Introduction to the Fractional Distribution Families

Stephen Horng-Twu Lihn
stevelihn@gmail.com
Version: Early Draft 2025-04

ORFE, Princeton University

Dedicated to Professor John M. Mulvey for his 80th birthday.

(as of April 30, 2025)

Contents

CHAPTER 1

Introduction

In quantitative finance, we often encounter data sets with prominent skewness and kurtosis. In the domain of portfolio optimization and the market regime model[**11, 24**], a showcase example is the S&P 500 Index (SPX) and the CBOE Volatility Index (VIX), whose daily prices are publicly available (since 1990). Such data sets are easy to obtain but hard to make sense even in the simplest style statistics. Surprise!

For example, the daily return distribution of VIX has a skewness of 2.0 and kurtosis of 16. It is not easy to find a two-sided distribution that can produce a good fit for it. The daily return distribution of SPX is even more peculiar (in Figure 12.2). In addition to its negative skewness and high kurtosis of 10, we point out that its standardized peak density is approximately 0.65. It is found that it takes a Student's t of about 3 degrees of freedom to produce a reasonable fit. But the theoretical kurtosis (the fourth moment) is not defined at 3 degrees of freedom. This is very puzzling.

When it comes to the multivariate study, how does one put the SPX and VIX return distributions into a parametric bivariate distribution? They have very different shapes, skewnesses, and tail behaviors. None of the existing multivariate distributions can handle it with ease.

This project was born out of an attempt to understand these strange financial data sets. Maybe their statistics could be captured by some kind of new distributions. In fact, the *fractional distribution* system may be the answer.

The word "fractional" can be roughly understood as adding the Lévy stability index $\alpha \in [0, 2]$ to a known distribution. For example, in the Mellin transform of the PDF of a distribution, $\Gamma(s + c)$ in the classic world becomes $\Gamma(\alpha s + c)$ or $\Gamma(s/\alpha + c)$ in the fractional world. When the coefficient of s is $\frac{1}{2}$, 1, or 2, the fractional distribution subsumes the classic distribution, since the Legendre duplication formula (A.2) becomes applicable.

The change may look simple in the Mellin space. But when it is transformed back to the x space, things become quite complicated. That is what makes it interesting and powerful.

The most important chapters of the book are

- Chapter 12 on the univariate GAS-SN distribution and
- Chapter 14 on the multivariate GAS-SN elliptical distribution.

The reader can think that the entire book is aimed at developing tools in order to create these two distributions.

The univariate GAS-SN distribution is supposed to be the most flexible two-sided distribution up to date for statisticians to fit a univariate data set, such as return distributions in finance.

The multivariate GAS-SN elliptical distribution is intended to be the most flexible multivariate distribution to date that extends the multivariate skew-t and skew-normal distributions[**1**].

This book is divided into three parts.

Part I describes the mathematical foundation needed for the construction of fractional distributions. It contains several higher transcendental functions. Several classic special functions are extended with a fractional parameter.

Each distribution has its density function (PDF) and distribution function (CDF). Its Mellin transform. The squared variable or quadratic forms. Therefore, new mathematical tools are needed to address them.

Part II contains the univariate one-sided fractional distributions that are invented. All of them have their classic counterparts. For example, the generalized gamma distribution (GG) is upgraded. All the χ and F related distributions are also upgraded.

Part III contains the two-sided univariate fractional distributions. The Azzalini (2013) book is used as the blueprint[**1**]. It is integrated with the symmetric distributions developed in my 2024 work[**13**].

This book can be viewed as an integration between the two works, literally going chapter-by-chapter. The consistency of such integration and harmony speaks volumes.

The fourth part contains the multivariate fractional distributions. These distributions are the super families of Part III. They subsumes and all the SN/ST distributions mentioned in Azzalini's book.

The major strength of fractional distributions integrated with SN is its ability to address a very wide range of skewness, kurtosis, and peak probability density. This allows a statistician to describe the statistics of her data set properly.

In the modern computer age, large amounts of data are collected in terms of both dimensionality and the number of samples. Tail behavior becomes more obvious. In the domain of finance, it is increasingly important to adequately capture the properties of the left tail.

An adaptive version of the multivariate distribution is developed to allow each dimension to have its own set of shape parameters. This distribution is where the rubber means the road. It is used to fit one of the most difficult data sets in finance: the daily returns from the SPX and VIX indices since 1990. And it works. The methodologies are presented.

Although the two multivariate distributions present new opportunities to fit the data sets that were thought impossible formerly, the outcomes post new challenges.

On the one hand, the maximum likelihood estimate (MLE) can be implemented in a straightforward manner for the elliptical distribution. The output (Figures 16.1, 16.2, 16.3) shows a very nice fit by MLE. But its choice of (α, k) lies in an area near infinite kurtosis when the bivariate distribution is projected to its two marginal 1D distributions. This behavior is quite puzzling.

