5.3.1 (Voiculescu et al., 1992)**.** *Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ be any real-valued probability measures. Let $\mathbf{A}_N$ and $\mathbf{B}_N$ be $N \times N$ random Hermitian matrices. Suppose that, almost surely, the ESD of $\mathbf{A}_N$ converges in distribution to μ and the ESD of $\mathbf{B}_N$ converges in distribution to ν. Finally, let $\mathbf{U}_N$ denote a sequence of $N \times N$ random orthogonal matrices. Suppose all matrices are independent from one another and for each N. We take*

$$\mathbf{C}_N = \mathbf{U}_N \mathbf{A}_N \mathbf{U}_N^\dagger + \mathbf{B}_N$$

Then the ESD of $\mathbf{C}_N$ converges in distribution almost surely to $\mu \boxplus \nu$. Furthermore, if $\mathbf{A}_N$ and $\mathbf{B}_N$ are nonnegative definite Hermitian random matrices with $\sqrt{\mathbf{B}_N}$ denoting the Hermitian square root of $\mathbf{B}_N$, and we consider

$$\mathbf{D}_N = \sqrt{\mathbf{A}_N} \mathbf{U}_N \mathbf{B}_N \mathbf{U}_N^\dagger \sqrt{\mathbf{A}_N} \sim \mathbf{A}_N \mathbf{U}_N \mathbf{B}_N \mathbf{U}_N^\dagger$$

then the ESD of $\mathbf{D}_N$ converges in distribution almost surely to $\mu \boxtimes \nu$.

Theorem 5.3.2 (Benaych-Georges, 2009a)**.** *Let $\mu, \nu \in \mathcal{P}_s(\mathbb{R})$ be any symmetric probability measures. Let $p = p(N)$ be a function $p : \mathbb{N} \to \mathbb{N}$ such that $p(N) \le N$ for all $N \in \mathbb{N}$ and $p(N)/N \to \lambda \in (0, 1)$. Let $\mathbf{A}_N$ and $\mathbf{B}_N$ be $N \times p$ random matrices. Suppose that, almost surely, the ESD of $\mathbf{A}_N^\dagger \mathbf{A}_N$ converges in distribution to μ^2 and the ESD of $\mathbf{B}_N^\dagger \mathbf{B}_N$ converges in distribution to ν^2. Finally, let $\mathbf{U}_N$ and $\mathbf{V}_N$ denote sequences of $N \times N$ and $p \times p$ random orthogonal matrices. Suppose all matrices described are independent. Take*

$$\mathbf{C}_N = \mathbf{U}_N \mathbf{A}_N \mathbf{V}_N^\dagger + \mathbf{B}_N$$

Then the ESD of $\mathbf{C}_N^\dagger \mathbf{C}_N$ converges in distribution almost surely to $(\mu \boxplus_\lambda \nu)^2$.

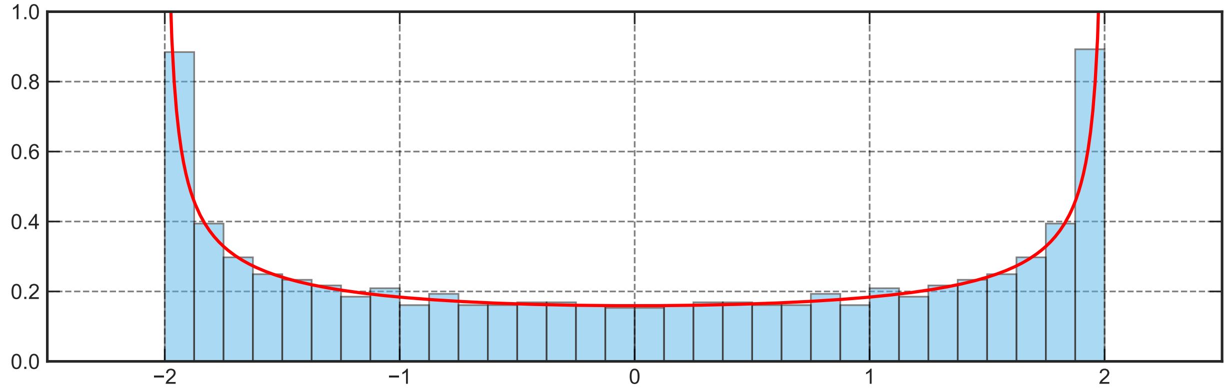


Figure 5.1: Comparison of the eigenvalues of a matrix of the type described in Example 5.3.3 with $N = 10^3$, and the density of the Arcsine distribution $\boldsymbol{a}$ (red line).

In other words, if two $N \times p$ random matrix ensembles have limiting (symmetrized) singular values following distributions μ and ν as $N, p \to \infty$ with $p/N \to \lambda \in (0,1)$, and if they are bi-unitarily independent, then their sum has limiting singular values following $\mu \boxplus_\lambda \nu$.

Example 5.3.3. We let $\boldsymbol{r} = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$ denote the Rademacher distribution. It is famously known that $\boldsymbol{r} \boxplus \boldsymbol{r} = \boldsymbol{r}^{\boxplus 2} = \boldsymbol{a}$, where $\boldsymbol{a}$ is the Arcsine distribution concentrated on the set $[-2, 2]$, which can be described by the density

$$\frac{d\boldsymbol{a}(x)}{dx} = \frac{1}{\pi\sqrt{4 - x^2}} \mathbb{1}_{[-2,2]}(x)$$

From Theorem 5.3.1, we can estimate the eigenvalues of a matrix of the form

$$\mathbf{C}_N = \mathbf{U}_N \mathbf{A}_N \mathbf{U}_N^\dagger + \mathbf{B}_N$$

where $\mathbf{A}_N$ and $\mathbf{B}_N$ are large $N \times N$ diagonal matrices with half of their diagonal entries equal to $+1$ and half equal to -1. The resulting eigenvalues of $\mathbf{C}_N$ will be approximately distributed like $\boldsymbol{a}$. An example is shown in Figure 5.1.

5.3.1 Generalized Marčenko–Pastur and free Poisson distributions

Lemma 5.3.4 (Pérez-Abreu and Sakuma, 2012)**.** *Let $\boldsymbol{\pi}(t,\nu)$ be a free compound Poisson distribution, which is to say that $\boldsymbol{\pi}(t,\nu) = \Lambda(P(t,\nu))$ where $P(t,\nu)$ is the classical compound Poisson distribution with rate $t > 0$ and jump distribution $\nu \in \mathcal{P}(\mathbb{R})$. If ν is either nonnegative or symmetric, then*

$$\boldsymbol{\pi}(t,\nu)^{\boxplus 1/t} = \boldsymbol{\pi} \boxtimes \nu$$

for $t > 0$. In particular, we also have that for $t \geq 1$,

$$\boldsymbol{\pi}(t,\nu) = D_{1/t}\left(\boldsymbol{\pi}^{\boxplus t}\right) \boxtimes \nu^{\boxplus t}$$

Example 5.3.5. Comparing (2.6) and (5.1), it is clear that the Marčenko–Pastur distribution mp_λ is a dilation of the Free Poisson $\boldsymbol{\pi}^{\boxplus t}$, explicitly

$$\mathrm{mp}_\lambda = D_\lambda\left(\boldsymbol{\pi}^{\boxplus 1/\lambda}\right)$$

for any $\lambda > 0$. Looking at the limiting distribution described by Theorem 2.2.4, where $\lambda \in (0,1)$, we can see from Theorem 5.3.1 that $\mu_{\lambda,\nu} = \mathrm{mp}_\lambda \boxtimes \nu$. If the distribution $\nu \in \mathcal{P}(\mathbb{R}^+)$ is Free Regular, so that $\nu^{\boxplus t}$ is well defined for any $t > 0$, we can write

$$\mu_{\lambda,\nu} = \mathrm{mp}_\lambda \boxtimes \nu = D_\lambda\left(\boldsymbol{\pi}^{\boxplus 1/\lambda}\right) \boxtimes \nu = \boldsymbol{\pi}(1/\lambda, \nu^{\boxplus \lambda})$$

In this case, by Theorem 5.2.7 we also have that the (symmetrized) distribution $\sqrt{\mu_{\lambda,\nu}}$ of the singular values of $\frac{1}{\sqrt{N}}\mathbf{Y}_N\sqrt{\mathbf{T}_N}$ is a distribution in $\mathrm{ID}(\boxplus_\lambda)$.

Example 5.3.6. From the preceding discussion, we can also observe that the Free Poisson distribution $\boldsymbol{\pi}$ plays the role of the exponential distribution Exp(1) as described in Section 3.2.5 in the context of free probability. In particular, if $\nu \in \mathcal{P}(\mathbb{R}^+)$ is any nonnegative distribution, then $\boldsymbol{\pi} \boxtimes \nu \in \mathrm{FR}$, so Free independent "scale-mixtures" with the Free Poisson are necessarily Free Regular.

Chapter 6

Performance of SLCE for EGGC and CGGC

6.1 Quadratic Variation of a Lévy Process

To motivate this section, we consider the structure of the proof of the Marčenko–Pastur theorem along the lines of Yao et al. (2015). We take the ensemble of matrices $\mathbf{Y}_N$ described in Theorem 2.2.3: a random variable Y with mean zero and unit variance is chosen, and the matrices $\mathbf{Y}_N$ are $N \times p$ with i.i.d. entries $[\mathbf{Y}_N]_{ij} \stackrel{d}{=} Y$ for all $N \in \mathbb{N}$. For brevity, we omit N and simply write $\mathbf{Y}$, letting $S_N(z)$ denote the Stieltjes transform of $\mathbf{Y}^\dagger \mathbf{Y}$. Let $\mathbf{Y}_k$ denote the $N \times (p-1)$ matrix equal to $\mathbf{Y}$ with the k^{th} column removed, and denote that column vector by $\mathbf{y}_k$. Then the resolvent formula can be rewritten in the following form (Yao et al., 2015, Theorem A.4):

$$S_N(z) = \frac{1}{p}\sum_{k=1}^{p} \frac{1}{\mathbf{y}_k^\dagger \mathbf{y}_k - z - \mathbf{y}_k^\dagger \mathbf{Y}_k(\mathbf{Y}_k^\dagger \mathbf{Y}_k - z\mathbf{I}_{p-1})^{-1}\mathbf{Y}_k^\dagger \mathbf{y}_k} \tag{6.1}$$

The goal of the proof is to show that the denominator of the expression converges to its expectation, and then to compute an expression for that expectation. The final form of the proof is to show that

$$S_N(z) \approx \frac{1}{p}\sum_{k=1}^{p} \frac{1}{1 - z - (\lambda + \lambda z S_N(z))}$$

which simplifies immediately to $S_N(z) \approx 1/(1 - z - (\lambda + \lambda z S_N(z)))$. The distribution mp_λ emerges precisely as the one whose Stieltjes transform satisfies this equation.

There are two critical steps in the proof that may be violated by the substitution of an SLCE type sequence $\mathbf{X}$ in place of the MP type $\mathbf{Y}$. Both rely essentially on the behavior of the denominator in (6.1). We will address the second in Section 6.2. The first relies on the fact that each $\mathbf{y}_k^\dagger \mathbf{y}_k$, converges almost surely to 1. This is true for MP-type matrices, since the constant distribution of the entries in the growing columns is necessarily Gaussian-like. In contrast, the sum of the squares of a column in $\mathbf{X}$ will converge almost surely to a random variable following a distribution depending on the driving process X_t.

The structure of the SLCE is such that each of the p independent columns represents the fluctuations of a Lévy process X_t as it is sampled at N equally spaced points over an interval $[0, T]$. These observations are given by

$$0 = X_0,\ X_{T/N},\ X_{2\cdot T/N},\ \ldots,\ X_{N\cdot T/N} = X_T$$

The i.i.d. entries $[\mathbf{X}]_{jk} \overset{d}{=} X_{T/N}$ are related to the fluctuations over the j^{th} subinterval due to the time-invariance of the Lévy process distribution, namely that

$$[\mathbf{X}]_{jk} \stackrel{d}{=} X_{j\cdot T/N} - X_{(j-1)\cdot T/N} \stackrel{d}{=} X_{T/N}$$

Let us consider letting the number of columns p remain fixed and take $N \to \infty$, so that $\lambda \to 0^+$. In the classical setting, the M–P law converges weakly to the point mass at 1, reflecting the fact that the sample covariance of any two columns converges to 0, while the sample variance of each column converges to 1. For a non-Gaussian Lévy process, however, this will not be the case.

Let X_t be a Lévy process; we will assume a finite fourth moment for the time being. Suppose that the mean μ of X_1 is known, but we are trying to estimate the variance σ^2 of X_1 from the $N+1$ observations of X_t described above. The corresponding sample variance $\widehat{\sigma}^2$ from these observations can be expressed by first computing the sample quadratic variation process

$$Y_T^{(N)} = \sum_{j=1}^{N} Z_j^2$$

where $Z_j = X_{j\cdot T/N} - X_{(j-1)\cdot T/N} \sim X_{T/N}$ are i.i.d. From here we note that

$$\begin{aligned}
\mathbb{E}\left[Y_T^{(N)}\right] &= N \cdot \mathbb{E}\left[X_{T/N}^2\right] \\
&= N \cdot \left(\operatorname{var}\left[X_{T/N}\right] + \mathbb{E}\left[X_{T/N}\right]^2\right) \\
&= N \cdot \left(\frac{T}{N}\sigma^2 + \frac{T^2}{N^2}\mathbb{E}\left[X\right]^2\right) \\
&= T \cdot \sigma^2 + \frac{T^2}{N}\mu^2
\end{aligned}$$

so that it is clear that the following expression for the sample variance estimator $\widehat{\sigma}^2$ is unbiased:

$$\widehat{\sigma}^2 = \frac{1}{T}\sum_{j=1}^{N}\left(Z_j - \frac{T}{N}\mu\right)^2 = \frac{1}{T}Y_T^{(N)} + \frac{\mu}{N}\left(T\mu - 2X_T\right)$$

To compute the variance of $\widehat{\sigma}^2$, we calculate

$$\begin{aligned}
\operatorname{var}\left[\widehat{\sigma}^2\right] &= \mathbb{E}[\left(\widehat{\sigma}^2\right)^2] - \mathbb{E}\left[\widehat{\sigma}^2\right]^2 \\
&= \mathbb{E}\left[\left(\frac{1}{T}\sum_{j=1}^{N}\widetilde{Z}_j^2\right)^2\right] - \sigma^4 \\
&= \frac{N}{T^2}\mathbb{E}\left[\widetilde{Z}_1^4\right] + \frac{N(N-1)}{T^2}\mathbb{E}\left[\widetilde{Z}_1^2\right]^2 - \sigma^4 \\
&= \frac{1}{T}\kappa_4 + \frac{3}{N}\sigma^4 + \sigma^4 - \frac{1}{N}\sigma^4 - \sigma^4 \\
&= \frac{1}{T}\kappa_4 + \frac{2}{N}\sigma^4
\end{aligned}$$

where $\widetilde{Z}_j = Z_j - (T/N)\mu$ and κ_4 is the fourth cumulant of X.

If X_1 is Gaussian and X_t is Brownian motion, then $\kappa_4[X_1] = 0$ and the variance of $\widehat{\sigma}^2$ is given simply by $2\sigma^4/N$. On the other hand, if X_t is any process other than Brownian motion, the fourth cumulant $\kappa_4[X_1] > 0$ will be positive. Under these circumstances, accuracy of the sample variance estimator $\widehat{\sigma}^2$ improves as the number of samples N and the horizon T are increased simultaneously.

Suppose that we fix $T > 0$ and let $N \to \infty$. Then the sample quadratic variation converges in probability to the true quadratic variation (Pascucci, 2011), which is to say that

$$Y_T^{(N)} \to [X]_T$$

where the convergence here is in probability. This similarly implies that the variance estimator $\widehat{\sigma}^2$ converges to the normalized quadratic variation process $(1/T)[X]_T$ in distribution.

Therefore, as $\lambda \to 0^+$ we expect the covariance matrix to converge to a diagonal $p \times p$ matrix whose diagonal entries are i.i.d. with distribution equal to $(1/T)[X]_T$. We will therefore be concerned with the quadratic variation of an arbitrary Lévy process X_t.

When the true mean μ is unknown, the sample variance estimator can still be constructed using the sample mean. Although the calculations are much longer, a similar result holds. Here the sample variance estimator is defined as

$$\begin{aligned}
\widehat{\sigma}^2 &= \frac{1}{T} \sum_{j=1}^{N} \left(Z_j - \frac{1}{N} \sum_{k=1}^{N} Z_k \right)^2 \\
&= \frac{N-2}{TN} Y_T^{(N)} - \frac{4}{TN} \sum_{j=2}^{N} \sum_{k<j} Z_j Z_k + \frac{1}{TN} X_T^2
\end{aligned}$$

Computing the expectation of $\widehat{\sigma}^2$ in this case, we find that $\mathbb{E}[\widehat{\sigma}^2] = \frac{N-1}{N}\sigma^2$, as expected. The computation of the variance of $\widehat{\sigma}^2$ is long, eventually leading to

$$\operatorname{var}\left[\frac{N}{N-1}\widehat{\sigma}^2\right] = \frac{1}{T}\kappa_4 + \frac{2}{N-1}\sigma^4$$

6.2 Approximation by Free Poisson Distributions

The second key step in the proof is that the spectrum of the submatrices $\mathbf{Y}_k^\dagger \mathbf{Y}_k$ are very close to the the spectrum of $\mathbf{Y}^\dagger \mathbf{Y}$. This is to say, removing individual columns of $\mathbf{Y}$ does not effect its singular values too much. This may or may not hold for our matrices $\mathbf{X}$. For instance, if X_t is a Poisson process over an exceptionally short interval $[0, T]$ then our matrix $\mathbf{X}$ will likely be sparse, and the removal of certain columns may have a large effect on its singular values. If this problem can be circumvented by restricting ourselves to a particular subclass of ID($*$), however, then we can consider the modified form of (6.1), which becomes

$$S_N(z) \approx \frac{1}{p}\sum_{k=1}^{p} \frac{1}{Z_k - z - Z_k(\lambda + \lambda z S_N(z))}$$

where Z_k are i.i.d. random variables following the distribution $\mathcal{L}([X]_T)$. In this case, the limiting distribution would have a Stieltjes transform which satisfies

$$S(z) = \int_{\mathbb{R}} \frac{1}{x - z - x(\lambda + \lambda z S(z))}\, d\nu(x)$$

where $\nu = \mathcal{L}([X]_T)$. Limiting distributions like this are precisely those occurring in Theorem 2.2.4 when considering matrix products $\mathbf{Y}\sqrt{\mathbf{T}}$ such that $\mathbf{T}$ is a $p \times p$ diagonal matrix whose i.i.d. diagonal entries follow the distribution ν. From Example 5.3.5, this has an interpretation in terms of free probability as the distribution arising from the free multiplicative convolution $\mathrm{mp}_\lambda \boxtimes \nu$. This then becomes a question of which classes of distributions will allow such a proof to go through, or rather to which class does the quadratic variation process $[X]_t$ need to belong.