On the other hand, the adaptive distribution suffers from the curse of dimensionality. A direct MLE approach is computationally prohibitive. A modified fitting algorithm is used. The output (Figures 16.4, 16.5, 16.6) is reasonable, but with a few flaws. The SPX marginal near $\alpha = 1, k = 3$ is intrinsically challenging. It is difficult to have a theoretical correlation coefficient that matches the empirical value (about -0.7). In the absolute term, the former is always lower than the latter. The quadratic form has not yet a matching F distribution.

Hope you enjoy this new statistical and mathematical adventure.

Part 1

Mathematical Functions

CHAPTER 2

Mellin Transform

We begin the book with some mathematical foundations. The reader who wishes to dive into the statistical distributions can skip the next two chapters.

The Mellin transform is crucial in the analysis of a statistical distribution. It is named after the Finnish mathematician Hjalmar Mellin, who first proposed it in 1897[**19**]. It provides insight into the inner workings of the statistical distribution, making it analytically tractable. Since once the Mellin transform is known, its moment formula is also known. In addition, derivatives of the probability density function (PDF) can be obtained as well.

In particular, the relations between the Wright function, the α-stable distribution, and the fractional χ distribution are best described by their Mellin transforms.

This chapter provides an overview of the Mellin transform. Following the notation of [**17**], the Mellin transform of a function $f(x)$ properly defined for $x \geq 0$ is

$$f^*(s) := \mathcal{M}\{f(x); s\} = \int_0^\infty f(x)\, x^{s-1}\, dx, \qquad c_1 < \mathbb{R}(s) < c_2. \tag{2.1}$$

The role of c_1, c_2 will be explained in the following.

If $f^*(s)$ has analytic continuation on the complex plane, the inverse Mellin transform is

$$f(x) = \mathcal{M}^{-1}\{f^*(s); x\} = \frac{1}{2\pi i} \int_{C-i\infty}^{C+i\infty} f^*(s)\, x^{-s}\, ds, \qquad c_1 < C < c_2. \tag{2.2}$$

The Mellin transform is directly related to the moments of a distribution. When $f(x)$ is the PDF of a one-sided distribution, its n-th moment is $\mathbb{E}(X^n|f) = f^*(n+1)$.

Hence, by modifying the Mellin transform $f^*(s)$, it is equivalent to constructing a new distribution based on the original distribution.

Introducing the juxtaposition notation $\overset{\mathcal{M}}{\longleftrightarrow}$, the above expressions, (2.1) and (2.2), are consolidated to a one-liner: $f(x) \overset{\mathcal{M}}{\longleftrightarrow} f^*(s)$, with a valid range $c_1 < C < c_2$ for C. This notation is much more concise. A correct specification for C is required when performing the Mellin integral in (2.2) numerically. Otherwise, it is irrelevant to the readers most of the time.

The main rules of Mellin transform used in this paper are:

$$f(ax) \overset{\mathcal{M}}{\longleftrightarrow} a^{-s} f^*(s), \qquad a > 0 \tag{2.3}$$

$$x^k f(x) \overset{\mathcal{M}}{\longleftrightarrow} f^*(s+k), \tag{2.4}$$

$$f(x^p) \overset{\mathcal{M}}{\longleftrightarrow} \frac{1}{p} f^*(s/p), \qquad p \neq 0 \tag{2.5}$$

$$h(x) = \int_0^\infty f(xs)g(s)\, s ds \overset{\mathcal{M}}{\longleftrightarrow} h^*(s) = f^*(s)g^*(2-s), \qquad \text{(ratio distribution)} \tag{2.6}$$

$$\gamma_f(x) = \int_0^x f(x)\, dx \overset{\mathcal{M}}{\longleftrightarrow} -s^{-1} f^*(s+1), \qquad \text{(lower incomplete function)} \tag{2.7}$$

$$\Gamma_f(x) = \int_x^\infty f(x)\, dx \overset{\mathcal{M}}{\longleftrightarrow} s^{-1} f^*(s+1). \qquad \text{(upper incomplete function)} \tag{2.8}$$

The ratio distribution rule (2.6) is widely used in our fractional distribution system. Notice that the argument of $g^*(s)$ is transformed via $s \to 2 - s$.

For (2.7) and (2.8), the valid range of C is decremented by one: $c_1 - 1 < C < c_2 - 1$.

Example 2.1. A simple exercise is the Mellin transform of the standard normal distribution. It starts with

$$e^{-x} \overset{\mathcal{M}}{\longleftrightarrow} \Gamma(s)$$

via the definition of the gamma function itself.

By applying (2.5) then (2.3), we get

$$\mathcal{N}(x) \overset{\mathcal{M}}{\longleftrightarrow} \mathcal{N}^*(s) = \frac{2^{(s-3)/2}}{\sqrt{\pi}}\, \Gamma\left(\frac{s}{2}\right) \tag{2.9}$$

where $\mathcal{N}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is our notation for the PDF of a standard normal distribution.

Example 2.2. A slightly more complicated exercise is the Mellin transform of the GSC distribution in Chapter 6. But we only work out its skeleton here.

Assume we have a function $F_\alpha(x)$ whose Mellin transform is

$$F_\alpha(x) \overset{\mathcal{M}}{\longleftrightarrow} \frac{\Gamma(s)}{\Gamma(\alpha s)}.$$

It undergoes the following transforms:

$$F_\alpha(x^p) \overset{\mathcal{M}}{\longleftrightarrow} \frac{1}{|p|} \frac{\Gamma(s/p)}{\Gamma(\alpha s/p)},$$

$$x^{d-1} F_\alpha(x^p) \overset{\mathcal{M}}{\longleftrightarrow} \frac{1}{|p|} \frac{\Gamma((s+d-1)/p)}{\Gamma(\alpha(s+d-1)/p)},$$

which is the prototype of GSC before further normalization.

2.1. Distribution Function and Moments

If $f(x)$ is a density function of a distribution, the two rules of incomplete functions provide a clear path to obtain its distribution function (CDF). On one hand, if the distribution is one-sided, then $\gamma_f(x)$ is its CDF obviously.

2.1.1. Mellin Transform of a Two-sided CDF. On the other hand, assume the distribution is two-sided and the density function satisfies the *reflection rule* based on a skew parameter:

$$f(-x;\beta) := f(x;-\beta) \quad \text{for } x > 0.$$

Further assume that

$$\int_0^\infty f(x;\beta)\,dx = c_\beta < 1.$$

which leads to $c_{-\beta} + c_\beta = 1$. Then we have

LEMMA 2.3. The Mellin transform of the CDF $\Phi(x)$ of a two-sided distribution has two parts. Both can be derived from its density function transform, $f(x;\beta) \overset{\mathcal{M}}{\longleftrightarrow} f^*(s;\beta)$, in the positive domain.

From (2.7), let $\gamma_f(x;\beta) \overset{\mathcal{M}}{\longleftrightarrow} \Phi^*(s;\beta) := -s^{-1} f^*(s+1;\beta)$. Then for $x > 0$, the Mellin transform of the CDF can be expressed as

$$\Phi(x) - \Phi(0) \overset{\mathcal{M}}{\longleftrightarrow} \Phi^*(s;\beta),$$
$$1 - \Phi(0) - \Phi(-x) \overset{\mathcal{M}}{\longleftrightarrow} \Phi^*(s;-\beta).$$

PROOF. Note that $\Phi(0) = c_{-\beta} = 1 - c_\beta$. When $x \geq 0$, its CDF is

$$\Phi(x) = \int_{-\infty}^x f(x;\beta)\,dx \;= c_{-\beta} + \int_0^x f(x;\beta)\,dx \;= \Phi(0) + \gamma_f(x;\beta).$$

In the negative domain, its CDF is

$$\begin{aligned}\Phi(-x) &= \int_{-\infty}^{-x} f(x;\beta)\,dx \;= \int_x^\infty f(x;-\beta)\,dx \\ &= 1 - \Phi(0) - \int_0^x f(x;-\beta)\,dx \;= 1 - \Phi(0) - \gamma_f(x;-\beta).\end{aligned}$$

□

The point is that, once the Mellin transform of either the PDF or CDF is known, the other one can be derived by simple algebraic rules.

2.1.2. From Mellin Transform to Moments. By assigning $s = n+1$, it is easy to show that its n-th moment is

$$\mathbb{E}(X^n|f) = f^*(n+1;\beta) + (-1)^n f^*(n+1;-\beta) \tag{2.10}$$
$$= -n\big[\Phi^*(n;\beta) + (-1)^n \Phi^*(n;-\beta)\big] \tag{2.11}$$

The moment formula is tightly linked to $\Phi^*(n;\beta)$.

The total density can be regarded as the zeroth moment. Hence,

$$c_\beta = \int_0^\infty f(x;\beta)dx = f^*(1;\beta). \tag{2.12}$$

Its application is in (10.9).

2.2. Ramanujan's Master Theorem

In order to keep things simple, we anchor all the distributions via the Mellin transform of their PDFs. Due to Ramanujan's master theorem[3], not only can the moments be obtained from the Mellin transform but also all the derivatives of the PDF at $x = 0$. We get its series representation "for free", so to speak.

LEMMA 2.4 (Ramanujan's master theorem). If $f(x)$ has an expansion of the form

$$f(x) = \sum_{n=0}^{\infty} \frac{\varphi(n)}{n!} (-x)^n \tag{2.13}$$

then its Mellin transform is given by

$$f(x) \overset{\mathcal{M}}{\longleftrightarrow} f^*(s) = \Gamma(s)\, \varphi(-s) \tag{2.14}$$

Assume that $g^*(s) := f^*(s)/\Gamma(s)$ exists on the complex plane, $s \in \mathbb{C}$. Its connection to the derivatives of the PDF at $x = 0$ is as follow.