We are interested then in the distributions of Lévy processes X_t such that the quadratic variation $[X]_T$ belongs to ID($*$) $\cap$ FR. Recent work in this area has suggested that there is a surprisingly large but complicated overlap (Bożejko and Hasebe, 2013; Hasebe, 2016; Morishita and Ueda, 2018) between FR and the classes of GGC and HCM distributions. In particular, Hasebe (2014) showed that many classically infinitely divisible distributions are also Freely infinitely divisible, although these properties are certainly not path dependent (and so may change for different choices of $T > 0$). For instance, the Gamma subordinator Γ_t is in ID($\boxplus$) for $t \in (0, 1/2] \cup [3/2, \infty)$, but fails to be in ID($\boxplus$) for a complicated union of intervals contained in $(1/2, 3/2)$, including the key case of the exponential distribution Exp(1) when $t = 1$. On the other hand, the inverse Gamma distributions are in ID($\boxplus$) for all values of $t > 0$. This scenario is discussed at length in Appendix B.5, where we conclude with the following conjecture (see also Figure B.5).

Conjecture 6.2.1. *Let $\mathcal{R}_\beta$ denote the class of distributions which can be written in the form X/Γ_β where $X \in \mathrm{GGC}$ and Γ_β is a Gamma subordinator independent from X, as introduced in Section 3.2.6. Recall that we have*

$$\mathcal{R}_\alpha \subseteq \mathcal{R}_\beta$$

for $0 < \alpha < \beta$, and

$$\mathcal{R}_\beta \to \mathrm{GGC}$$

as $\beta \to \infty$ (considering the weak closure). Then there is some small $0 < \alpha < 1$ such that $\mathcal{R}_\alpha \subset \mathrm{ID}(\boxplus)$.

6.2.1 Monte Carlo Simulations

We consider two examples of non-stable Lévy processes encountered in the financial modeling of asset returns. The first is the variance-gamma (VG) process X_t^{VG}, which can be realized as Brownian motion B_t subordinated to a gamma process Γ_t. This is the process such that $\Gamma_1 \stackrel{d}{=} \mathrm{Exp}(1)$, as discussed in Section 3.3.2. Consequently, $B_{\Gamma_1} \stackrel{d}{=} Z\sqrt{E}$ follows a Laplace distribution, where Z and E are independent with $Z \stackrel{d}{=} \mathrm{N}(0,1)$ and $E \stackrel{d}{=} \mathrm{Exp}(1)$. As a result, the VG process can be considered the time evolution of a Laplace distribution.

The second process we consider is the normal-inverse Gaussian (NIG) process X_t^{NIG}, which can be realized as Brownian motion subordinated to an inverse Gaussian process T_t. As discussed in Section 3.3.5, the Lévy measure for the NIG process can be expressed in terms of modified Bessel functions, but for the purpose of simulating random variables it is enough

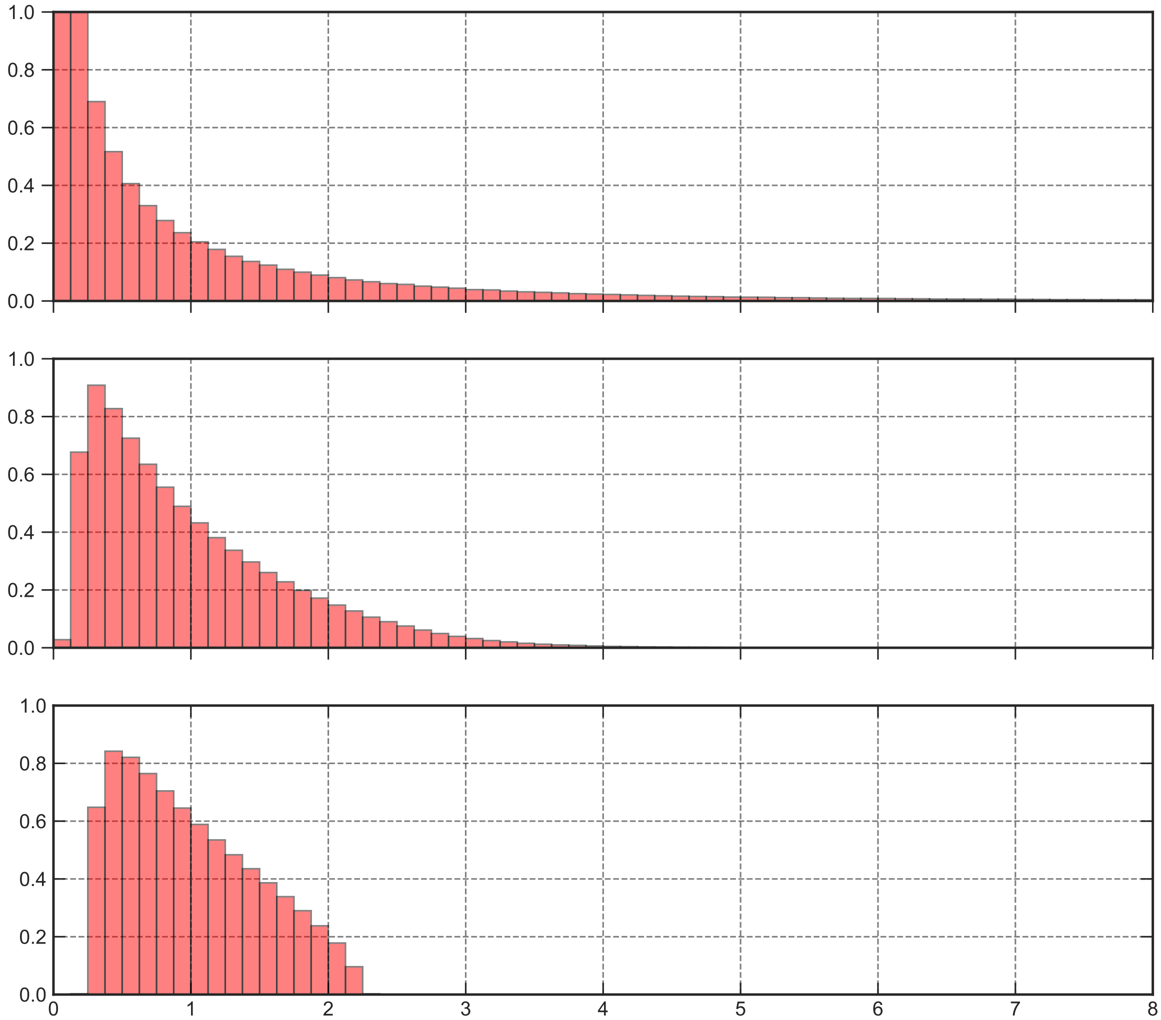


Figure 6.1: Histograms of aggregate eigenvalues of the VG SLCE when $N \times p = 2000 \times 500$, $\lambda = 1/4$, and $T = 1$ (top), $T = 10$ (middle), and $T = 100$ (bottom). The process has been normalized to have unit variance and kurtosis equal to $3/T$.

to know that

$$X_t^{\mathrm{NIG}} \stackrel{d}{=} Z\sqrt{T_t}$$

where $Z \stackrel{d}{=} \mathrm{N}(0,1)$ is independent of T_t.

Both processes are symmetric EGGCs, and so the entries of our matrices can be generated as the Hadamard product of a matrix with i.i.d. standard normal entries and one whose

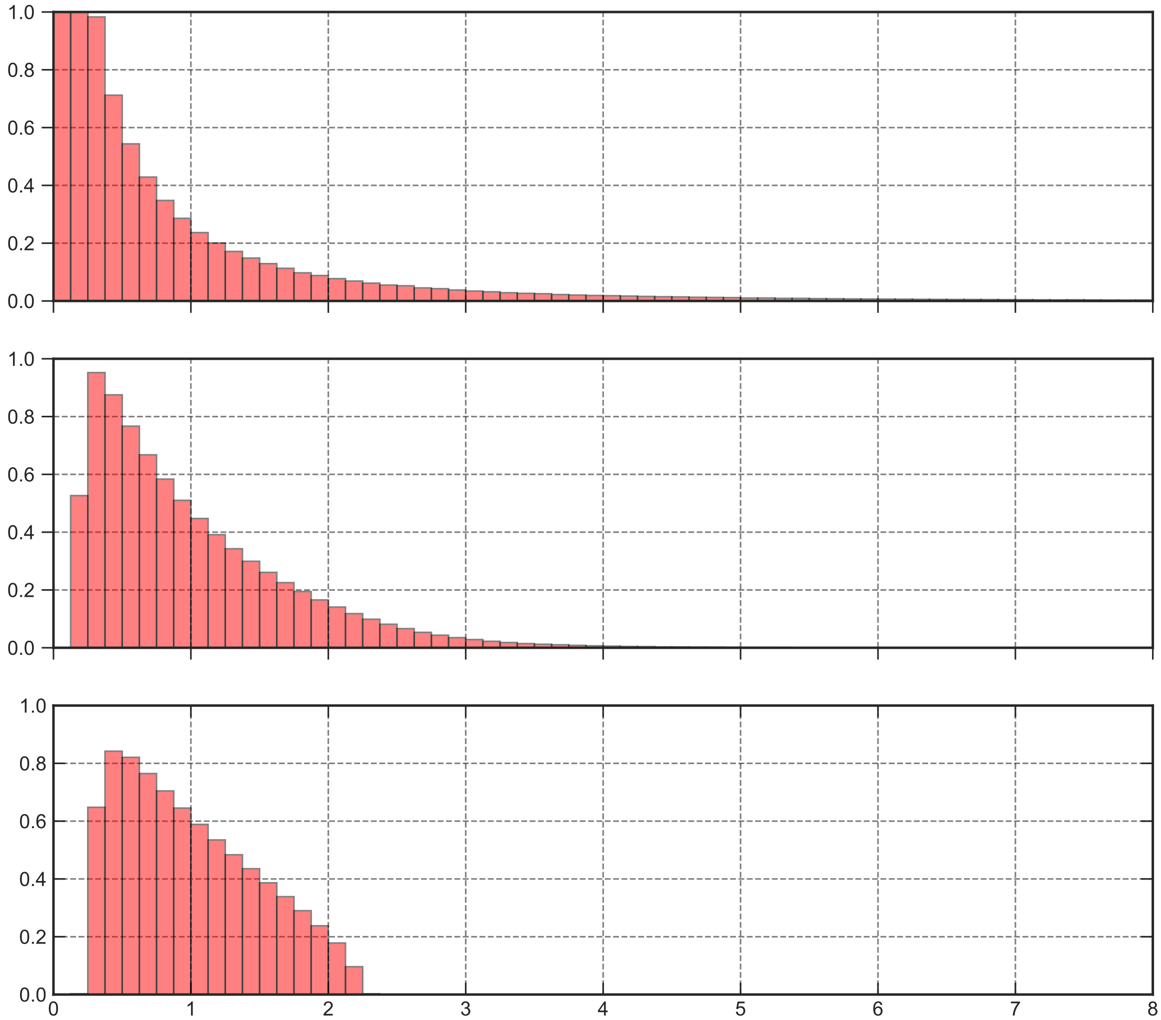


Figure 6.2: Histograms of aggregate eigenvalues of the NIG SLCE when $N \times p = 2000 \times 500$, $\lambda = 1/4$, and $T = 1$ (top), $T = 10$ (middle), and $T = 100$ (bottom). The process has been normalized to have unit variance and kurtosis equal to $3/T$.

entries are taken to be the square root of i.i.d. random variables drawn from a Gamma distribution or an Inverse Gaussian distribution, respectively. Data on the densities of the ESDs is collected by sampling matrices of various size (dependent on N), where $p = \lceil \lambda N \rceil$ for $\lambda = 1/4$. Experiments for different choices of λ show similar results; we display only $\lambda = 1/4$ for brevity, noting that the support of the MP distributions is the interval $[0.25, 2.25]$. For each choice of N, eigenvalues are aggregated from a total of $10^6/p$ Monte Carlo simulations in order to produce one million datapoints. Entries are normalized in order to be comparable to the M–P distribution mp_λ.

Table 6.1: Results of Monte Carlo Simulations for VG and NIG Driven SLCE Matrices

Entries	N	T	$\#\{\sigma : \sigma < \sigma_{\min} = 0.25\}$	$\#\{\sigma : \sigma > \sigma_{\max} = 2.25\}$	$d_{\text{K–S}}(\mu_N, m_\lambda)$	$d_{\text{K–S}}(\mu_N, \mu_{2000})$
VG	20	1	0.418500×10^6	0.121512×10^6	0.432708	0.003008
	200	1	0.419560×10^6	0.121549×10^6	0.432980	0.000919
	2000	1	0.419553×10^6	0.122387×10^6	0.433081	NA
	20	10	0.098902×10^6	0.074011×10^6	0.129765	0.013368
	200	10	0.088608×10^6	0.071761×10^6	0.121758	0.001301
	2000	10	0.088087×10^6	0.071791×10^6	0.121542	NA
NIG	20	1	0.322939×10^6	0.104397×10^6	0.362288	0.008482
	200	1	0.322989×10^6	0.104298×10^6	0.362920	0.000984
	2000	1	0.322906×10^6	0.104413×10^6	0.363221	NA
	20	10	0.081187×10^6	0.070340×10^6	0.117020	0.018738
	200	10	0.067578×10^6	0.067129×10^6	0.109260	0.002326
	2000	10	0.065810×10^6	0.066821×10^6	0.107971	NA

Results are displayed in Table 6.1, which shows that the distribution of the eigenvalues deviates significantly from the M–P distribution mp_λ. The VG process leads to sample covariance matrices which carry a huge portion of their spectrum to the left of the M–P bulk $[0.25, 2.25]$, as visualized in Figure 6.1. The normalization of both processes was chosen such that the excess kurtosis of individual entries in the matrices can be computed as $3N/T$ for both ensembles, demonstrating that a limiting distribution is affected but not exclusively determined by the fourth cumulant of X_t. Shrinkage of the spectrum when comparing $T = 1$ and $T = 10$ is expected, as the entries necessarily become more Gaussian over longer horizons. For datasets of equal size 10^6, the threshold for rejecting the null hypothesis (that the two distributions are identical) with a confidence of 99% for the K–S test is that $d_{\text{K–S}}$ exceeds

$$\sqrt{\log\left(\frac{2}{1-0.99}\right)\frac{1}{10^6}} \approx 0.002302$$

Although this is under the assumption of independent samples, the repellent behavior of eigenvalues should, if anything, increase the accuracy of the test statistic.

6.2.2 Algorithm for Free Poisson when $[X]_T$ is Unknown

Yao et al. (2015) proposes a numerical scheme for approximating densities of the form $\text{mp}_\lambda \boxtimes \nu$, which involves the computation of a modified form of (2.7) for the companion Stieltjes transform $\underline{S}(z) = -\frac{1-\lambda}{z} + \lambda S(z)$,

$$\underline{S}(z) = \frac{1}{-z + \lambda \int_{\mathbb{R}} \frac{w}{1+w\underline{S}(z)} d\nu(w)} \tag{6.2}$$

Specifically, (6.2) has a unique fixed point for all $z \in \mathbb{C}^+$, and $\frac{1}{\lambda\pi}\text{Im}[\underline{S}(x+i\epsilon)]$ converges to the continuous density of the limiting spectral distribution of $\mu_{\tilde{\mathbf{S}}\mathbf{T}}$ as $\epsilon \to 0^+$. An approximation can be produced by fixing some small $\epsilon > 0$, and iterating the map

$$s \mapsto \frac{1}{-(x+i\epsilon) + \lambda \int_{\mathbb{R}} \frac{w}{1+ws} d\nu(w)} \tag{6.3}$$

Under such a scheme, the integral $\int_{\mathbb{R}} \frac{w}{1+ws} d\nu(w)$ can be evaluated numerically when an analytic description of $\nu = \mathcal{L}([X]_T)$ is known. Consider, however, the example presented in Figure 1.1b, where the entries of a large 2000×500 matrix are i.i.d. Suzuki random variables. As discussed in Section 3.2.7, the Suzuki distribution is CGGC, so it can be realized as the distribution of a Lévy process at a fixed time. On the other hand, there is no convenient analytic expression for the distributions of the process for arbitrary $t > 0$, nor for the associated quadratic variation process. This is the case for many EGGC and CGGC processes derived from generalized inverse Gaussian distributions (Bondesson, 1992), such as the Student's-t distributions, generalized hyperbolic distributions, and skew generalized hyperbolic secant distributions (Fischer, 2014), as membership in such classes is not time-invariant.

We propose the following approach to this problem. First, fix some desired p and N. Consider the SLCE with parameters $\lambda = p/N$, $T = 1$ (for convenience), and X_t chosen so that $\mathcal{L}(X_{1/N})$ matches the desired distribution of the entries, such as a Suzuki distribution with appropriate parameters. Then, although the quadratic variation $[X]_1$ may not have an analytic description, it will be closely approximately by the sample variance of N i.i.d. samples following the distribution $X_{1/N}$ if N is large. We can now proceed by fixing some large $M \in \mathbb{N}$ and considering a large array of M samples of the sample variance

$$\widehat{\sigma}^2 = \left[\widehat{\sigma}_1^2,\ \widehat{\sigma}_2^2,\ \ldots,\ \widehat{\sigma}_M^2\right]$$

defined as

$$\frac{N}{N-1}\widehat{\sigma}_j^2 = \frac{1}{N-1}\sum_{k=1}^{N}\left(y_{jk} - \frac{1}{N}\sum_{l=1}^{N} y_{jl}\right)^2$$

where the y_{jk} for $1 \leq j \leq M$ and $1 \leq k \leq N$ are i.i.d. samples of the chosen distribution. Each $\widehat{\sigma}_j^2$ is approximately distributed according to $[X]_1$. The integral in (6.3) can now be approximated using the discrete measure

$$\widehat{\nu} = \frac{1}{M}\sum_{j=1}^{M}\delta_{\widehat{\sigma}_j^2}$$

This method is summarized in Algorithm 1.