LEMMA 2.5. The Taylor series of $f(x)$ is

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n$$

where $f^{(n)}(0)$ is the n-th derivative of $f(x)$ at $x = 0$.

Then $f^{(n)}(0)$ can be obtained from $g^*(s)$ by

$$f^{(n)}(0) = (-1)^n g^*(-n) \tag{2.15}$$

At $x = 0$, we have $f(0) = g^*(0)$.

The power of the master theorem is that, once the Mellin transform is known, the Taylor series is also known immediately. We provide a contrived example from next chapter as a showcase.

EXAMPLE 2.6. The Mellin transform of the Wright function from (3.5) is $f(-x) \overset{\mathcal{M}}{\longleftrightarrow} f^*(s) = \Gamma(s)/\Gamma(\delta - \lambda s)$. Then its $g^*(s) = 1/\Gamma(\delta - \lambda s)$.

According to Lemma 2.5, its Taylor series should be

$$f(-x) := \sum_{n=0}^{\infty} \frac{(-1)^n g^*(-n)}{n!} x^n = \sum_{n=0}^{\infty} \frac{g^*(-n)}{n!} (-x)^n$$

Replace $-x$ with z, and plug in $g^*(-n)$, we have

$$f(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!\, \Gamma(\lambda n + \delta)}$$

This is the series representation (3.1) where we essentially "derived" it from the master theorem.

The major application in this book is in Chapter 11. In the experimental construction of the generalized α-stable distribution, the theorem is used to remedy the discontinuity of the PDF in $x = 0$.

2.2.1. Distribution Function. The form of the Mellin transform in (2.14) has an important implication when $f(x)$ is a density function.

LEMMA 2.7. Assume $x > 0$, its complimentary distribution function $\Gamma_f(x) := \int_x^\infty f(x)\, dx$ has the series representation of

$$\Gamma_f(x) = \sum_{n=0}^{\infty} \frac{\varphi(n-1)}{n!} (-x)^n \tag{2.16}$$

PROOF. From (2.8), the Mellin transform of $\Gamma_f(x)$ is

$$\Gamma_f(x) = \int_x^\infty f(x)\,dx \overset{\mathcal{M}}{\longleftrightarrow} s^{-1} f^*(s+1)$$

which can be simplified to

$$\begin{aligned} s^{-1} f^*(s+1) &= s^{-1}\Gamma(s+1)\,\varphi(-s-1) \\ &= \Gamma(s)\,\varphi(-s-1). \end{aligned}$$

This is still in the form of (2.14), with a transformation rule of $s \to s+1$ in the function $\varphi(-s)$.

Applying the master theorem of (2.13), we get (2.16).

□

We use the CDF of the M-Wright function from (3.15) as an example.

LEMMA 2.8. The goal is to show

$$\int_x^\infty M_\alpha(t)dt = W_{-\alpha,1}(-x). \tag{2.17}$$

PROOF. We start with the Mellin transform of $M_\alpha(x)$ from (3.12),

$$M_\alpha(x) \overset{\mathcal{M}}{\longleftrightarrow} M_\alpha^*(s) = \frac{\Gamma(s)}{\Gamma((1-\alpha)+\alpha s)}$$

which yields $\varphi(-s) = 1/\Gamma((1-\alpha)+\alpha s)$.

Therefore, its $\Gamma_f(x)$ should be

$$\Gamma_f(x) = \sum_{n=0}^\infty \frac{1}{n!\,\Gamma((1-\alpha)-\alpha(n-1))}(-x)^n = \sum_{n=0}^\infty \frac{1}{n!\,\Gamma(-\alpha n+1)}(-x)^n$$

which is $W_{-\alpha,1}(-x)$ according to (3.1).

□

CHAPTER 3

The Wright Function

3.1. Definition

The Wright function is the most basic building block in our fractional distribution system. It was proposed by E. M. Wright in the 1930s[**28, 29**]. Bateman recorded this function together with the Mittag-Leffler function in the 1930s[**2**].

Its importance was gradually noticed since the late 1980's, especially through the works of F. Mainardi, who proposed the M-Wright function $M_\alpha(x)$. $M_\alpha(x)$ is considered the fractional extension of the exponential function e^{-x}. Such logic appears in many places of this book. This chapter provides an overview.

Definition 3.1. The series representation of the Wright function is

$$W_{\lambda,\delta}(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!\,\Gamma(\lambda n+\delta)} \qquad (\lambda \geq -1, z \in \mathbb{C}) \tag{3.1}$$

Its shape parameters are pairs (λ, δ). The apparent limit is $W_{0,1}(z) = e^z$.

The author used four variants extensively. The first group of two are

- $M_\alpha(z) := W_{-\alpha,1-\alpha}(-z)$
- $F_\alpha(z) := W_{-\alpha,0}(-z)$

where $\alpha \in [0, 1]$. They are related to each other by $M_\alpha(z) = F_\alpha(z)/(\alpha z)$.

In particular, $M_\alpha(z)$ is called *the M-Wright function* or simply *the Mainardi function*[**14, 18, 15**]. See Section 3.3 for further details. Conceptually, *fractional extension* of a classic exponential-based function is based on two important properties: $M_0(z) = \exp(-z)$ and $M_{\frac{1}{2}}(z) = \frac{1}{\sqrt{\pi}} \exp(-z^2/4)$.

The second group of the two are

- $W_{-\alpha,-1}(-z)$
- $-W_{-\alpha,1-2\alpha}(-z)$

The author discovers their usefulness. They are associated with the derivatives of $F_\alpha(z)$ and $M_\alpha(z)$, for the generation of random variables, such as in (3.16) and Section 11 of [**13**]. In some cases, they lead to beautiful polynomial solutions.

3.2. Classic Results

The recurrence relations of the Wright function are (Chapter 18, Vol 3 of [**2**])

$$\lambda z\, W_{\lambda,\lambda+\mu}(z) = W_{\lambda,\mu-1}(z) + (1-\mu) W_{\lambda,\mu}(z) \tag{3.2}$$

$$\frac{d}{dz} W_{\lambda,\mu}(z) = W_{\lambda,\lambda+\mu}(z) \tag{3.3}$$

The moments of the Wright function are (See (1.4.28) of [**18**])

$$\mathbb{E}(X^{d-1}) = \int_0^\infty x^{d-1} W_{-\lambda,\delta}(-x)\,dx = \frac{\Gamma(d)}{\Gamma(d\lambda+\delta)} \tag{3.4}$$

The way it is written is in fact its Mellin transform:

$$W_{\lambda,\delta}(-x) \overset{\mathcal{M}}{\longleftrightarrow} W^*_{\lambda,\delta}(s) = \frac{\Gamma(s)}{\Gamma(\delta - \lambda s)} \tag{3.5}$$

$W_{\lambda,\delta}(z)$ has the following Hankel integral representation:

$$W_{\lambda,\delta}(z) = \frac{1}{2\pi i} \int_H dt\, \frac{\exp\left(t + z t^{-\lambda}\right)}{t^\delta} \tag{3.6}$$

The four-parameter Wright function is defined as

$$W \begin{bmatrix} a, & b \\ \lambda, & \mu \end{bmatrix} (z) := \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(n+1)} \frac{\Gamma(an+b)}{\Gamma(\lambda n + \mu)} \tag{3.7}$$

This function is a higher-order Wright function. It was used seriously for the first time by the author[13].

3.3. The M-Wright Functions

Mainardi has introduced two auxiliary functions of Wright type (see F.2 of [14]):

$$F_\alpha(z) := W_{-\alpha,0}(-z) \qquad (z > 0) \tag{3.8}$$

$$M_\alpha(z) := W_{-\alpha,1-\alpha}(-z) = \frac{1}{\alpha z} F_\alpha(z) \qquad (z > 0) \tag{3.9}$$

The relation between $M_\alpha(z)$ and $F_\alpha(z)$ in (3.9) is an application of (3.2) by setting $\lambda = -\alpha, \mu = 1$.

$F_\alpha(z)$ has the following Hankel integral representation:

$$F_\alpha(z) = \frac{1}{2\pi i} \int_H dt\, \exp\left(t - z t^\alpha\right) \tag{3.10}$$

Both functions have simple Mellin transforms from (3.5):

$$F_\alpha(x) \overset{\mathcal{M}}{\longleftrightarrow} F^*_\alpha(s) = \frac{\Gamma(s)}{\Gamma(\alpha s)} \tag{3.11}$$

$$M_\alpha(x) \overset{\mathcal{M}}{\longleftrightarrow} M^*_\alpha(s) = \frac{\Gamma(s)}{\Gamma((1-\alpha) + \alpha s)} \tag{3.12}$$

$F_\alpha(z)$ is used to define fractional one-sided distributions. But its series representation isn't very useful computationally. It requires many more terms to converge to a prescribed precision.

On the other hand, $M_\alpha(z)$ has a more computationally friendly series representation, especially for small α's:

$$M_\alpha(z) = \sum_{n=0}^{\infty} \frac{(-z)^n}{n!\, \Gamma(-\alpha n + (1-\alpha))} = \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{(-z)^{n-1}}{(n-1)!} \Gamma(\alpha n) \sin(\alpha n \pi) \qquad (0 < \alpha < 1) \tag{3.13}$$

$M_\alpha(z)$ also has very nice analytic properties at $\alpha = 0, 1/2$, where $M_0(z) = \exp(-z)$ and $M_{\frac{1}{2}}(z) = \frac{1}{\sqrt{\pi}} \exp(-z^2/4)$.