Algorithm 1 Scheme for Approximating the Limiting Spectral Density

procedure ApproxDensity$(x, p, N, M, \text{ProbDist}, \epsilon, \epsilon')$ ▷ Approximate the density $f(x)$
  $\lambda \leftarrow p/N$
  for $j = 1, \ldots, M$ **do** ▷ Sample the N-sample variance of ProbDist
    $\widehat{\sigma}^2[j] \leftarrow \mathrm{Var}\left(\mathrm{Sample}(\mathrm{ProbDist}, \mathrm{size} = N)\right)$
  $s \leftarrow i$
  $s_{\text{last}} \leftarrow i + i\epsilon'$
  while $|s - s_{\text{last}}| \geq \epsilon'$ **do** ▷ Stop when consecutive iterations are close
    $s_{\text{last}} \leftarrow s$
    $s \leftarrow \left(-x - i\epsilon + \lambda \times M^{-1} \times \sum_{j=1}^{M}\left[\widehat{\sigma}^2[j] \times \left(1 + \widehat{\sigma}^2[j] \times s_{\text{last}}\right)^{-1}\right]\right)^{-1}$
  return $\mathrm{Im}(s)/\lambda\pi$ ▷ By Stieltjes inversion

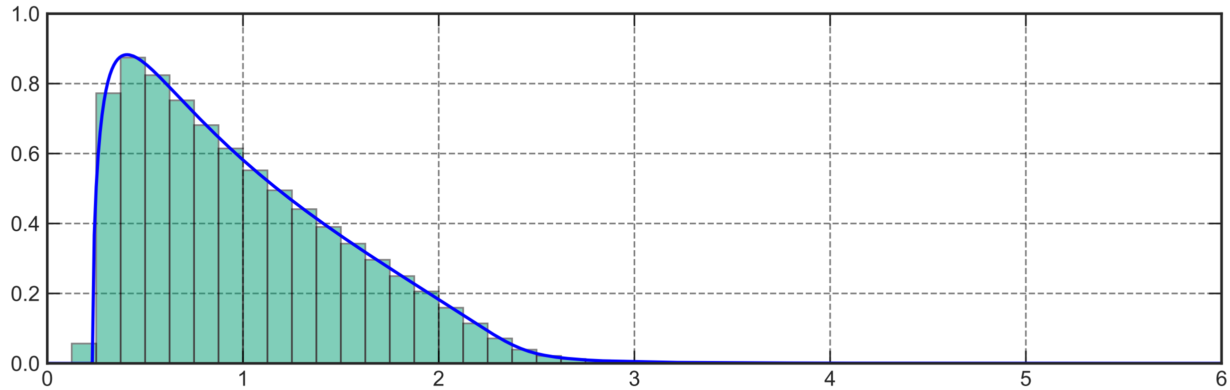


Figure 6.3: Comparison of the histogram of 10^6 eigenvalues collated from 2000 matrices of the type described in Figure 1.1b with the estimate given there as well.

We see that those outliers in Figure 1.1b are not anomalous, and the eigenvalues of matrices of this type lie in a bulk which can be quite accurately predicted by this method. The bulk no longer exhibits a right edge, as predicted by Theorem 4.2.1, and a few rare eigenvalues as large as 34, 42, and 60 (approximately) were observed in the random matrices generated. On the other hand, the smallest eigenvalues observed clustered around a left edge of about 0.22, while the approximated values of $f(x)$ jump from 9.47×10^{-7} at $x = 0.23$ to 3.68×10^{-1} at $x = 0.24$.

6.3 Applications to Financial Data

We now analyze the empirical structure of asset returns on daily and intraday timescales, and discuss an example of the similar scaling of covariance noise which occurs under the SLCE model. We consider the universe of the S&P 500 (SPX) and Nikkei 225 (NKY) indices over two timeframes: an extended daily period of June 2013 through May 2017, and a shorter intraday minute-by-minute period from January through May of 2017. The daily timeframe provides 908 datapoints for the SPX versus 895 for the NKY, while the intraday timeframe exhibits approximately 40,000 minutes containing price-changing tick data on the SPX asset versus 31,000 on those in the NKY. Linear returns are computed from prices and then standardized by factoring out the cross sectional volatility, acting as a first approximation in order to isolate stationary behavior. Expected returns are not removed through subtracting the sample mean or by more sophisticated methods, as they are many orders of magnitude smaller than the volatility. If $\mathbf{r}_j$ for $j = 1, \ldots, N$ denotes the vector of returns over the j^{th} period, then the standardized return vector $\underline{\mathbf{r}}_j$ is given by

$$\underline{\mathbf{r}}_j = \frac{\mathbf{r}_j}{\|\mathbf{r}_j\|}$$

An $N \times p$ matrix $\mathbf{X}$ is then composed of the N row vectors $\underline{\mathbf{r}}_j$, $j = 1, \ldots, N$. Figure 6.4 shows the histograms corresponding to the logarithms of the eigenvalues of the sample covariance matrix $N^{-1}\mathbf{X}^{\dagger}\mathbf{X}$.

6.3.1 Empirical Deviation from M–P Law

Figures 6.4a and 6.4c are similar to those demonstrated in previous applications of random matrix theory to such datasets (for recent examples, see Livan et al., 2011; Singh and Xu, 2016; Bun et al., 2017). Although the overlayed M–P distributions do not immediately coincide with the histograms, it is possible that a rescaling (represented by a horizontal shift of the solid red lines) might capture a decent portion of the bulk. On the other hand, a rescaling cannot widen or shrink the densities on a logarithmic plot. The constant width of the logarithmic M–P distribution is equal to

$$w_\lambda = \log_{10}\left(\left(1+\sqrt{\lambda}\right)^2\right) - \log_{10}\left(\left(1-\sqrt{\lambda}\right)^2\right) = 2\log_{10}\frac{1+\sqrt{\lambda}}{1-\sqrt{\lambda}}$$

For Figures 6.4a and 6.4c, the values w_λ are approximately 1.5914 and 0.9471, respectively. Compared to the lengths of the ticks in these figures, this is large enough to contain at least some of the bulk. For the right Figures 6.4b and 6.4d, however, the values of w_λ

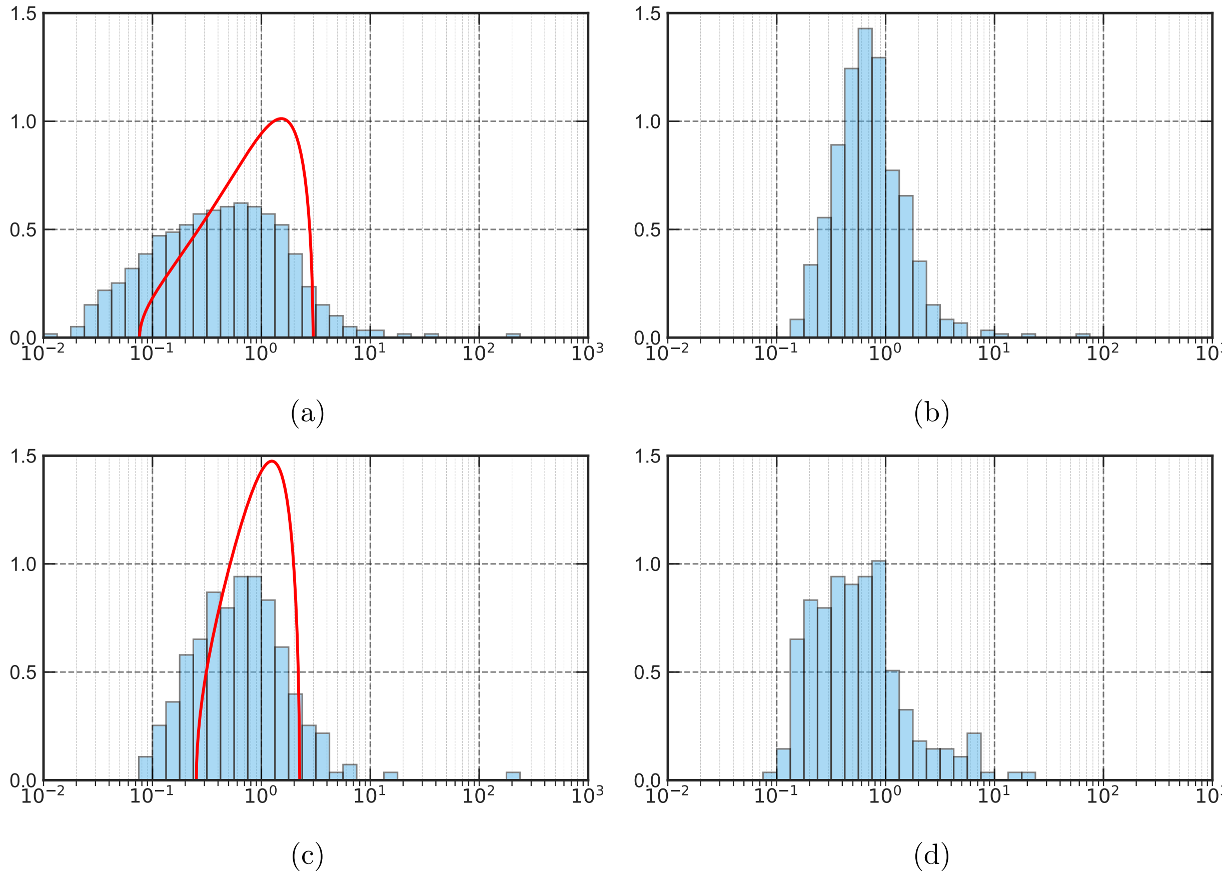


Figure 6.4: Log-plot of the empirical eigenvalues for sample covariance matrices of assets belonging to the S&P 500 (top row) and Nikkei 225 (bottom row) indices, for the periods June 2013-May 2017 (daily, left column) and January 2017-May 2017 (minute-by-minute, right column). Left column figures show the density of the logarithm of the M–P density (red) for appropriate values of λ. (a) 476 assets belonging to the S&P 500 Index with daily return data recorded from June 2013-May 2017. $N = 908$, $\lambda = 476/908 \approx 0.5242$. (b) Assets from (a) with intraday minute-by-minute data taken from January to May of 2017. $N = 40156$, $\lambda = 476/40156 \approx 0.0119$. (c) 221 assets belonging to the Nikkei 225 Index over the same period as (a), $N = 895$, $\lambda = 221/895 \approx 0.2469$. (d) Assets from (c) with intraday minute-by-minute data taken from January to May of 2017. $N = 30717$, $\lambda = 221/30717 \approx 0.0072$.

are minuscule (0.1899 and 0.1477), and cannot account for any reasonable subset of the eigenvalues observed.

6.3.2 Covariance Noise Modeling with SLCE

Modeling using SLCE given a particular Lévy process can be done by selecting appropriate values of T, N, and p. In the daily dataset, the window is approximately 900 days long with a total of $N = 900$ datapoints, while for the second it is around 100 days long with a total of 31,000 (for NKY) datapoints. As a toy model, we consider the pure noise case where returns are stationary and independent, following identical NIG processes and their corresponding ensembles. Scaling is chosen by taking $T = 900\tau$ when modeling the first window and $T = 100\tau$ when modeling the second, where $\tau = 5 \times 10^{-3}$ is chosen so that the kurtosis of the entries is of the same order of magnitude as that observed in the NKY dataset. Figures 6.5c and 6.5d show the eigenvalues of a single sample from each model, along with the density estimated according to the techniques outlined in Section 6.2.2.

Unlike the M–P case, whose bulk nearly disappears in the intraday parameter range $\lambda \approx 0.0072$, the SLCE maintains a shape similar to NKY as scaling occurs. The approximated density for the ensemble on the minute-scale in Figure 6.5d is much closer to the shape of the bulk visible in the actual NKY eigenvalues in Figure 6.5b. Under analysis motivated by the M–P distribution, one would necessarily conclude that nearly all eigenvalues in the minute-by-minute NKY data represent significant factors if all other assumptions on the returns held true, while for the SLCE it becomes unclear whether there are any at all.

The similarities between the upper a lower rows of Figure 6.5 are not the result of any complicated modeling or parameter fitting of the underlying asset behavior. The bottom figures are constructed under the (certainly false) hypothesis that fluctuations in the market are the result of complete noise, with no underlying covariance structure or factors. One interpre-

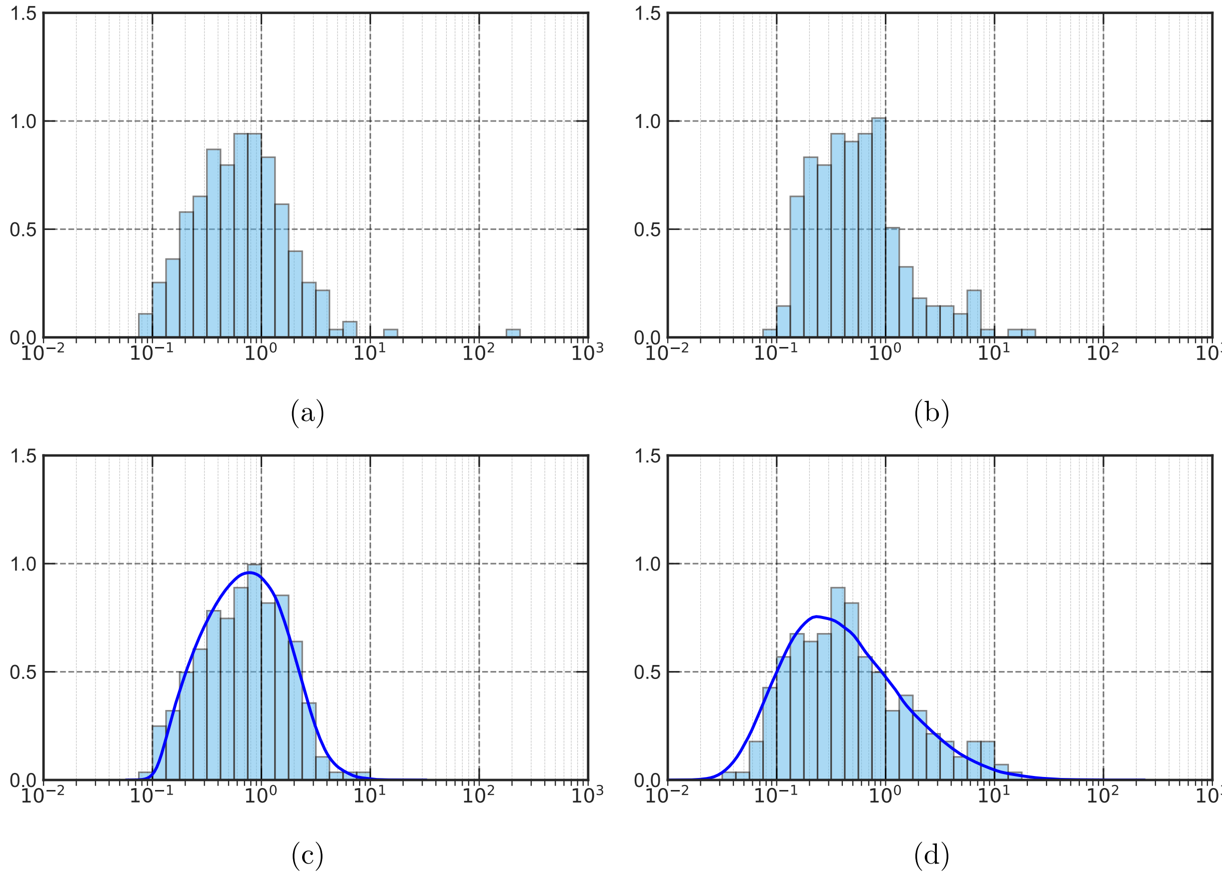


Figure 6.5: Log-plot of the empirical eigenvalues for sample covariance matrices of assets belonging to the Nikkei 225 (top row) indices, and randomly generated data (bottom row), for daily-scaled data (June 2013-May 2017, left column) and and minute-scaled data (January 2017-May 2017, right column). (a) 221 assets belonging to the Nikkei 225 Index, $N = 895$, $\lambda = 221/895 \approx 0.2469$. (b) Assets from (a) with intraday minute-by-minute data taken from January to May of 2017. $N = 30717$, $\lambda = 221/30717 \approx 0.0072$. (c) Eigenvalues from a matrix drawn from the SLCE with i.i.d. NIG entries, $T = 0.005 \times 900$, where N and p match (a), along with the estimated density for the ensemble (solid blue line). (d) Eigenvalues from a matrix drawn from the SLCE with i.i.d. NIG entries, $T = 0.005 \times 100$, where N and p match (b), along with the estimated density for the ensemble (solid blue line).

tation of the remarkable similarities is that there are significantly fewer factors present in the market than were previously inferred by modeling noisy factors in principal component analysis on the M–P law. It would be interesting to see figures produced using well fitted Lévy processes based on higher frequency data, as techniques in this area have become quite advanced (see Feng and Lin, 2013).

Bibliography

L. N. Andersen, S. Asmussen, F. Aurzada, P. W. Glynn, M. Maejima, M. Pihlsgård, and T. Simon. *Lévy matters. V*, volume 2149 of *Lecture Notes in Mathematics*. Springer, Cham, 2015. ISBN 978-3-319-23137-2; 978-3-319-23138-9. URL https://doi.org/10.1007/978-3-319-23138-9. Functionals of Lévy processes, Lévy Matters.

O. Arizmendi and T. Hasebe. Semigroups related to additive and multiplicative, free and Boolean convolutions. *Studia Math.*, 215(2):157–185, 2013. ISSN 0039-3223. doi: 10.4064/sm215-2-5. URL https://doi.org/10.4064/sm215-2-5.

O. Arizmendi and T. Hasebe. Free subordination and Belinschi-Nica semigroup. *Complex Anal. Oper. Theory*, 10(3):581–603, 2016a. ISSN 1661-8254. doi: 10.1007/s11785-015-0500-9. URL https://doi.org/10.1007/s11785-015-0500-9.

O. Arizmendi and T. Hasebe. Classical scale mixtures of Boolean stable laws. *Trans. Amer. Math. Soc.*, 368(7):4873–4905, 2016b. ISSN 0002-9947. doi: 10.1090/tran/6792. URL https://doi.org/10.1090/tran/6792.