$M_\alpha(z)$ can be computed to high accuracy when properly implemented with arbitrary-precision floating point library, such as the `mpmath` package[20]. In this regard, it is much more "useful" than $F_\alpha(z)$.

This is particularly important in working with large degrees of freedom and extreme values of α, mainly close to 0. Typical 64-bit floating point is quickly overflowed.

Both functions can be derived from the one-sided α-stable distribution $L_\alpha(x)$ of Section 4.2, which is implemented in `scipy.stats.levy_stable` package[27]. For example, $M_\alpha(x)$ can be computed using $L_\alpha(x) = \alpha x^{-\alpha-1} M_\alpha(x^{-\alpha})$ where $x > 0$.

$M_\alpha(z)$ has the asymptotic representation in a Generalized Gamma (GG) style: (See F.20 of [14])

$$M_\alpha\left(\frac{x}{\alpha}\right) = A\, x^{d-1}\, e^{-B\, x^p} \tag{3.14}$$

$$\text{where } p = 1/(1-\alpha), d = p/2,\ A = \sqrt{p/(2\pi)},\ B = 1/(\alpha p).$$

This formula is important in guiding (3.13) to high precision for large x.

$M_\alpha(x)$ can be used as the density function of a one-sided distribution[15]. In such case, $\int_0^\infty M_\alpha(x) dx = 1$, and its CDF is another Wright function:

$$\int_0^x M_\alpha(t) dt = 1 - W_{-\alpha,1}(-x). \tag{3.15}$$

This is proved in Lemma 2.8.

Differentiating $M_\alpha(z)$, and from (3.13), we get

$$\frac{d}{dz} M_\alpha(z) = -W_{-\alpha,1-2\alpha}(-z) = \frac{-1}{\pi} \sum_{n=2}^{\infty} \frac{(-z)^{n-2}}{(n-2)!} \Gamma(\alpha n) \sin(\alpha n \pi) \tag{3.16}$$

Note that $\frac{d}{dz} M_\alpha(0) = -\frac{1}{\pi} \Gamma(2\alpha) \sin(2\alpha\pi)$. This also indicates that

$$\frac{d}{dz} F_\alpha(z) = \alpha \left(1 + z \frac{d}{dz}\right) M_\alpha(z) \tag{3.17}$$

which can be implemented from $M_\alpha(z)$ through (3.13) and (3.16).

3.4. The Fractional Gamma-Star Function

The so-called γ^* function is documented in 8.2.6 and 8.2.7 of DLMF[5]. It is defined as follows:

$$\gamma^*(s,x) := \frac{1}{\Gamma(s)} \int_0^1 t^{s-1}\, e^{-xt}\, dt = \frac{x^{-s}}{\Gamma(s)} \gamma(s,x)$$

The finite integral in $t \in [0,1]$ is transformed from the incomplete gamma function, which takes the form of $\gamma(s,x) = \int_0^x t^{s-1}\, e^{-t}\, dt$.

$\gamma^*(s,x)$ can be extended fractionally in a straightforward manner. It is used to calculate the CDF of the GSC in Chapter 6. See (6.7) for details.

Definition 3.2 (The fractional γ^* function). It is defined by replacing e^{-xt} with $M_\alpha(xt)$ such that

$$\gamma^*_\alpha(s,x) := \frac{\Gamma((1-\alpha)+\alpha s)}{\Gamma(s)} \int_0^1 dt\, t^{s-1}\, M_\alpha(xt) \tag{3.18}$$

The $\alpha \to 0$ limit of $\gamma^*_\alpha(s,x)$ subsumes the classic γ^* function, that is, $\gamma^*_0(s,x) = \gamma^*(s,x)$. This is reflected in the simple fact that $M_0(xt) = \exp(-xt)$.

The γ^* function is a subset of the fractional confluent hypergeometric function in Lemma 5.4.

CHAPTER 4

The Alpha-Stable Distribution - Review

The two-sided distributions in this book are based on the α-stable distribution, which was published in the seminal 1925 book of Paul Lévy[**10**]. These distributions have a major parameter, among others, called *the stability index* $\alpha \in (0, 2]$. We call it the *fractional* parameter.

In this chapter, we provide a review of the α-stable distribution based on the Mellin transform framework. This framework lays the foundation for further generalization in subsequent chapters.

The ratio distribution approach for its density function in Section 4.3 is invented by the author.

4.1. Classic Result

The α-stable distribution has two shape parameters. There are many parametrizations that have been studied (see p.5 of [**21**]). We are primarily concerned with Feller's (α, θ) parametrization[**6, 7**], where α is called the stability index with a range of $0 < \alpha \leq 2$, and θ is an angle that injects skewness to the distribution when it is not zero.

An innovative approach is to study its Mellin transform. This presentation is used because it is *simpler* and provides great insight into its structure.