O. Arizmendi, T. Hasebe, and N. Sakuma. On the law of free subordinators. *ALEA Lat. Am. J. Probab. Math. Stat.*, 10(1):271–291, 2013. ISSN 1980-0436.

Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010. ISBN 978-1-4419-0660-1. doi: 10.1007/978-1-4419-0661-8. URL https://doi.org/10.1007/978-1-4419-0661-8.

Z. Bai, H. Liu, and W.-K. Wong. Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory. *Math. Finance*, 19(4):639–667, 2009. ISSN 0960-1627. doi: 10.1111/j.1467-9965.2009.00383.x. URL https://doi.org/10.1111/j.1467-9965.2009.00383.x.

Z. Bai, Z. Fang, and Y.-C. Liang. *Spectral theory of large dimensional random matrices and its applications to wireless communications and finance statistics*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ; University of Science and Technology of China Press, Hefei, 2014. ISBN 978-981-4579-05-6. doi: 10.1142/9049. URL https://doi.org/10.1142/9049. Random matrix theory and its applications.

J. Baik and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.*, 97(6):1382–1408, 2006. ISSN 0047-259X. URL https://doi.org/10.1016/j.jmva.2005.08.003.

O. E. Barndorff-Nielsen. Normal inverse gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of statistics*, 24(1):1–13, 1997.

S. Belinschi, A. Dembo, and A. Guionnet. Spectral measure of heavy tailed band and covariance random matrices. *Comm. Math. Phys.*, 289(3):1023–1055, 2009a. ISSN 0010-3616. doi: 10.1007/s00220-009-0822-4. URL https://doi.org/10.1007/s00220-009-0822-4.

S. T. Belinschi. The Lebesgue decomposition of the free additive convolution of two probability distributions. *Probab. Theory Related Fields*, 142(1-2):125–150, 2008. ISSN 0178-8051. doi: 10.1007/s00440-007-0100-3. URL https://doi.org/10.1007/s00440-007-0100-3.

S. T. Belinschi and H. Bercovici. Atoms and regularity for measures in a partially defined free convolution semigroup. *Math. Z.*, 248(4):665–674, 2004. ISSN 0025-5874. doi: 10.1007/s00209-004-0671-y. URL https://doi.org/10.1007/s00209-004-0671-y.

S. T. Belinschi and A. Nica. On a remarkable semigroup of homomorphisms with respect to free multiplicative convolution. *Indiana Univ. Math. J.*, 57(4):1679–1713, 2008. ISSN 0022-2518. doi: 10.1512/iumj.2008.57.3285. URL https://doi.org/10.1512/iumj.2008.57.3285.

S. T. Belinschi, F. Benaych-Georges, and A. Guionnet. Regularization by free additive convolution, square and rectangular cases. *Complex Anal. Oper. Theory*, 3(3):611–660, 2009b. ISSN 1661-8254. doi: 10.1007/s11785-008-0080-z. URL https://doi.org/10.1007/s11785-008-0080-z.

G. Ben Arous and A. Guionnet. The spectrum of heavy tailed random matrices. *Comm. Math. Phys.*, 278(3):715–751, 2008. ISSN 0010-3616. doi: 10.1007/s00220-007-0389-x. URL https://doi.org/10.1007/s00220-007-0389-x.

F. Benaych-Georges. Rectangular random matrices, related convolution. *Probab. Theory Related Fields*, 144(3-4):471–515, 2009a. ISSN 0178-8051. doi: 10.1007/s00440-008-0152-z. URL https://doi.org/10.1007/s00440-008-0152-z.

F. Benaych-Georges. Rectangular random matrices, entropy, and Fisher's information. *J. Operator Theory*, 62(2):371–419, 2009b. ISSN 0379-4024.

F. Benaych-Georges. On a surprising relation between the Marchenko-Pastur law, rectangular and square free convolutions. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(3):644–652, 2010. ISSN 0246-0203. doi: 10.1214/09-AIHP324. URL https://doi.org/10.1214/09-AIHP324.

F. Benaych-Georges and T. Cabanal-Duvillard. Marčenko-Pastur theorem and Bercovici-Pata bijections for heavy-tailed or localized vectors. *ALEA Lat. Am. J. Probab. Math. Stat.*, 9(2):685–715, 2012. ISSN 1980-0436.

H. Bercovici. On Boolean convolutions. In *Operator theory 20*, volume 6 of *Theta Ser. Adv. Math.*, pages 7–13. Theta, Bucharest, 2006.

H. Bercovici and V. Pata. Stable laws and domains of attraction in free probability theory. *Ann. of Math. (2)*, 149(3):1023–1060, 1999. ISSN 0003-486X. doi: 10.2307/121080. URL `https://doi.org/10.2307/121080`. With an appendix by Philippe Biane.

D. Berend and T. Tassa. Improved bounds on Bell numbers and on moments of sums of random variables. *Probab. Math. Statist.*, 30(2):185–205, 2010. ISSN 0208-4147.

P. J. Bickel and E. Levina. Regularized estimation of large covariance matrices. *Ann. Statist.*, 36(1):199–227, 2008. ISSN 0090-5364. URL `https://doi.org/10.1214/009053607000000758`.

P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. ISBN 0-471-19745-9. doi: 10.1002/9780470316962. URL `https://doi.org/10.1002/9780470316962`. A Wiley-Interscience Publication.

V. I. Bogachev. *Measure theory. Vol. I, II*. Springer-Verlag, Berlin, 2007. ISBN 978-3-540-34513-8; 3-540-34513-2. doi: 10.1007/978-3-540-34514-5. URL `https://doi.org/10.1007/978-3-540-34514-5`.

R. Bøgvad and B. Shapiro. On mother body measures with algebraic Cauchy transform. *Enseign. Math.*, 62(1-2):117–142, 2016. ISSN 0013-8584. doi: 10.4171/LEM/62-1/2-8. URL `https://doi.org/10.4171/LEM/62-1/2-8`.

L. Bondesson. *Generalized gamma convolutions and related classes of distributions and densities*, volume 76 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1992. ISBN 0-387-97866-6. doi: 10.1007/978-1-4612-2948-3. URL `https://doi.org/10.1007/978-1-4612-2948-3`.

L. Bondesson. A class of probability distributions that is closed with respect to addition as well as multiplication of independent random variables. *J. Theoret. Probab.*, 28(3):1063–1081, 2015. ISSN 0894-9840. doi: 10.1007/s10959-013-0523-y. URL `https://doi.org/10.1007/s10959-013-0523-y`.

P. Bosch and T. Simon. A proof of Bondesson's conjecture on stable densities. *Ark. Mat.*, 54(1):31–38, 2016. ISSN 0004-2080. doi: 10.1007/s11512-015-0216-0. URL `https://doi.org/10.1007/s11512-015-0216-0`.

J.-P. Bouchaud and M. Potters. *Theory of financial risks*. Cambridge University Press, Cambridge, 2000. ISBN 0-521-78232-5. From statistical physics to risk management, With a foreword by Nick Dunbar.

J.-P. Bouchaud and M. Potters. Financial applications of random matrix theory: a short review. In *The Oxford handbook of random matrix theory*, pages 824–850. Oxford Univ. Press, Oxford, 2011.

M. Bożejko and T. Hasebe. On free infinite divisibility for classical Meixner distributions. *Probab. Math. Statist.*, 33(2):363–375, 2013. ISSN 0208-4147.

J. Bun, J.-P. Bouchaud, and M. Potters. Cleaning large correlation matrices: tools from random matrix theory. *Phys. Rep.*, 666:1–109, 2017. ISSN 0370-1573. doi: 10.1016/j.physrep.2016.10.005. URL `https://doi.org/10.1016/j.physrep.2016.10.005`.

Z. a. Burda, A. Jarosz, M. A. Nowak, J. Jurkiewicz, G. Papp, and I. Zahed. Applying free random variables to random matrix analysis of financial data. Part I: The Gaussian case. *Quant. Finance*, 11(7):1103–1124, 2011. ISSN 1469-7688. doi: 10.1080/14697688.2010.484025. URL `https://doi.org/10.1080/14697688.2010.484025`.

P. Carr, H. Geman, D. B. Madan, and M. Yor. The fine structure of asset returns: An empirical investigation. *The Journal of Business*, 75(2):305–332, 2002.

G. P. Chistyakov and F. Götze. The arithmetic of distributions in free probability theory. *Cent. Eur. J. Math.*, 9(5):997–1050, 2011. ISSN 1895-1074. doi: 10.2478/s11533-011-0049-4. URL `https://doi.org/10.2478/s11533-011-0049-4`.

Y.-G. Choi, J. Lim, and S. Choi. High-dimensional Markowitz portfolio optimization problem: empirical comparison of covariance matrix estimators. *J. Stat. Comput. Simul.*, 89(7):1278–1300, 2019. ISSN 0094-9655. doi: 10.1080/00949655.2019.1577855. URL `https://doi.org/10.1080/00949655.2019.1577855`.

R. Couillet and M. Debbah. *Random matrix methods for wireless communications*. Cambridge University Press, 2011.

N. El Karoui. On the realized risk of high-dimensional Markowitz portfolios. *SIAM J. Financial Math.*, 4(1):737–783, 2013. ISSN 1945-497X. doi: 10.1137/090774926. URL `https://doi.org/10.1137/090774926`.

J. Eriksson, E. Ollila, and V. Koivunen. Statistics for complex random variables revisited. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 3565–3568. IEEE, 2009.

L. Feng and X. Lin. Inverting analytic characteristic functions and financial applications. *SIAM J. Financial Math.*, 4(1):372–398, 2013. ISSN 1945-497X. URL `https://doi.org/10.1137/110830319`.

M. J. Fischer. *Generalized hyperbolic secant distributions*. SpringerBriefs in Statistics. Springer, Heidelberg, 2014. ISBN 978-3-642-45137-9; 978-3-642-45138-6. URL `https://doi.org/10.1007/978-3-642-45138-6`. With applications to finance.

S. Fourati. alpha-stable densities are hcm for alpha\in] 0, 1/4] and [1/3, 1/2]. *arXiv preprint arXiv:1309.1045*, 2013.

U. Franz. Monotone and Boolean convolutions for non-compactly supported probability measures. *Indiana Univ. Math. J.*, 58(3):1151–1185, 2009. ISSN 0022-2518. doi: 10.1512/iumj.2009.58.3578. URL `https://doi.org/10.1512/iumj.2009.58.3578`.

D. P. Gaver and P. A. W. Lewis. First-order autoregressive gamma sequences and point processes. *Adv. in Appl. Probab.*, 12(3):727–745, 1980. ISSN 0001-8678. doi: 10.2307/1426429. URL `https://doi.org/10.2307/1426429`.

T. Hasebe. Levy processes and their distributions in terms of independence (mathematical studies on independence and dependence structure: A functional analytic point of view). 2012.

T. Hasebe. Free infinite divisibility for beta distributions and related ones. *Electron. J. Probab.*, 19:no. 81, 33, 2014. ISSN 1083-6489. doi: 10.1214/EJP.v19-3448. URL `https://doi.org/10.1214/EJP.v19-3448`.

T. Hasebe. Free infinite divisibility for powers of random variables. *ALEA Lat. Am. J. Probab. Math. Stat.*, 13(1):309–336, 2016. ISSN 1980-0436.

T. Hasebe, T. Simon, and M. Wang. Some properties of free stable distributions. *arXiv preprint arXiv:1805.01133*, 2018.

M. Jeanblanc, M. Yor, and M. Chesney. *Mathematical methods for financial markets*. Springer Finance. Springer-Verlag London, Ltd., London, 2009. ISBN 978-1-85233-376-8. URL `https://doi.org/10.1007/978-1-84628-737-4`.

E. Jondeau, S.-H. Poon, and M. Rockinger. *Financial modeling under non-Gaussian distributions*. Springer Finance. Springer-Verlag London, Ltd., London, 2007. ISBN 978-1-84628-419-9; 1-84628-419-8.

E. Kaniuth. *A course in commutative Banach algebras*, volume 246 of *Graduate Texts in Mathematics*. Springer, New York, 2009. ISBN 978-0-387-72475-1. doi: 10.1007/978-0-387-72476-8. URL `https://doi.org/10.1007/978-0-387-72476-8`.

I. Koponen. Analytic approach to the problem of convergence of truncated lévy flights towards the gaussian stochastic process. *Physical Review E*, 52(1):1197, 1995.

L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters. Noise dressing of financial correlation matrices. *Physical review letters*, 83(7):1467, 1999.

Y. Liu, N. H. Chan, C. T. Ng, and S. P. S. Wong. Shrinkage estimation of mean-variance portfolio. *Int. J. Theor. Appl. Finance*, 19(1):1650003, 25, 2016. ISSN 0219-0249. doi: 10.1142/S0219024916500035. URL `https://doi.org/10.1142/S0219024916500035`.

G. Livan, S. Alfarano, and E. Scalas. Fine structure of spectral properties for random correlation matrices: An application to financial markets. *Physical Review E*, 84(1):016113, 2011.

E. Lukacs. *Characteristic functions*. Hafner Publishing Co., New York, 1970. Second edition, revised and enlarged.

D. B. Madan and E. Seneta. The variance gamma (vg) model for share market returns. *Journal of business*, pages 511–524, 1990.

M. Maejima. Classes of infinitely divisible distributions and examples. In *Lévy Matters V*, pages 1–65. Springer, 2015.

R. N. Mantegna and H. E. Stanley. Stochastic process with ultraslow convergence to a Gaussian: the truncated Lévy flight. *Phys. Rev. Lett.*, 73(22):2946–2949, 1994. ISSN 0031-9007. doi: 10.1103/PhysRevLett.73.2946. URL `https://doi.org/10.1103/PhysRevLett.73.2946`.

R. N. Mantegna and H. E. Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376(6535):46, 1995.

V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114):507–536, 1967.

A. Meucci. *Risk and asset allocation*. Springer Science & Business Media, 2009.

J. Morishita and Y. Ueda. Free infinite divisibility for generalized power distributions with free poisson term. *arXiv preprint arXiv:1806.07738*, 2018.

J. Parsons. Characterisation of multipath phenomena. *The Mobile Radio Propagation Channel, Second Edition*, pages 114–163, 2000.

A. Pascucci. *PDE and martingale methods in option pricing*, volume 2 of *Bocconi & Springer Series*. Springer, Milan; Bocconi University Press, Milan, 2011. ISBN 978-88-470-1780-1. doi: 10.1007/978-88-470-1781-8. URL `https://doi.org/10.1007/978-88-470-1781-8`.

D. Paul and A. Aue. Random matrix theory in statistics: a review. *J. Statist. Plann. Inference*, 150:1–29, 2014. ISSN 0378-3758. URL `https://doi.org/10.1016/j.jspi.2013.09.005`.

V. Pérez-Abreu and N. Sakuma. Free infinite divisibility of free multiplicative mixtures of the Wigner distribution. *J. Theoret. Probab.*, 25(1):100–121, 2012. ISSN 0894-9840. doi: 10.1007/s10959-010-0288-5. URL `https://doi.org/10.1007/s10959-010-0288-5`.

M. Politi, E. Scalas, D. Fulger, and G. Germano. Spectral densities of Wishart-Lévy free stable random matrices: analytical results and Monte Carlo validation. *Eur. Phys. J. B*, 73(1):13–22, 2010. ISSN 1434-6028. doi: 10.1140/epjb/e2009-00360-7. URL `https://doi.org/10.1140/epjb/e2009-00360-7`.

Y. Rakvongthai, A. P. N. Vo, and S. Oraintara. Complex Gaussian scale mixtures of complex wavelet coefficients. *IEEE Trans. Signal Process.*, 58(7):3545–3556, 2010. ISSN 1053-587X. doi: 10.1109/TSP.2010.2046698. URL `https://doi.org/10.1109/TSP.2010.2046698`.

V. A. Rohlin. On the fundamental ideas of measure theory. *Amer. Math. Soc. Translation*, 1952(71):55, 1952. ISSN 0065-9290.

K.-i. Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2013. ISBN 978-1-107-65649-9. Translated from the 1990 Japanese original, Revised edition of the 1999 English translation.

J. W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.*, 55(2):331–339, 1995. ISSN 0047-259X. doi: 10.1006/jmva.1995.1083. URL https://doi.org/10.1006/jmva.1995.1083.

A. Singh and D. Xu. Random matrix application to correlations amongst the volatility of assets. *Quant. Finance*, 16(1):69–83, 2016. ISSN 1469-7688. URL https://doi.org/10.1080/14697688.2015.1014400.

R. S. Strichartz. *A guide to distribution theory and Fourier transforms*. World Scientific Publishing Co., Inc., River Edge, NJ, 2003. ISBN 981-238-430-8. doi: 10.1142/5314. URL https://doi.org/10.1142/5314. Reprint of the 1994 original [CRC, Boca Raton; MR1276724 (95f:42001)].

H. Suzuki. A statistical model for urban radio propogation. *IEEE Transactions on communications*, 25(7):673–680, 1977.

O. Thorin. On the infinite divisibility of the lognormal distribution. *Scand. Actuar. J.*, (3):121–148, 1977a. ISSN 0346-1238. URL https://doi.org/10.1080/03461238.1977.10405635.

O. Thorin. On the infinite divisibility of the Pareto distribution. *Scand. Actuar. J.*, (1): 31–40, 1977b. ISSN 0346-1238.

N. G. Ushakov. *Selected topics in characteristic functions*. Modern Probability and Statistics. VSP, Utrecht, 1999. ISBN 90-6764-307-6. doi: 10.1515/9783110935981. URL https://doi.org/10.1515/9783110935981.

D. V. Voiculescu, K. J. Dykema, and A. Nica. *Free random variables*, volume 1 of *CRM Monograph Series*. American Mathematical Society, Providence, RI, 1992. ISBN 0-8218-6999-X. A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups.

J. Voit. *The statistical mechanics of financial markets*. Texts and Monographs in Physics. Springer-Verlag, Berlin, third edition, 2005. ISBN 978-3-540-26285-5; 3-540-26285-7.

J. Yao, S. Zheng, and Z. Bai. *Large sample covariance matrices and high-dimensional data analysis*, volume 39 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York, 2015. ISBN 978-1-107-06517-8. doi: 10.1017/CBO9781107588080. URL https://doi.org/10.1017/CBO9781107588080.

P. Yaskov. Necessary and sufficient conditions for the Marchenko-Pastur theorem. *Electron. Commun. Probab.*, 21:Paper No. 73, 8, 2016a. ISSN 1083-589X. URL https://doi.org/10.1214/16-ECP4748.

P. Yaskov. A short proof of the Marchenko-Pastur theorem. *C. R. Math. Acad. Sci. Paris*, 354 (3):319–322, 2016b. ISSN 1631-073X. URL https://doi.org/10.1016/j.crma.2015.12.008.

V. M. Zolotarev. *One-dimensional stable distributions*, volume 65 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1986. ISBN 0-8218-4519-5. Translated from the Russian by H. H. McFaden, Translation edited by Ben Silver.

Appendix A

Additional Probabilistic Topics

A.1 Probabilistic Framework for Random Variables and Paths

We assume that the reader is familiar with the measure theoretic foundations of probability theory. Our goal is to establish a simple framework for dealing with countable collections of random variables, vectors, sequences, and càdlàg (right continuous with left limits) stochastic processes, which is robust enough to guarantee conditional probability densities. In this sense, it is sufficient to consider a single probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taken to be standard in the sense of Rohlin (1952). We remind the reader that Ω here is the set of possible outcomes, $\mathcal{F}$ is the collection of measurable events, and $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a probability measure defined for each event. We take our random objects to be measurable maps $X : \Omega \to F$ where F is a Polish space, a topological space homeomorphic to a complete separable metric space, equipped with its Borel σ-algebra. The interested reader may consult Bogachev (2007, Def. 9.4.6 and Ch. 10) for the precise definition of a standard probability space and its relation to the existence of conditional probabilities, which we will use throughout. Filtrations on the space can be constructed once a stochastic process is given, but will typically be unnecessary.

For a random obect $X : \Omega \to F$, the choice of the Polish space F determines the type of object under consideration. If $n = 1, 2, \ldots, \infty$ and F is taken to be $\mathbb{R}^n$ or $\mathbb{C}^n$, equipped with the topology of componentwise convergence, then X is understood to be a random variable, vector, or sequence. In order to consider càdlàg stochastic processes, we fix a complete separable metric space E and let $\mathrm{D}(\mathbb{R}^+, E)$ denote the space of functions from $\mathbb{R}^+ = [0, \infty)$ into E which are right continuous with left limits at each point $t \in [0, \infty)$. The space $\mathrm{D}(\mathbb{R}^+, E)$ is called the Skorokhod space, and can be equipped with a metric under which it is itself a complete separable metric space. For a definition of the metric and further details, see Billingsley (1999, Ch. 3). A measurable function $X : \Omega \to \mathrm{D}(\mathbb{R}^+, E)$ is called a càdlàg stochastic process with values in E. Under the topology induced on $\mathrm{D}(\mathbb{R}_+, E)$, the cylindrical projections

$$X_t(\omega) \triangleq X(\omega)(t)$$

are measurable, and so for each $t \in [0, \infty)$ the function $X_t : \Omega \to E$ is a random variable. We mention that the space of continuous functions from $\mathbb{R}^+$ into E, denoted by $\mathrm{C}(\mathbb{R}^+, E)$, is a closed subspace of $\mathrm{D}(\mathbb{R}^+, E)$. Furthermore, the Skorokhod topology on $\mathrm{D}(\mathbb{R}^+, E)$ restricted to $\mathrm{C}(\mathbb{R}^+, E)$ coincides with the topology of uniform convergence on compact subsets $[0, T] \subseteq [0, \infty)$.

In this framework, all random object throughout the document can be understood to have been drawn from a single collection $\{X_{(j)}\}_{j \in \mathcal{I}}$, where $\mathcal{I}$ is some countable index set. Each object is a measurable map $X_{(j)} : \Omega \to F_j$, where each codomain F_j is Polish. The F_j are naturally probability spaces when equipped with their Borel σ-algebras $\mathcal{B}(F_j)$ and the pushforward measures

$$\mu_j(E) \triangleq \mathbb{P}[X_{(j)}^{-1}(E)], \quad E \in \mathcal{B}(F_j)$$

If $X_{(j)}$ is any random object in our collection, then by Bogachev (2007, Cor 10.4.6) there exists a system of regular conditional measures μ_j^y for $y \in F_j$ such that each μ_j^y is a probability measure on $(\Omega, \mathcal{F})$ and

$$\mathbb{P}[A \cap X_{(j)}^{-1}(E_j)] = \int_E \mu_j^y(A)\ d\mu_j(y), \quad A \in \mathcal{F},\ E_j \in \mathcal{B}(F_j)$$

Consequently, if $X_{(k)}$ is any other random object, it becomes meaningful to write the conditional probability

$$\mathbb{P}[X_{(k)} \in E_k \mid X_{(j)} = y] \triangleq \mu_j^y(X_{(k)}^{-1}(E_k)), \quad E_k \in \mathcal{B}(F_k),\ y \in \operatorname{supp}(\mu_j) \subseteq F_j$$

Throughout the text, all uses of the phrase *almost surely* refer to the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Many theorems contain statements to the effect that the empirical spectral distribution $\mu_{\mathbf{M}_N} \in \mathcal{P}(\mathbb{R})$ of a $p(N) \times p(N)$ Hermitian matrix $\mathbf{M}_N$ indexed by $N \in \mathbb{N}$ converges in distribution to a probability measure $\mu \in \mathcal{P}(\mathbb{R})$ almost surely. Here we understand the entries $[\mathbf{M}_N]_{i,j}$ for $1 \leq i, j \leq p(N)$ as random variables on our underlying space $(\Omega, \mathcal{F}, \mathbb{P})$. Since convergence in distribution is metrizable, we can consider some appropriate metric $d : \mathcal{P}(\mathbb{R}) \times \mathcal{P}(\mathbb{R}) \to \mathbb{R}^+$, and define a sequence of random variables Δ_N

$$\Delta_N = d(\mu_{\mathbf{M}_N}, \mu)$$

The statement that $\mu_{\mathbf{M}_N} \xrightarrow{d} \mu$ almost surely is equivalent to

$$\mathbb{P}\left[\lim_{N\to\infty} \Delta_N = 0\right] = 1$$

A.2 Topologies on the Space of Probability Measures

If F is a Polish space, then we let $\mathrm{C}(F)$ denote the class of continuous functions from F to $\mathbb{C}$. In order to form a Banach space in a general setting, we can restrict ourselves to the subclass $\mathrm{C}_b(F) \subseteq \mathrm{C}(F)$ of bounded continuous functions, which can be equipped with the norm

$$\|f\|_\infty = \sup_{x \in F} |f(x)|$$

where $f \in \mathrm{C}(F)$. If F is additionally locally compact, then we can consider the further subclass $\mathrm{C}_0(A)$ of $\mathbb{C}$-valued continuous functions on F that vanish at infinity, so that for each $f \in \mathrm{C}_0(F)$ and any $\epsilon > 0$ there exists some compact subset $K_{f,\epsilon} \subseteq F$ such that

$$\sup_{x \in F \setminus K_{f,\epsilon}} |f(x)| < \epsilon$$

The space $\mathrm{C}_0(F)$ is a closed subspace of $\mathrm{C}_b(F)$ under the same norm, coinciding when F is itself compact. We let $\mathrm{C}_0(F)^+$ denote the functions whose range is strictly nonnegative real numbers. We note that, in addition to being a Banach space, $\mathrm{C}_0(F)$ is also a C^*-algebra by taking the algebra multiplication to be pointwise multiplication and involution to be pointwise conjugation. The positive elements of a C^*-algebra are those elements whose spectrum is a subset of the nonnegative reals. For $\mathrm{C}_0(F)$, this coincides with $\mathrm{C}_0(F)^+$.

The dual space of $\mathrm{C}_0(F)$ can be described in terms of Borel measures on F. We will henceforth use the word measure to refer to a $\mathbb{C}$-valued, σ-additive function μ on the Borel subsets of F which can be expressed as

$$\mu = \mu_1 - \mu_2 + i(\mu_3 - \mu_4)$$

where each μ_i is a finite (positive) Radon measure. The support of a measure μ, denoted supp(μ), is the union of the supports of the μ_i. The total variation of μ is the quantity

$$|\mu|(F) = \sup \sum_j |\mu(F_j)|$$

where the supremum is taken over all countable collections of disjoint Borel subsets of F. We denote the vector-space of all such measures as $\mathcal{M}(F)$. The Riesz—Markov theorem states that the Banach space dual of $\mathrm{C}_0(F)$ is isomorphic to $\mathcal{M}(F)$ through the bilinear pairing

$$\langle \cdot, \cdot \rangle : \mathrm{C}_0(F) \times \mathrm{C}_0(F)^* \to \mathbb{C}$$

$$\langle f, \mu \rangle = \int_F f \ d\mu$$

The norm on $\mathcal{M}(F)$ as a Banach dual space agrees with the total variation.

Definition A.2.1. Let μ_n denote a sequence of finite Radon measures on the space F. We say that μ_n converges vaguely to some finite Radon measure μ if it converges in the weak-$*$ topology induced by $\mathrm{C}_0(F)$, which is to say if for any $f \in \mathrm{C}_0(F)$ we have

$$\int_F f \ d\mu_n \to \int_F f \ d\mu$$

Similarly, we say that μ_n converges weakly (or in distribution) to some finite Radon measure μ if the same condition holds for all $f \in \mathrm{C}_b(F)$. We denote weak convergence by $\mu_n \xrightarrow{d} \mu$ and vague convergence by $\mu_n \xrightarrow{v} \mu$.

As $\mathrm{C}_0(F) \subseteq \mathrm{C}_b(F)$, weak convergence implies vague convergence, but not conversely.

Definition A.2.2. We say that a measure μ is a

- positive measure or in $\mathcal{M}(F)^+$ if $\langle f, \mu \rangle \geq 0$ for any $f \in \mathrm{C}_0(F)^+$.
- sub-probability measure or in $\mathcal{M}_{\leq 1}(F)$ if it is positive and has total variation less than or equal to 1.

- ▶ probability measure or in $\mathcal{M}_1(F)$ if it is positive and has total variation equal to 1.

Lemma A.2.3. *The set of sub-probability measures $\mathcal{M}_{\leq 1}(F)$ equipped with the vague topology is compact and metrizable. In particular, any sequence of sub-probability measures has a subsequence that converges vaguely to another sub-probability measure.*

Proof. The space $\mathrm{C}_0(F)$ is a separable Banach space whose dual is $\mathcal{M}(F)$. By the Banach—Alaoglu theorem, the closed unit ball in $\mathcal{M}(F)$ is compact in the weak-$*$ topology and metrizable.

Now the closed unit ball in $\mathcal{M}(F)$ is the set of all elements whose total variation is less than or equal to 1, its positive elements being the measures $\mathcal{M}_{\leq 1}(F)$. Since the vague limit of positive measures is positive, the set $\mathcal{M}_{\leq 1}(F)$ is a closed subset of the closed unit ball, making it compact and metrizable. ■

Lemma A.2.4. *Suppose a sequence of probability measures $\mu_n \in \mathcal{M}_1(F)$ converges vaguely to a probability measure $\mu \in \mathcal{M}_1(F)$. Then μ_n converges to μ weakly, which is to say in distribution.*

Proof. If F is compact then $\mathrm{C}_b(F) = \mathrm{C}_0(F)$, so weak convergence and vague convergence are identical and there is nothing to show. Otherwise, since F is locally compact and separable, there exists an exhausting sequence of nested compact sets $K_1 \subset K_2 \subset \ldots$ such that $\bigcup_j K_j = F$.

Fix some $\epsilon > 0$. Since μ is Radon and thus inner regular, it follows that there exists some j_ϵ such that $\mu(K_{j_\epsilon}) \geq 1 - \epsilon$. By the Tietze extension theorem, for any such $\epsilon > 0$ there exists some continuous function f_{j_ϵ} taking only values in $[0, 1]$ which is compactly supported on $K_{j_\epsilon+1}$ and takes the value 1 at all points in K_{j_ϵ}. Since μ_n converges to μ vaguely and

$f_{j_\epsilon} \in \mathrm{C}_0(E)$ it follows that

$$\begin{aligned}\mu_n(K_{j_\epsilon+1}) = \int_{K_{j_\epsilon+1}} d\mu_n \geq \int_{K_{j_\epsilon+1}} f_{j_\epsilon}\, d\mu_n \\ \to \int_{K_{j_\epsilon+1}} f_{j_\epsilon}\, d\mu \geq \int_{K_{j_\epsilon}} d\mu = \mu(K_{j_\epsilon}) \geq 1-\epsilon\end{aligned}$$

Now fix some $g \in \mathrm{C}_b(F)$. If we can show that $\int_F g\, d(\mu_n - \mu) \to 0$ then we have proved the claim. Then we have

$$\begin{aligned}\left|\int_F g\, d(\mu_n-\mu)\right| &= \left|\int_F (1 - f_{j_\epsilon+1} + f_{j_\epsilon+1}) g\, d(\mu_n-\mu)\right| \\ &\leq \left|\int_F (1 - f_{j_\epsilon+1}) g\, d(\mu_n-\mu)\right| + \left|\int_F f_{j_\epsilon+1} g\, d(\mu_n-\mu)\right| \\ &= \left|\int_{F\setminus K_{j_\epsilon+1}} (1 - f_{j_\epsilon+1}) g\, d(\mu_n-\mu)\right| + \left|\int_F f_{j_\epsilon+1} g\, d(\mu_n-\mu)\right| \\ &\leq \Big(\mu_n(F\setminus K_{j_\epsilon}) + \mu(F\setminus K_{j_\epsilon})\Big)\|g\|_\infty + \left|\int_F f_{j_\epsilon+1} g\, d(\mu_n-\mu)\right| \\ &\leq \Big(\mu_n(F\setminus K_{j_\epsilon}) + \epsilon\Big)\|g\|_\infty + \left|\int_F f_{j_\epsilon+1} g\, d(\mu_n-\mu)\right|\end{aligned}$$

In the first term $1 - f_{j_\epsilon+1}$ is supported on $F\setminus K_{j_\epsilon}$. Taking large enough n, the measure $\mu_n(F\setminus K_{j_\epsilon})$ can be made smaller than 2ϵ by the convergence described above. The second term is the integral of a $\mathrm{C}_0(F)$ function which can be made smaller than ϵ for large n since μ_n converges to μ vaguely. So for large enough n,

$$\left|\int_F g\, d(\mu_n-\mu)\right| \leq (3\|g\|_\infty + 1)\,\epsilon$$

Since $\epsilon > 0$ and $g \in \mathrm{C}_b(F)$ were arbitrary, this proves the claim. ■

Example A.2.5. The importance of the condition that the limiting distribution μ be a probability measure can be explained by the fact that a sequence of probability measures $\mu_n \in \mathcal{M}_1(F)$ can converge vaguely to a sub-probability measure $\mathcal{M}_{\leq 1}(F)$ which is not itself a probability measure, in which case Lemma A.2.4 need not hold. If $\mu_n = \delta_n$, a traveling point mass, then it is easy to see that μ_n converges vaguely to the zero measure on $\mathbb{R}$, which is clearly not a probability measure. On the other hand, μ_n does not converge to any Radon measure in the weak sense. Thus, we can understand that the weak and vague notions of convergence of probability measures coincide if we already know that the limiting measure in question is itself a probability measure.

A.3 Sampling Random Variables from Characteristic Functions

For a probability measure $\mu \in \mathcal{P}(\mathbb{R})$, the characteristic function φ_μ is defined in Section 2.1.2 and inherits the properties described therein. For a treatise on the general L^p theory, we refer the reader to Kaniuth (2009). If μ has a density f_μ, then φ_μ is simply the (conjugate) Fourier transform of μ. Since any f_μ is necessarily in $L^1(\mathbb{R})$ with respect to Lebesgue measure, the corresponding φ_μ will be in the class $\mathrm{C}_0(\mathbb{R})$ of continuous $\mathbb{C}$-valued functions which decay to zero at infinity. If we additionally have that f_μ is in $L^2(\mathbb{R})$, then φ_μ will be as well. The recovery of a law from its characteristic function can be accomplished with the following inversion theorem.

Theorem A.3.1 (Gil–Pelaez Theorem, Ushakov, 1999)**.** *If $x \in \mathbb{R}$ is a continuity point of F_μ, then*

$$F_\mu(x) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\mathrm{Im}[e^{-ixz}\varphi_\mu(z)]}{z}\, dz \tag{A.1}$$

By definition, $\varphi_\mu(-z) = \overline{\varphi_\mu(z)}$, which is to say that φ_μ is Hermitian. If we know that $\mu \in \mathcal{P}_s(\mathbb{R})$, then φ_μ is necessarily real-valued. In this case, we can make the following modification which may be used to speed up numerical computations by avoiding complex variables.

Corollary A.3.2. *If X is symmetric and x is a continuity point of F_X, then*

$$F_\mu(x) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\sin(xz)\varphi_\mu(z)}{z}\, dz \tag{A.2}$$

Now suppose that we have an analytic expression for the characteristic function φ_μ of a distribution μ, and we want to generate independent samples of μ using a random number generator. One way to do this is to sample a uniform random variable U on $[0,1]$, and then compute $F_\mu^{-1}(U) \sim \mu$. This process can be accomplished with the following algorithm:

- We choose a method to calculate $F_\mu(x)$ based on a numerical approximation for either (A.1) or (A.2). Most numerical packages incorporate sophisticated techniques to perform this task quickly and accurately. Since $|\varphi_\mu(z)|$ is often decreasing as $|z| \to \infty$ (for instance, when μ has a density) and heavily concentrated near the origin, it is recommended against calling a numerical integrator on $[0,\infty)$, as this is usually accomplished by transforming the interval into $[0,1)$ by passing through the function $z \mapsto z/(1-z)$. Instead, we compute a value $\theta_1 > 0$ such that $|\varphi_\mu(\theta_1)| < \epsilon_1$ for some small $\epsilon_1 > 0$, and then call a numerical integrator on the interval $[0,\theta_1]$. Whatever method is used, we will denote this approximation by $\widetilde{F}_\mu(x)$.