LEMMA 4.1. The Mellin transform of its PDF is

$$L_\alpha^\theta(x) \overset{\mathcal{M}}{\longleftrightarrow} \epsilon \, \frac{\Gamma(s)\Gamma(\epsilon(1-s))}{\Gamma(\gamma(1-s))\Gamma(1-\gamma+\gamma s)} \quad \text{where } \epsilon = \frac{1}{\alpha}, \gamma = \frac{\alpha-\theta}{2\alpha}. \tag{4.1}$$

where $0 < C < 1$ implicitly. This is defined for $x \geq 0$. The reflection rule is used for $x < 0$ such that $L_\alpha^\theta(x) := L_\alpha^{-\theta}(-x)$.

This result was first derived in 1986 by Schneider[**23**], then rediscovered in 2001 by Mainardi et al.[**16**], and summarized by Mainardi and Pagnini in (2.8) of [**17**], from which we quote.

In (4.1), instead of using (α, θ) directly, it uses a different representation, which we call the (ϵ, γ) representation. In the Mellin transform space, such representation is often more elegant.

The constraint on θ in the Feller parameterization: $|\theta| \leq \min\{\alpha, 2-\alpha\}$, is called the "Feller-Takayasu diamond". In the (ϵ, γ) parametrization, the constraint becomes (a) $0 \leq \gamma \leq 1$ when $\epsilon > 1$; and (b) $1 - \epsilon \leq \gamma \leq \epsilon$ when $\epsilon \leq 1$.[1]

4.1.1. The Reflection Rule. Note that the reflection of $\theta \to -\theta$ in the (α, θ) parametrization is equivalent to the reflection of $\gamma \to 1 - \gamma$ in the (ϵ, γ) parametrization.

Since we often mingle the two parameterizations, this alternative view can be very helpful in certain scenarios. For example, the total density in the positive domain is $\int_0^\infty L_\alpha^\theta(x) = \gamma$. By the reflection rule, $\int_0^\infty L_\alpha^{-\theta}(x) = 1 - \gamma$. Hence, the total density $\int_{-\infty}^\infty L_\alpha^\theta(x) = \gamma + (1 - \gamma) = 1$.

---

[1]Conversely, if γ is fixed, (b) puts a constraint on the largest α allowed: $\alpha \leq \min\{1/\gamma, 1/(1-\gamma)\}$.

4.2. Extremal Distributions

There are two types of the so-called "extremal distributions", where θ is pushed to the limit, so to speak. They are especially intriguing because the M-Wright functions, $F_\alpha(x), M_\alpha(x)$ in Section 3.3, can be derived from them.

They can be understood from (4.1). The first kind of extremal distribution lies in $\gamma = 0$ or $\gamma = 1$ when $\theta = \pm\alpha \leq 1$. Due to the reflection rule, we only need to study the case of $\theta = -\alpha$, that is, $\gamma = 1$. This defines the one-sided α-stable distribution:

$$L_\alpha(x) := L_\alpha^{-\alpha}(x) \overset{\mathcal{M}}{\longleftrightarrow} \epsilon \, \frac{\Gamma(\epsilon(1-s))}{\Gamma(1-s)}$$

Apply three manipulations of Mellin transform on $F_\alpha(x)$: First, $x \to x^\alpha$; second, multiply x; third, $x \to x^{-1}$. We obtain the classic result of

$$L_\alpha(x) = x^{-1} F_\alpha(x^{-\alpha}) \quad (x \geq 0 \text{ and } 0 < \alpha \leq 1) \tag{4.2}$$

and $L_1(x) = \delta(x-1)$ is the upper bound of this relation.

$L_\alpha(x)$ can be computed via `scipy.stats.levy_stable`[27] using 1-Parameterization with `beta`=1, `scale`= $\cos(\alpha\pi/2)^{1/\alpha}$ for $0 < \alpha < 1$.[2] It might seem somewhat peculiar that we can use the existing implementation of $L_\alpha(x)$ to develop all the new fractional distributions for proof of concept.

The second kind of extremal distribution (but not necessarily one-sided) occurs when $\theta = \alpha - 2$, which leads to $\epsilon = \gamma = 1/\alpha$ and

$$L_\alpha^{\alpha-2}(x) \overset{\mathcal{M}}{\longleftrightarrow} \epsilon \, \frac{\Gamma(s)}{\Gamma(1-\epsilon+\epsilon s))}$$

Compare it to (3.12), we get the classic result of (e.g. see (F.49) of [14])

$$L_\alpha^{\alpha-2}(x) = \frac{1}{\alpha} M_{1/\alpha}(x) \quad (x \in \mathbb{R} \text{ and } 1 < \alpha \leq 2) \tag{4.3}$$

Notice that it extends the M-Wright function to $x < 0$ because $L_\alpha^{\alpha-2}(x)$ is two-sided.