- We now discretize the function F_μ with a long $M \times 1$ vector

  $$\mathbf{F} = \begin{bmatrix} \widetilde{F}_\mu(x_1) & \widetilde{F}_\mu(x_2) & \ldots & \widetilde{F}_\mu(x_M) \end{bmatrix}^\dagger$$

  where $a_\mu = x_1 < x_2 < \ldots < x_M = b_\mu$ is some mesh for the interval $[a_\mu, b_\mu]$. Values for

a_μ and b_μ should be chosen so that $\widetilde{F}_\mu(a_\mu) < \epsilon_2$ and $\widetilde{F}_\mu(b_\mu) > 1 - \epsilon_2$ for some small $\epsilon_2 > 0$.

- Fix some granularity level $\epsilon_3 > 0$. If the values $\widetilde{F}_\mu(x_{j+1}) - \widetilde{F}_\mu(x_j) > \epsilon_3$ then the mesh above should be refined and the additional values of $\widetilde{F}_\mu$ need to be computed. Once all differences are within this threshold ϵ_3, we can proceed with the vector $\mathbf{F}$. Note that this sequence of steps only needs to be completed once, and the final $\mathbf{F}$ can be saved and used repeatedly.

- Sampling μ is now accomplished by sampling a uniform random U on $[0, 1]$, and then performing a binary search to determine for which j we have

$$\widetilde{F}_\mu(x_j) < U \leq \widetilde{F}_\mu(x_{j+1})$$

The algorithm should return the value x_{j+1}.

Appendix B

Auxiliary Operations in Free Probability

This is a companion section to Chapter 5, introducing the operations of Boolean and Monotone additive and multiplicative convolutions. These operations can be used to simplify some expressions in Free Probability, as they can be calculated in terms of the Stieltjes transforms S_μ of probability measures rather than the more complicated R-transforms that appear in the literature. The key point here is that manipulating the Stieltjes transform is simple numerically due to Stieltjes inversion (Theorem 2.2.2), whereas inverting the Stieltjes transform in order to find the R-transform may be quite difficult. Some applications to eigenvalues of random matrices are discussed. We continue with the notation of S_μ, F_μ, and η_μ from Chapter 5, as well as the spaces $\mathcal{N}$, $\mathcal{F}$, and $\mathcal{K}$. The operations of Free, Boolean, and Monotone convolutions are often called "noncommutative" due to their origins in the theory of noncommutative operator algebras. The statement that a particular function is an η-function simply means that it can be expressed as $\eta(z) = 1 - zF(1/z)$ for $z \in \mathbb{C}\backslash\mathbb{R}$ where $F \in \mathcal{F}$ is in the reciprocal Cauchy class.

B.1 Additive Convolutions on Probability Distributions

Definition B.1.1. Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ be any probability distributions, and let $s \geq 0$. Then by definition of η_μ and η_ν, the functions

$$\begin{aligned} \eta_\mu(z) + \eta_\nu(z) \\ s\eta_\mu(z) \end{aligned}$$

represent η-functions as well. By Lemma 5.1.4, there exists unique probability measures, denote by $\mu \uplus \nu, \mu^{\uplus s} \in \mathcal{P}(\mathbb{R})$, such that

$$\left.\begin{aligned} \eta_{\mu \uplus \nu}(z) &= \eta_\mu(z) + \eta_\nu(z), \\ \eta_{\mu^{\uplus s}}(z) &= s\eta_\mu(z), \end{aligned}\right. \qquad z \in \mathbb{C} \backslash \mathbb{R}$$

This makes $\uplus$ into an associative, symmetric binary operation on the space of probability measures $\mathcal{P}(\mathbb{R})$, called Boolean additive convolution, and assigns to every probability measure $\mu \in \mathcal{P}(\mathbb{R})$ a $\uplus$-semigroup $\mu^{\uplus s} \in \mathcal{P}(\mathbb{R})$ for $s \geq 0$ with $\mu^{\uplus 0} = \delta_0$ and $\mu^{\uplus 1} = \mu$, such that for any $s_1, s_2 \geq 0$,

$$\mu^{\uplus s_1} \uplus \mu^{\uplus s_2} = \mu^{\uplus (s_1 + s_2)}$$

In particular, every probability distribution is infinitely divisible with respect to Boolean additive convolution, so that we can write $\mathcal{P}(\mathbb{R}) = \mathrm{ID}(\uplus)$.

Definition B.1.2. Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$. By the definition of $\mathcal{F}$, the function $F_\mu(F_\nu(z))$ is in $\mathcal{F}$ as well. By Lemma 5.1.4, there is a unique probability measure, denoted by $\mu \rhd \nu \in \mathcal{P}(\mathbb{R})$, such that

$$F_{\mu \rhd \nu}(z) = F_\mu(F_\nu(z)), \qquad z \in \mathbb{C}^+$$

This makes $\rhd$ into an associative binary operation on the space of probability measures $\mathcal{P}(\mathbb{R})$, called Monotone additive convolution.

Definition B.1.3. Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$, and by Lemma 5.1.4 let $\zeta_1, \zeta_2 \in \mathcal{P}(\mathbb{R})$ be the unique probability measures that correspond to Z_1 and Z_2. Then by the definition of $\mathcal{F}$, the function $F_\mu(F_{\zeta_1}(z)) = F_\nu(F_{\zeta_2}(z))$ is in $\mathcal{F}$ as well. Furthermore, there is a unique probability measure, denoted by $\mu \boxplus \nu \in \mathcal{P}(\mathbb{R})$, such that

$$F_{\mu \boxplus \nu}(z) = F_\mu(F_{\zeta_1}(z)) = F_\nu(F_{\zeta_2}(z)), \qquad z \in \mathbb{C}^+$$

This makes $\boxplus$ into an associative, symmetric binary operation on the space of probability measures $\mathcal{P}(\mathbb{R})$, called Free additive convolution. Furthermore, we have the relationship

$$\mu \boxplus \nu = \mu \rhd \zeta_1 = \nu \rhd \zeta_2 = \zeta_1 \uplus \zeta_2$$

Lemma B.1.4. *Let $\mu \in \mathcal{P}(\mathbb{R})$ and $x \in \mathbb{R}$. Then classical, Free, and Monotone additive convolutions (in the correct order) of μ and δ_x all agree, such that*

$$\mu * \delta_x = \mu \boxplus \delta_x = \mu \rhd \delta_x$$

The preceding lemma is clear from the definitions of the operations. The Boolean case is slightly different, as is the Monotone additive convolution in the opposite order. Curiously, these two exceptions coincide.

Lemma B.1.5 (Franz, 2009)**.** *Let $\mu \in \mathcal{P}(\mathbb{R})$ and $x \in \mathbb{R}$. Then*

$$\delta_x \rhd \mu \stackrel{d}{=} \delta_x \uplus \mu$$

Furthermore, there exists a unique probability distribution $\nu \in \mathcal{P}(\mathbb{R})$, defined explicitly by

$$\nu = \delta_x \triangleright \mu \triangleright \delta_{-x}$$

such that

$$\nu \boxplus \delta_x = \nu \triangleright \delta_x = \delta_x \triangleright \mu$$

Although we now have two types of convolution with point-mass distributions, the following lemma shows that they commute with one another.

Lemma B.1.6. *Let $\mu \in \mathcal{P}(\mathbb{R})$ and $x, y \in \mathbb{R}$. Then we have that*

$$(\mu \uplus \delta_x) \boxplus \delta_y = (\mu \boxplus \delta_y) \uplus \delta_x$$

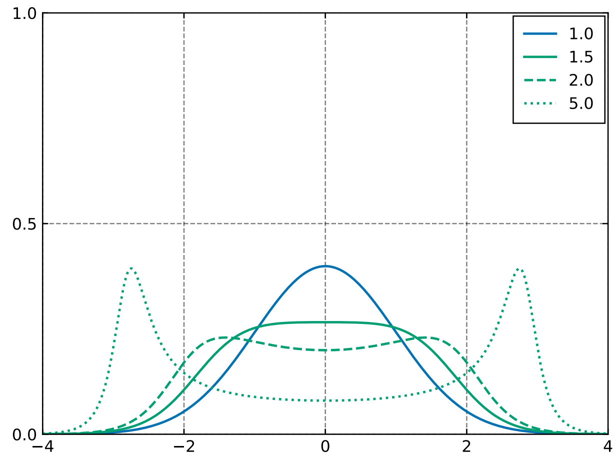
Consequently, we will omit writing the parenthesis when performing multiple operations with point mass distributions, and instead write $\delta_x \triangleright \mu \triangleright \delta_y$.

Proof. Notice that for $z \in \mathbb{C}^+$ we have

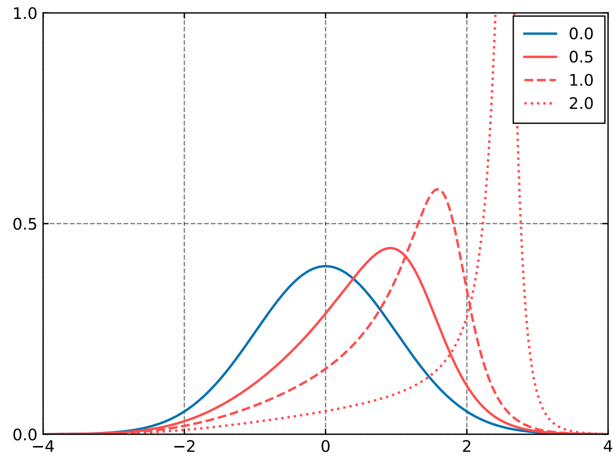
$$S_{\mu \uplus \delta_x}(z) = \frac{S_\mu(z)}{1 + x S_\mu(z)}$$

and

$$S_{\mu \boxplus \delta_y} = S_\mu(z - y)$$



(a)



(b)

Figure B.1: Approximate densities of (a) $\boldsymbol{g}^{\uplus s}$ and (b) $\delta_x \rhd \boldsymbol{g} = \boldsymbol{g} \uplus \delta_x$, for various values of $s > 0$ and $x \geq 0$. Note that $D_{1/\sqrt{s}}(\boldsymbol{g}^{\uplus s})$ converges weakly to the Rademacher distribution $\boldsymbol{r}$ as $s \to \infty$.

Therefore,

$$S_{(\mu \uplus \delta_x) \boxplus \delta_y}(z) = S_{\mu \uplus \delta_x}(z - y) = \frac{S_\mu(z-y)}{1 + x S_\mu(z-y)}$$
$$= \frac{S_{\mu \boxplus \delta_y}(z)}{1 + x S_{\mu \boxplus \delta_y}(z)} = S_{(\mu \boxplus \delta_y) \uplus \delta_x}(z)$$

■

Example B.1.7. The standard normal distribution $\boldsymbol{g} \stackrel{d}{=} \mathrm{N}(0,1)$ has Stieltjes transform

$$S_{\boldsymbol{g}}(z) = i\sqrt{\frac{\pi}{2}} e^{-z^2/2} \left(1 + \mathrm{Erf}\left(i \frac{z}{\sqrt{2}}\right)\right)$$

where Erf is the complex Error function. Figures B.1a and B.1b show the effects of the Boolean semigroup and Boolean sum involving $\boldsymbol{g}$.

B.2 Multiplicative Convolutions on Probability Distributions

Lemma B.2.1 (Arizmendi and Hasebe, 2016b). *Let $\mu \in \mathcal{P}(\mathbb{R}^+)$ and $\nu \in \mathcal{P}(\mathbb{R})$ with $\nu \neq \delta_0$. Then the function $\eta_\mu(\eta_\nu(z))$ is an η-function, and so by Lemma 5.1.6 there exists a unique probability measure, denoted by $\mu \circlearrowright \nu \in \mathcal{P}(\mathbb{R})$, such that*

$$\eta_{\mu \circlearrowright \nu}(z) = \eta_\mu(\eta_\nu(z)), \qquad z \in \mathbb{C}^+$$

Furthermore, if $\nu \in \mathcal{P}(\mathbb{R}^+)$ then $\mu \circlearrowright \nu \in \mathcal{P}(\mathbb{R}^+)$ as well, which makes $\circlearrowright$ into an associative binary operation on the space of nonnegative probability measures $\mathcal{P}(\mathbb{R}^+)$, called the Monotone multiplicative convolution.

Definition B.2.2. Suppose $\mu, \nu \in \mathcal{P}(\mathbb{R}^+)$, and that

$$\frac{\eta_\mu(z)\eta_\nu(z)}{z} \in \mathcal{K}$$

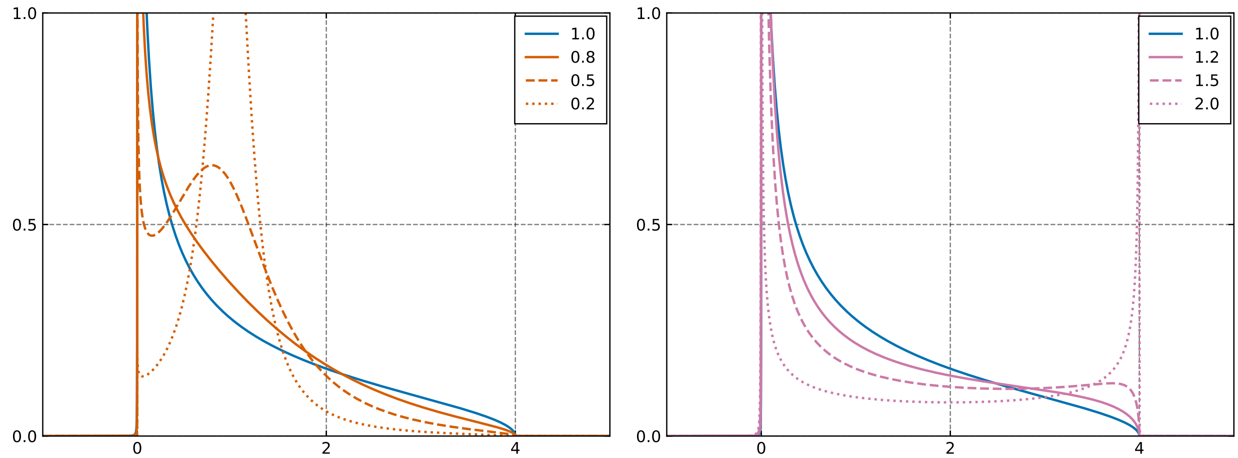
Then there exists a unique probability measure, denoted by $\mu \uplus\!\!\!\times \nu \in \mathcal{P}(\mathbb{R})$, such that

$$\eta_{\mu \uplus\!\!\times \nu}(z) = \frac{\eta_\mu(z)\eta_\nu(z)}{z}, \qquad z \in \mathbb{C}^+$$

This associative symmetric binary operation $\uplus\!\!\!\times$, when it is well defined, is called the Boolean multiplicative convolution.

Lemma B.2.3 (Bercovici, 2006). *For any $\mu \in \mathcal{P}(\mathbb{R}^+)$ and $0 \leq s \leq 1$, there exists a unique probability measure $\mu^{\uplus\!\!\times s} \in \mathcal{P}(\mathbb{R}^+)$ which forms a $\uplus\!\!\!\times$-semigroup $\mu^{\uplus\!\!\times s}$ with $\mu^{\uplus\!\!\times 0} \stackrel{d}{=} \delta_1$ and $\mu^{\uplus\!\!\times 1} \stackrel{d}{=} \mu$, such that if $0 \leq s_1, s_2 \leq 1$ with $s_1 + s_2 \leq 1$, then*

$$\mu^{\uplus\!\!\times s_1} \uplus\!\!\!\times \mu^{\uplus\!\!\times s_2} = \mu^{\uplus\!\!\times (s_1+s_2)}$$



(a) (b)

Figure B.2: Approximate densities of $\boldsymbol{\pi}^{\cup\!\!\!\times s}$ for (a) $0 < s \le 1$ and (b) $s \ge 1$. In (b), note that we have from Example B.4.6 that $\boldsymbol{\pi}^{\cup\!\!\!\times 2} = \boldsymbol{\pi} \cup\!\!\!\times \boldsymbol{\pi} = \frac{1}{2}\delta_0 + \frac{1}{2}\boldsymbol{\pi}^{\uplus 2} = \frac{1}{2}\delta_0 + \frac{1}{2}(\boldsymbol{a} \boxplus \delta_2)$.

Definition B.2.4. Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^+)$, and let $\xi_1, \xi_2 \in \mathcal{P}(\mathbb{R}^+)$ be the unique probability measures corresponding to K_1 and K_2 in Lemma 5.1.6. Then by Lemma B.2.1, there exists a unique probability measure, denoted by $\mu \boxtimes \nu \in \mathcal{P}(\mathbb{R}^+)$, such that

$$\eta_{\mu\boxtimes\nu}(z) = \eta_\mu(\eta_{\xi_1}(z)) = \eta_\nu(\eta_{\xi_2}(z)), \qquad z \in \mathbb{C}^+$$

This makes $\boxtimes$ into an associative symmetric binary operation on the space of nonnegative probability measures $\mathcal{P}(\mathbb{R}^+)$, called the Free multiplicative convolution. Furthermore, we have the relationship

$$\mu \boxtimes \nu = \mu \circlearrowright \xi_1 = \nu \circlearrowright \xi_2 = \xi_1 \cup\!\!\!\times \xi_2$$

where the final Boolean multiplicative convolution is always well defined.

B.3 Belinschi–Nica Semigroup and Free Divisibility Indicators

An early question in Free Probability was to address the issue of the proximity of a distribution to the class ID($\boxplus$), as Free infinitely divisible distributions are highly regular with analytic densities. Since regularity increases as we consider $\mu^{\boxplus t}$ for $t \geq 1$ larger and larger, a natural question to ask was how far "backwards" could you wind time: for which $0 < t < 1$ does the expression $\mu^{\boxplus t}$ exist as a distribution for a particular choice of μ? The following examples show that the situation may be non-trivial.

Example B.3.1. Let $\boldsymbol{r} = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$ be a Rademacher distribution. Like all distributions, the Free additive convolution semigroup $\boldsymbol{r}^{\boxplus t}$ exists for all $t \geq 1$. Is it possible to extend $\boldsymbol{r}^{\boxplus t}$ for $0 < t_0 \leq t \leq 1$? This amounts to showing the existence of some distribution $\tau \in \mathcal{P}(\mathbb{R})$ with a Free additive convolution semigroup $\tau^{\boxplus t}$ such that $\boldsymbol{r} \stackrel{d}{=} \tau^{\boxplus 1/t_0}$ where $1/t_0 > 1$. If this were possible, however, it would violate Theorem 5.1.16, as $\boldsymbol{r}$ has atoms at ± 1 but $\boldsymbol{r}((-1,1)) = 0$.

Example B.3.2. From the preceding example, consider $\boldsymbol{a} \stackrel{d}{=} \boldsymbol{r}^{\boxplus 2}$, where $\boldsymbol{a}$ has an Arcsine distribution on $[-2,2]$. Then it is possible to extend the Free additive convolution semigroup $\boldsymbol{a}^{\boxplus t}$ to $t \geq 1/2$, but no further. Consequently, we see that every distribution has some minimal value for which its Free additive convolution semigroup can be extended. For a Free additive infinitely divisible distribution like $\boldsymbol{\pi}$, this value is zero.

Definition B.3.3. The Belinschi–Nica semigroup corresponding to a distribution $\mu \in \mathcal{P}(\mathbb{R})$ is defined as

$$\mathbb{B}_t(\mu) = \left(\mu^{\boxplus(1+t)}\right)^{\uplus \frac{1}{1+t}}$$

where $t \geq 0$ and $\mathbb{B}_0(\mu) \stackrel{d}{=} \mu$.

Theorem B.3.4 (Belinschi and Nica, 2008)**.** *The operator* $\mathbb{B}_t(\mu)$ *defines an actual semigroup, in the sense that*

$$\mathbb{B}_t(\mathbb{B}_s(\mu)) = \mathbb{B}_{t+s}(\mu)$$

Furthermore, if $\mu, \nu \in \mathcal{P}(\mathbb{R}^+)$*, then* $\mathbb{B}_t$ *is a* $\boxtimes$ *homomorphism in the sense that*

$$\mathbb{B}_t(\mu \boxtimes \nu) = \mathbb{B}_t(\mu) \boxtimes \mathbb{B}_t(\nu)$$

The Belinschi–Nica semigroup became useful in studying Free regularity through the connection introduced in Lemma B.3.7.

Definition B.3.5. If $\circledast_1$ and $\circledast_2$ are two different operations from among $\{*, \boxplus, \uplus\}$, then we let $\Gamma_{\circledast_1 \to \circledast_2} : \mathrm{ID}(\circledast_1) \to \mathrm{ID}(\circledast_2)$ denote the bijection that takes the triplet (γ, σ, Π) in one collection to the other.

Theorem B.3.6 (Hasebe, 2012)**.** *We have that for* $\mu \in \mathcal{P}(\mathbb{R}) = \mathrm{ID}(\uplus)$*, there exists* $\gamma \in \mathbb{R}$ *and a nonnegative, finite Borel measure* τ *on* $\mathbb{R}$ *such that*

$$\eta_\mu(z) = \gamma z - \sigma^2 z^2 - \int_{\mathbb{R}} \left(\frac{1}{1 - xz} - 1 - xz \mathbb{1}_{[-1,1]}(x) \right) d\Pi(x)$$

for a triplet (γ, σ, Π) *where* $\gamma \in \mathbb{R}$*,* $\sigma \geq 0$*, and* Π *is a Lévy measure. This uniquely determines the distribution of* μ*.*

Lemma B.3.7 (Belinschi and Nica, 2008)**.** *The* $t = 1$ *Belinschi–Nica semigroup operation is the Boolean-to-Free map,* $\mathbb{B}_1 = \Gamma_{\uplus \to \boxplus}$*. In particular,* $\mu \in \mathrm{ID}(\boxplus)$ *if and only if it is in the image of* $\mathbb{B}_1(\mathcal{P}(\mathbb{R}))$*, so that there exists a unique* $\nu \in \mathcal{P}(\mathbb{R})$ *such that*

$$\nu \boxplus \nu = \nu \triangleright \mu = \mu \uplus \mu$$

As a result, an object called the Free divisibility indicator was introduced. This value, $\phi_{\boxplus}(\mu)$ was initially defined as

$$\phi_{\boxplus}(\mu) = \sup\{t \geq 0 : \mu \in \mathbb{B}_t(\mathcal{P}(\mathbb{R}))\}$$

This is to say, $\phi_{\boxplus}(\mu)$ is a measure of the largest class of Belinschi–Nica time evolutions that μ belongs to. The following connection between Boolean additive convolution and the divisibility indicator shows the effects of Free and Boolean convolution on regularity of a distribution. Most notably, the Free additive semigroup $\mu^{\boxplus t}$ is regularizing as $t \geq 1$ grows, while the Boolean additive semigroup $\mu^{\uplus}$ is regularizing as $0 < s \leq 1$ shrinks.

Theorem B.3.8 (Arizmendi and Hasebe, 2013)**.** *If $\mu \in \mathcal{P}(\mathbb{R})$ and $t > 0$, then*

$$\phi_{\boxplus}(\mu^{\uplus t}) = \frac{1}{t}\phi_{\boxplus}(\mu)$$

Consequently, $\phi_{\boxplus}(\mu)$ has two equivalent definitions

$$\phi_{\boxplus}(\mu) = \sup\{t \geq 0 : \mu \in \mathbb{B}_t(\mathcal{P}(\mathbb{R}))\} = \sup\{t \geq 0 : \mu^{\uplus t} \in \mathrm{ID}(\boxplus)\}$$

Furthermore, $\mu^{\boxplus t}$ exists for all $t \geq \max\{1 - \phi_{\boxplus}(\mu), 0\}$, and

$$\phi_{\boxplus}(\mathbb{B}_t(\mu)) = \phi_{\boxplus}(\mu) + t$$

so that the semigroup draws distributions closer to the class ID($\boxplus$)*.*

Lemma B.3.9. *Let $\mu \in \mathcal{P}(\mathbb{R})$ and $x \in \mathbb{R}$. Then*

$$\phi_{\boxplus}(\mu) = \phi_{\boxplus}(\mu \boxplus \delta_x) = \phi_{\boxplus}(\mu \uplus \delta_x)$$

so that convolution with point masses does not effect the regularity of distributions.

Lemma B.3.10. *Suppose $\mu \in \mathrm{ID}(\uplus)$ for (γ, σ, Π) is its Lévy triplet. Then $\mu^{\uplus t} \in \mathcal{P}(\mathbb{R}^+)$ for some $t \geq 0$ if and only if $\Gamma_{\uplus \to *}(\mu)$ is the distribution of a classical subordinator. As a result, $\mu^{\uplus t} \in \mathcal{P}(\mathbb{R}^+)$ for some $t > 0$ if and only if $\mu^{\uplus t} \in \mathcal{P}(\mathbb{R}^+)$ for all $t > 0$.*

B.4 Taxonomy of Distributions in Noncommutative Probability

We have already introduced the key distributions arising in Free Probability in Chapter 5, and use this section to discuss some more obscure connections between them. We note that for any probability distribution $\mu \in \mathcal{P}(\mathbb{R})$, the dilation $D_a(\mu)$ by a factor $a > 0$ has a convenient relationship with the Stieltjes and η transforms

$$S_{D_a(\mu)}(z) = \frac{1}{a} S_\mu(z/a) \qquad \eta_{D_a(\mu)}(z) = \eta(az)$$

Recall that we use μ^2 to denote the measure corresponding to the pushforward of the measure μ by the squaring process, and $\sqrt{\mu}$ for the symmetrized pullback. We consequently have that

$$S_{\mu^2}(z) = \frac{1}{\sqrt{z}} S_\mu\left(\sqrt{z}\right) \qquad \eta_{\mu^2}(z) = \eta_\mu\left(\sqrt{z}\right)$$

and for a nonnegative distribution $\mu \in \mathcal{P}(\mathbb{R}^+)$,

$$S_{\sqrt{\mu}}(z) = z S_\mu\left(z^2\right) \qquad \eta_{\sqrt{\mu}}(z) = \eta_\mu\left(z^2\right)$$

Lemma B.4.1 (Arizmendi and Hasebe, 2016a)**.** *Let $\mu \in \mathcal{P}(\mathbb{R})$, $x, y \in \mathbb{R}$, $s \geq 0$, and $t \geq 1$. Furthermore, let $p \geq 1$ and $q > 1 - 1/p$. Then the following relationships hold:*

- $(\mu \boxplus \delta_x)^{\uplus s} = \mu^{\uplus s} \uplus \delta_{(s-1)x} \boxplus \delta_x$

- ► $(\mu \uplus \delta_x)^{\boxplus t} = \mu^{\boxplus t} \uplus \delta_x \boxplus \delta_{(t-1)x}$
- ► $(\mu^2)^{\uplus s} = (\mu^{\uplus s})^2$
- ► *If* $p' = pq/(1-p+pq)$ *and* $q' = 1-p+pq$ *then*

$$\left(\mu^{\boxplus p}\right)^{\uplus q} = \left(\mu^{\uplus q'}\right)^{\boxplus p'}$$

- ► *If* $\mu, \nu \in \mathcal{P}(\mathbb{R}^+)$ *then*

$$D_t\left((\mu \boxtimes \nu)^{\boxplus t}\right) = \mu^{\boxplus t} \boxtimes \nu^{\boxplus t}$$

$$D_s\left((\mu \boxtimes \nu)^{\uplus s}\right) = \mu^{\uplus s} \boxtimes \nu^{\uplus s}$$

We recall some of the distributions already discussed: the Semicircle $\boldsymbol{s}$ (Example 5.1.9) distribution on $[-2, 2]$, the Free Poisson $\boldsymbol{\pi}$ distribution (Examples 5.1.11 and 5.3.5), and the Rademacher $\boldsymbol{r}$ and Arcsine $\boldsymbol{a}$ distributions (Examples 5.3.3, B.3.1 and B.3.2). The standard Gaussian distribution $\boldsymbol{g} \stackrel{d}{=} \mathrm{N}(0,1)$ is also mentioned in Example B.1.7. From the forms of the Stieltjes and η-transforms, we can see that the Semicircle $\boldsymbol{s}$, Rademacher $\boldsymbol{r}$ and Arcsine $\boldsymbol{a}$ play the role of the Gaussian (2-stable) distributions in the Free, Boolean, and Monotone additive convolutions. The role of the Poisson distribution in the Boolean case is played by the Boolean Poisson distributions $\rho^{\uplus s} = \frac{1}{1+s}\delta_0 + \frac{s}{1+s}\delta_{1+s}$.

The infinite divisibility bijections $\Gamma_{\circledast_1 \to \circledast_2}$ imply that α-stable distributions exist in the Free, Boolean, and Monotone additive convolutions. Curiously, the symmetric 1-stable distributions are the same as in classical probability, and coincide with the Cauchy distributions $\boldsymbol{c}_{a,b}$ whose densities are given by

$$\frac{d\boldsymbol{c}_{a,b}(x)}{dx} = \frac{1}{\pi} \frac{b}{(x-a)^2 + b^2}$$

As in the classical case, the α-stable densities are hard to describe. In the noncommutative cases, they are presented with the stability parameter $0 < \alpha < 2$ and skewness parameter

$$\rho = \mathbb{P}[X \geq 0]$$

with $0 \leq \rho \leq 1$, where $\rho = 1$ corresponds to nonnegative distributions when $0 < \alpha < 1$. For the cases $1 \leq \alpha < 2$, the further restrictions on the stability parameters are that $1 - 1/\alpha \leq \rho \leq 1/\alpha$. For the Free, Boolean, and Monotone cases, we write these distributions as $\boldsymbol{f}_{\alpha,\rho}$, $\boldsymbol{b}_{\alpha,\rho}$, and $\boldsymbol{m}_{\alpha,\rho}$, respectively.

In the Boolean case, the nonnegative distributions $\boldsymbol{b}_{\alpha,1}$ are absolutely continuous for $0 < \alpha \leq 1/2$, and have densities given by

$$\frac{d\boldsymbol{b}_{\alpha,1}(x)}{dx} = \frac{1}{\pi} \frac{\sin(\alpha\pi)x^{\alpha-1}}{x^{2\alpha} + 2\cos(\alpha\pi)x^{\alpha} + 1} \mathbb{1}_{(0,\infty)}(x)$$

Based on the properties described in Theorem B.5.1 below, we restrict ourselves to these parameters.

In the Free case, the only stable density known, apart from the Cauchy and Semicircle distributions, is the Inverse Beta distribution appearing as

$$\frac{d\boldsymbol{f}_{1/2,1}(x)}{dx} = \frac{\sqrt{4x-1}}{2\pi x^2} \mathbb{1}_{[1/4,\infty)}(x)$$

We note that if X is a random variable following a Beta distribution $\text{Beta}(\alpha, \beta)$ with parameters $\alpha = 1/2$ and $\beta = 3/2$, then $1/(4X) \stackrel{d}{=} \boldsymbol{f}_{1/2,1}$.

Similarly, in the Monotone case, the only additional stable density which is known is the 1/2-stable positive distribution

$$\frac{d\boldsymbol{m}_{1/2,1}(x)}{dx} = \frac{1}{\pi} \frac{\sqrt{x}}{x^2 - x + 1} \mathbb{1}_{(0,\infty)}(x)$$

Table B.1: Table of Noncommutative Probability Transforms for Popular Densities

μ	$d\mu^{\text{abs}}(x)/dx$	μ^{discrete}	$S_\mu(z)$	$\eta_\mu(z)$	$\phi_{\boxplus}(\mu)$
$\boldsymbol{r}$		$\frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$	$\frac{z}{1-z^2}$	z^2	0
$\boldsymbol{\rho}^{\uplus s}$		$\frac{1}{1+s}\delta_0 + \frac{s}{1+s}\delta_{1+s}$	$\frac{1-z}{z(z-1-s)}$	$\frac{sz}{1-z}$	0
$\boldsymbol{a}$	$\frac{1}{\pi\sqrt{4-x^2}}\mathbb{1}_{[-2,2]}(x)$		$-\frac{1}{\sqrt{z^2-4}}$	$1-\sqrt{1-4z^2}$	$\frac{1}{2}$
$\boldsymbol{s}$	$\frac{1}{2\pi}\sqrt{4-x^2}\mathbb{1}_{[-2,2]}(x)$		$\frac{\sqrt{z^2-4}-z}{2}$	$1-\frac{2z^2}{1-\sqrt{1-4z^2}}$	1
$\boldsymbol{\pi}^{\boxplus t}$	$\frac{\sqrt{(t_+-x)(x-t_-)}}{2\pi x}\mathbb{1}_{[t_-,t_+]}(x)$	$\max\{(1-t),0\}\delta_0$	$\frac{\sqrt{(z-t_+)(z-t_-)}-z-1+t}{2z}$		1
$\boldsymbol{c}_{a,b}$	$\frac{1}{\pi}\frac{b}{(x-a)^2+b^2}$		$\frac{1}{a-ib-z}$	$(a-ib)z$	∞
$\boldsymbol{g}$	$\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$		$i\sqrt{\frac{\pi}{2}}e^{-z^2/2}\left(1+\text{Erf}\left(i\frac{z}{\sqrt{2}}\right)\right)$		1
$\boldsymbol{b}_{\alpha,1}$	$\frac{1}{\pi}\frac{\sin(\alpha\pi)x^{\alpha-1}}{x^{2\alpha}+2\cos(\alpha\pi)x^\alpha+1}\mathbb{1}_{(0,\infty)}(x)$				∞
$\boldsymbol{f}_{1/2,1}$	$\frac{\sqrt{4x-1}}{2\pi x^2}\mathbb{1}_{[1/4,\infty)}(x)$		$\frac{2}{1-2z-i\sqrt{4z-1}}$	$\frac{z-i\sqrt{4z-z^2}}{2}$	≥ 1
$\boldsymbol{m}_{1/2,1}$	$\frac{1}{\pi}\frac{\sqrt{x}}{x^2-x+1}\mathbb{1}_{(0,\infty)}(x)$		$-\frac{1}{\left(\sqrt{z}+i\right)^2}$	$z-2i\sqrt{z}$	≥ 1

The Monotone α-stale distributions are known to be ID($\boxplus$) for $0 < \alpha \leq 1/2$, however the nonnegative versions are not FR (Arizmendi and Hasebe, 2013).

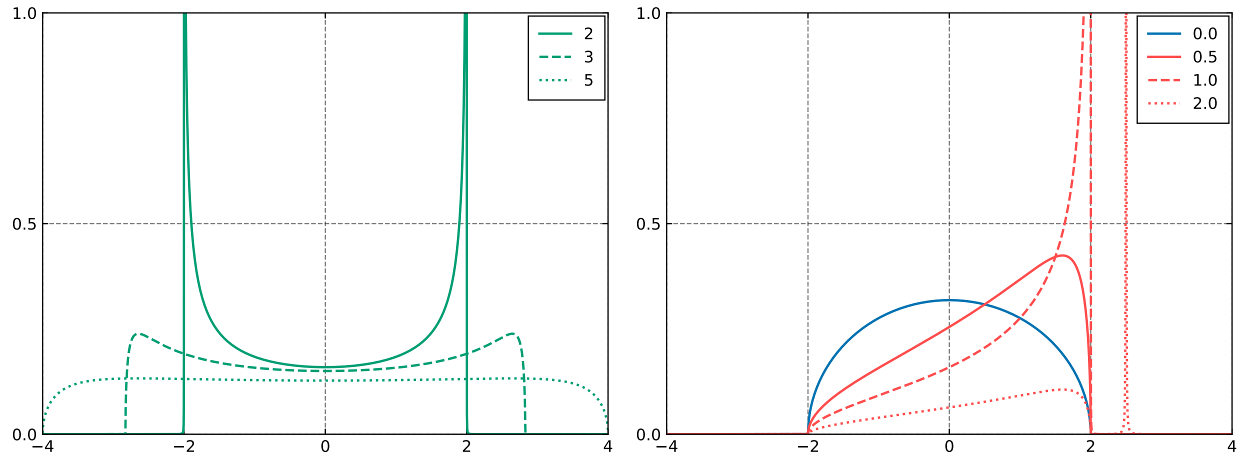
Example B.4.2. The distributions $\boldsymbol{s}$, $\boldsymbol{r}$, and $\boldsymbol{a}$ play the roles of the Gaussian (2-stable) distribution for the purpose of the noncommutative central limit theorems, in the sense that they satisfy the scaling relationships

$$\boldsymbol{s}^{\boxplus t} = D_{\sqrt{t}}(\boldsymbol{s}) \qquad \boldsymbol{r}^{\uplus s} = D_{\sqrt{s}}(\boldsymbol{r}) \qquad \boldsymbol{a}^{\triangleright n} = D_{\sqrt{n}}(\boldsymbol{a})$$

where $t, s > 0$ and $n \in \mathbb{N}$.

Example B.4.3. The fact that the square of the Semicircle (Free 2-stable) distribution is the Free Poisson is unique to the Free case, and has no analog in the classical world. In fact, each of the noncommutative 2-stable distributions has an interesting squaring rule:

$$\boldsymbol{s}^2 = \boldsymbol{\pi} \qquad \boldsymbol{r}^2 = \delta_1 \qquad \boldsymbol{a}^2 = \boldsymbol{a} \boxplus \delta_2$$



(a) (b)

Figure B.3: Approximate densities of (a) $\boldsymbol{r}^{\boxplus t} = D_{\sqrt{t-1}}\left(\boldsymbol{s}^{\uplus \frac{t}{t-1}}\right)$ and (b) $\delta_x \triangleright \boldsymbol{s} = \boldsymbol{s} \uplus \delta_x$, for various values of $t > 1$ and $x \geq 0$.

They are also related through other powers of 2

$$\boldsymbol{s}^{\uplus 2} = \boldsymbol{a} \qquad \boldsymbol{r}^{\boxplus 2} = \boldsymbol{a} \qquad \boldsymbol{\pi}^{\uplus 2} = \left(\boldsymbol{s}^2\right)^{\uplus 2} = \left(\boldsymbol{s}^{\uplus 2}\right)^2 = \boldsymbol{a} \boxplus \delta_2$$

which can easily be seen from the η-transforms of each of the distributions. The first two relationships further imply that

$$\boldsymbol{s} \uplus \boldsymbol{s} = \boldsymbol{r} \boxplus \boldsymbol{r} = \boldsymbol{r} \triangleright \boldsymbol{s} = \boldsymbol{a} \tag{B.1}$$

Example B.4.4. In Figure B.1b we can see that Boolean convolutions of the form $\delta_x \triangleright \mu = \mu \uplus \delta_x$ have the effect of skewing distributions in the direction and magnitude of $x \in \mathbb{R}$. When applied to the Semicircle distribution $\boldsymbol{s}$, an interesting shape appears, as seen in Appendix B.4. For large enough values of $x > 0$, the compact distribution emits a point mass. Furthermore, the skewed Semicircle resembles the M–P or Free Poisson laws $\boldsymbol{\pi}^{\boxplus t}$. By skewing and rescaling, it is not hard to show from the Stieltjes and η-transforms that

$$\delta_{-1} \rhd \boldsymbol{s}^{\boxplus t} \rhd \delta_{1+t} = \boldsymbol{\pi}^{\boxplus t}$$

where $t > 0$. Interestingly, a similar relation exists in the Boolean case:

$$\delta_{s-1} \rhd \boldsymbol{r}^{\uplus s} \rhd \delta_1 = \boldsymbol{\rho}^{\uplus s}$$

where $s > 0$. These do not appear to have been noticed or published previously.

Example B.4.5. The Monotone multiplicative convolution $\circlearrowright$ can be used in conjunction with various η-transforms to express some common measure transforms. For instance, if $\mu \in \mathcal{P}_s(\mathbb{R})$ is symmetric, then

$$\mu^2 \circlearrowright \boldsymbol{r} = \mu$$

Since $\eta_{\delta_a}(z) = az$, it is easy to see that $\mu \circlearrowright \delta_a = D_a(\mu)$ for $\mu \in \mathcal{P}(\mathbb{R}^+)$. The opposite direction mirrors the additive case: $\delta_a \circlearrowright \mu = \mu \mathbin{\cup\mkern-13mu\times} \delta_a$. However, an even stranger relation also occurs, which is obvious in retrospect:

$$\delta_a \circlearrowright \mu = \mu \mathbin{\cup\mkern-13mu\times} \delta_a = \mu^{\uplus a}$$

This holds when $\mu \in \mathcal{P}(\mathbb{R})$ is not necessarily nonnegative (the left and right sides of the expression always exists), since the expression $\mathbin{\cup\mkern-13mu\times}$ makes sense when considering the product

$$\eta_{\mu \mathbin{\cup\mkern-13mu\times} \delta_a}(z) = \frac{\eta_\mu(z)\eta_{\delta_a}(z)}{z} = \frac{az\eta_\mu(z)}{z} = a\eta_\mu(z) = \eta_{\mu^{\uplus a}}(z)$$

Similar manipulations of the transforms shows that

$$\boldsymbol{\rho}^{\uplus s} \circlearrowright \mu = \frac{1}{1+s}\delta_0 + \frac{s}{1+s}\mu^{\uplus 1+s}$$

Example B.4.6. We have the strange observation that

$$\boldsymbol{\pi} \mathbin{\cup\!\!\!\!\times} \boldsymbol{\pi} = \frac{1}{2}\delta_0 + \frac{1}{2}\left(\boldsymbol{a} \boxplus \delta_2\right)$$

which implies, given the previous example, that $\boldsymbol{\pi}^{\cup\!\!\!\!\times 2} = \frac{1}{2}\delta_0 + \frac{1}{2}\boldsymbol{\pi}^{\uplus 2} = \boldsymbol{\rho}^{\uplus 2} \circlearrowright \boldsymbol{\pi}$. According to Definition B.2.4, this means that there exists some distribution $\nu \in \mathcal{P}(\mathbb{R}^+)$ such that

$$\boldsymbol{\pi} \mathbin{\cup\!\!\!\!\times} \boldsymbol{\pi} = \boldsymbol{\rho}^{\uplus 2} \circlearrowright \boldsymbol{\pi} = \nu \circlearrowright \boldsymbol{\pi} = \boldsymbol{\rho}^{\uplus 2} \boxtimes \nu = \frac{1}{2}\delta_0 + \frac{1}{2}\left(\boldsymbol{a} \boxplus \delta_2\right)$$

although it is not clear from the right side what the distribution ν should be.

Example B.4.7. Suppose

$$\mu = p\delta_0 + (1-p)\nu$$

where $0 < p < 1$ and $\nu \in \mathcal{P}(\mathbb{R})$ such that $\nu(\{0\}) = 0$. Then we can write

$$\mu = \boldsymbol{\rho}^{\uplus \frac{1-p}{p}} \circlearrowright \nu^{\uplus p}$$

It is well known, for instance, that for $0 < t < 1$, we have

$$\boldsymbol{\pi}^{\boxplus t} = ((1-t)\delta_0 + t\delta_1) \boxtimes \boldsymbol{\pi} = D_{1-t}\left(\boldsymbol{\rho}^{\uplus \frac{t}{1-t}}\right) \boxtimes \boldsymbol{\pi} = \boldsymbol{\rho}^{\uplus \frac{t}{1-t}} \boxtimes D_{1-t}\left(\boldsymbol{\pi}\right)$$

and

$$\boldsymbol{\pi}^{\boxplus t} = (1-t)\delta_0 + tD_t\left(\boldsymbol{\pi}^{\boxplus 1/t}\right)$$

Therefore, it follows that

$$\boldsymbol{\pi}^{\boxplus t} = \boldsymbol{\rho}^{\uplus \frac{t}{1-t}} \circlearrowright D_t\left(\boldsymbol{\pi}^{\boxplus 1/t}\right)^{\uplus 1-t}$$

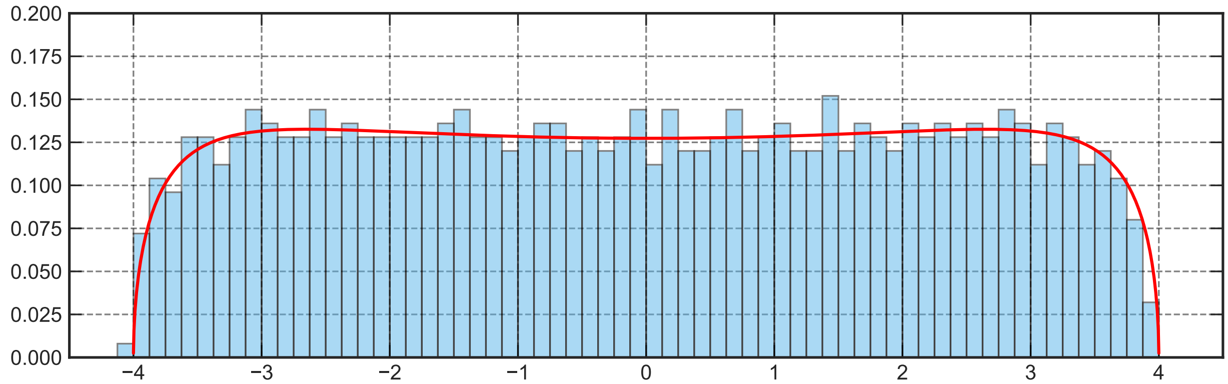


Figure B.4: Comparison of the eigenvalues of a matrix of the type described in Example B.4.8 with $N = 10^3$ and $k = 5$, with the estimation given by $\boldsymbol{r}^{\boxplus 5} = D_2\left(\boldsymbol{s}^{\uplus 5/4}\right)$ (red line).

According to Definition B.2.4, this implies that there exists some distribution $\xi \in \mathcal{P}(\mathbb{R}^+)$ such that

$$\begin{aligned}\boldsymbol{\rho}^{\uplus \frac{t}{1-t}} \boxtimes D_{1-t}\left(\boldsymbol{\pi}\right) = \boldsymbol{\pi}^{\boxplus t} &= \boldsymbol{\rho}^{\uplus \frac{t}{1-t}} \circlearrowright D_t\left(\boldsymbol{\pi}^{\boxplus 1/t}\right)^{\uplus 1-t} \\ &= D_{1-t}\left(\boldsymbol{\pi}\right) \circlearrowright \zeta = D_t\left(\boldsymbol{\pi}^{\boxplus 1/t}\right)^{\uplus 1-t} \mathbin{\times\!\!\!\!\!\cup} \xi\end{aligned}$$

Once again, it is not clear what distribution ξ should have.

Example B.4.8. From (B.1) it is not hard to see that

$$\mathbb{B}_t(\boldsymbol{r}) = D_{\sqrt{t}}\left(\boldsymbol{s}^{\uplus \frac{1}{t}}\right)$$

from which we get that

$$\boldsymbol{r}^{\boxplus t} = D_{\sqrt{t-1}}\left(\boldsymbol{s}^{\uplus \frac{t}{t-1}}\right)$$

where $t \geq 1$. The distribution $\boldsymbol{r}^{\boxplus k}$ for $k \in \mathbb{N}$ is notable because it expresses the eigenvalues of large random matrices of the form

$$\mathbf{U}_1\mathbf{A}_1\mathbf{U}_1^\dagger + \mathbf{U}_2\mathbf{A}_2\mathbf{U}_2^\dagger + \ldots + \mathbf{U}_k\mathbf{A}_k\mathbf{U}_k^\dagger$$

Here the $\mathbf{U}_j$ are taken to be independent Haar distributed random orthogonal matrices, and the $\mathbf{A}_j$ are diagonal matrices with half their diagonal entries equal to $+1$ and the other equal to -1. All matrices are taken to be size $N \times N$ for some particularly large N. This is the situation discussed in Example 5.3.3 and expressed in Figure 5.1. In the case of $k = 2$, $\boldsymbol{r}^{\boxplus 2} = \boldsymbol{a}$ and so the distribution is well understood. In higher order cases, however, there is no convenient analytic description for these distributions. In particular, trying to compute $\boldsymbol{r}^{\boxplus k}$ by inverting the reciprocal Cauchy transform may not be tractable. On the other hand, an expression involving $\boldsymbol{s}^{\uplus k/(k-1)}$ is tractable, since it only involves computing the new Stieltjes transform of the form

$$S_{\boldsymbol{s}^{\uplus k/(k-1)}}(z) = \frac{1}{\frac{k}{(k-1)S_{\boldsymbol{s}}(z)} + \frac{k}{k-1}z - z}$$

Since the Stieltjes transform for $\boldsymbol{s}$ is fast to compute, Stieltjes inversion can be used to approximate the density to these types of distributions. Furthermore, the properties of noncommutative convolutions tell us immediately that the support of $\boldsymbol{r}^{\boxplus k}$ will be precisely $[-2\sqrt{k-1}, 2\sqrt{k-1}]$. A comparison of such an estimation to the actual eigenvalues observed in such a matrix for $k = 5$ is demonstrated in Figure B.4.

B.5 Intersections of Classical and Free Infinite Divisibility

Significant interest has emerged over the last decade in investigating the intersection between classical and Free infinite divisibility, and whether a satisfying theory could even exist at all. Common laws in Free Probability, such as the Free Poisson (M–P), Semicircle, and Arcsine distributions cannot be classically infinitely divisible, as their densities are continuous but compactly supported. Very surprisingly, the standard Gaussian distribution $\boldsymbol{g} \stackrel{d}{=} \mathrm{N}(0,1)$ was

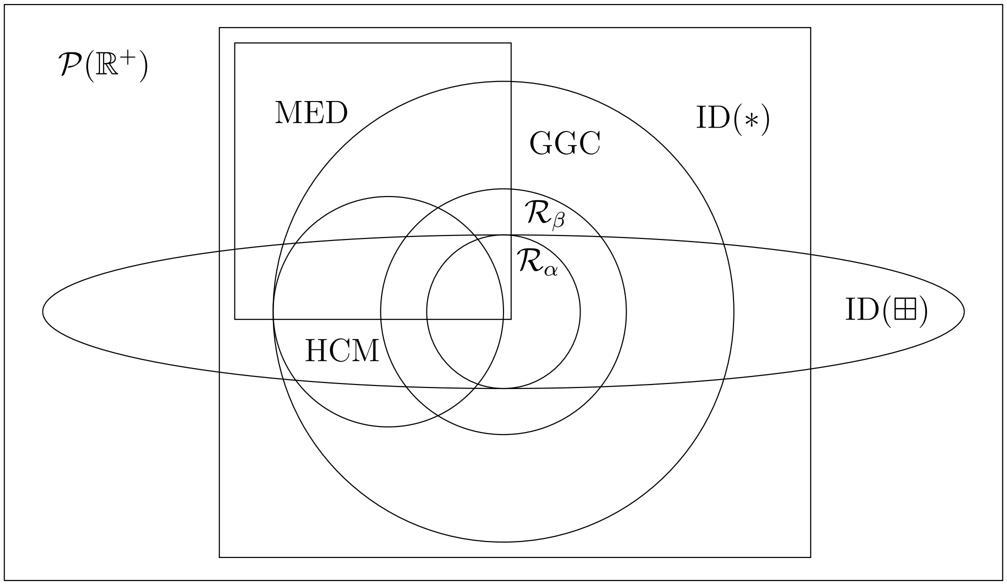


Figure B.5: Conjectured intersection between the classes of nonnegative ID($*$) and ID($\boxplus$) distributions, according to Conjecture B.5.3.

shown to be Freely infinitely divisible, and can even be shown to have a Free divisibility indicator $\phi_\boxplus(\boldsymbol{g}) = 1$. Specifically, we have that $\boldsymbol{g}^{\uplus s} \in$ ID($\boxplus$) if and only if $0 \leq s \leq 1$. On the other hand, it was also possible to show (Hasebe, 2014) that $\boldsymbol{g}^{\uplus s} \in$ ID($*$) if and only if $s \in \{0, 1\}$, so the interactions between the classical and noncommutative evolutions may be quite strange.

The first non-trivial family of distributions from the classical world which were shown to live in ID($*$) $\cap$ ID($\boxplus$) appeared in Hasebe (2014). Interestingly, these properties are not path dependent. For instance, the Gamma subordinator Γ_t is in ID($\boxplus$) for $t \in (0, 1/2] \cup [3/2, \infty)$, but fails to be in ID($\boxplus$) for a complicated union of intervals contained in $(1/2, 3/2)$, including the key case of the exponential distribution Exp(1) when $t = 1$. On the other hand, the inverse Gamma distributions, which follow $1/X$ when $X \stackrel{d}{=} \Gamma_t$, are in ID($\boxplus$) for all values of $t > 0$.

A huge family of distributions living in ID($*$) $\cap$ ID($\boxplus$) finally appeared in Arizmendi and Hasebe (2016b), in terms of mixture models of Boolean α-stable distributions.

Theorem B.5.1 (Arizmendi and Hasebe, 2016b)**.** *Let X be any nonnegative random variable independent from a random variable B, and suppose B follows $\boldsymbol{b}_{\alpha,\rho}$ for $0 < \alpha \leq 1/2$ and $0 \leq \rho \leq 1$, or $\rho = 1/2$ (the symmetric case) with $0 \leq \alpha \leq 2/3$. Then the (classical) independent product $XB \in \mathrm{ID}(*) \cap \mathrm{ID}(\boxplus)$.*

The intersection between $\mathrm{ID}(\boxplus)$ and the EGGC class has just recently been shown to contain a rich family of Free α-stable distributions. These distributions are clearly Free infinitely divisible.The following result shows that many are also classically infinitely divisible.

Theorem B.5.2 (Hasebe et al., 2018)**.** *For every $0 < \alpha \leq 1$ and $0 \leq \rho \leq 1$, the distributions $\boldsymbol{f}_{\alpha,\rho} \in \mathrm{ID}(*)$. Furthermore, if $0 < \alpha \leq 3/4$, then we have the stronger result that $\boldsymbol{f}_{\alpha,\rho} \in \mathrm{EGGC}$. On the other hand, for every $1 < \alpha < 2$, the symmetric distributions $\boldsymbol{f}_{\alpha,1/2} \notin \mathrm{ID}(*)$.*

In particular, the distributions $\boldsymbol{f}_{\alpha,1}$ for $0 < \alpha \leq 3/4$ are in the intersection $\mathrm{FR} \cap \mathrm{GGC}$. Given the strange but interesting interactions between the two probabilities, we propose the following conjecture.

Conjecture B.5.3. *Let $\mathcal{R}_\beta$ denote the class of distributions which can be written in the form X/Γ_β where $X \in \mathrm{GGC}$ and Γ_β is a Gamma subordinator independent from X, as introduced in Section 3.2.6. Recall that we have*

$$\mathcal{R}_\alpha \subseteq \mathcal{R}_\beta$$

for $0 < \alpha < \beta$, and

$$\mathcal{R}_\beta \to \mathrm{GGC}$$

as $\beta \to \infty$ (considering the weak closure). Then there is some small $0 < \alpha < 1$ such that $\mathcal{R}_\alpha \subset \mathrm{ID}(\boxplus)$.