4.3. Ratio Distribution Approach

Important insight can be obtained by interpreting (4.1) as a ratio distribution (2.6). We split (4.1) into two components:

$$L_\alpha^\theta(x) \overset{\mathcal{M}}{\longleftrightarrow} \epsilon \left[\frac{\Gamma(s)}{\Gamma(1-\gamma+\gamma s))}\right] \left[\frac{\Gamma(\epsilon(1-s))}{\Gamma(\gamma(1-s))}\right] \tag{4.4}$$

The first bracket is the Mellin transform of the M-Wright function (3.12).

The second bracket comes from the Mellin transform of the PDF of the fractional χ-mean distribution (FCM) at $k = 1$:

$$\begin{aligned} \overline{\chi}_{\alpha,1}^\theta(x) &\overset{\mathcal{M}}{\longleftrightarrow} \overline{\chi}_{\alpha,1}^{\theta}{}^*(s) \\ &= \epsilon \gamma^{\gamma(s-1)-1} \, \frac{\Gamma(\epsilon(s-1))}{\Gamma(\gamma(s-1))} \end{aligned} \tag{4.5}$$

According to the Mellin transform rule of a ratio distribution, s should be replaced by $2-s$ in $\overline{\chi}_{\alpha,1}^{\theta}{}^*(s)$. Therefore, $s-1$ in the second line of (4.5) becomes $1-s$ in the second bracket of (4.4).

[2]See Chapter 1 of [21] for more detail on different parameterizations. We would not go into the issue of stable parameterizations.

4.3.1. Rescaled M-Wright Function. Additionally, a small nuance here is to deal with scaling factors. Define the rescaled M-Wright function

$$\tilde{M}_\gamma(x) := \gamma^{1-\gamma} M_\gamma(x/\gamma^\gamma) \tag{4.6}$$

such that it matches the standard normal distribution: $\tilde{M}_{1/2}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ of $\mathcal{N}(0,1)$. And $\int_0^\infty \tilde{M}_\gamma(x)\,dx = \gamma$ since $\int_0^\infty M_\gamma(x) = 1$.

Notice that, according to the reflection rule, $\int_0^\infty \tilde{M}_\gamma(-x)\,dx = \int_0^\infty \tilde{M}_{1-\gamma}(x)\,dx = 1-\gamma$. We get $\int_{-\infty}^\infty \tilde{M}_\gamma(x)\,dx = 1$. Hence, $\tilde{M}_\gamma(x)$ is a valid two-sided density function.

According to (2.3), the rescaling of PDF modifies the Mellin transform from (3.12) to

$$\begin{aligned} \tilde{M}_\gamma(x) &\overset{\mathcal{M}}{\longleftrightarrow} \tilde{M}^*_\gamma(s) \\ &= \gamma^{1-\gamma+\gamma s} \, \frac{\Gamma(s)}{\Gamma(1-\gamma+\gamma s)} \end{aligned} \tag{4.7}$$

from which the $\gamma^{1-\gamma+\gamma s}$ term cancels out its counterpart in $\overline{\chi}^{\theta}_{\alpha,1}{}^*(2-s)$ nicely.

Therefore, we find a new method to construct the α-stable distribution using the following integral.

Lemma 4.2 (The ratio-distribution representation of the α-stable distribution). The Mellin transform of the PDF (4.1) becomes

$$L^\theta_\alpha(x) \overset{\mathcal{M}}{\longleftrightarrow} \tilde{M}^*_\gamma(s)\, \overline{\chi}^{\theta}_{\alpha,1}{}^*(2-s) \tag{4.8}$$

from which the PDF can be written in a ratio distribution form of

$$L^\theta_\alpha(x) := \int_0^\infty \tilde{M}_\gamma(xs)\, \overline{\chi}^{\theta}_{\alpha,1}(s)\, s\, ds \quad (x \ge 0) \tag{4.9}$$

Since the Mellin integral is only valid for $x > 0$, it is supplemented with *the reflection rule*:

$$L^\theta_\alpha(-x) := L^{-\theta}_\alpha(x) \tag{4.10}$$

This construction places $\overline{\chi}^{\theta}_{\alpha,1}$ in the central role. We define it at one degree of freedom $k = 1$. In Chapter 7, we will add *degrees of freedom* k to it and make it $\overline{\chi}^{\theta}_{\alpha,k}$, which is the fractional extension of the classic χ distribution.

Subsequently, in Chapter 11, we will add *degrees of freedom* k to the α-stable distribution and merge it with Student's t distribution.

4.4. SaS

Note that $\theta = 0$ is equivalent to $\gamma = 1/2$. The distribution is symmetric, with the nickname of "SaS", which stands for "Symmetric α-Stable".

Its Mellin transform is simplified to

$$\begin{aligned} L^0_\alpha(x) &\overset{\mathcal{M}}{\longleftrightarrow} \epsilon \left[\frac{\Gamma(s)}{\Gamma((1+s)/2)}\right] \left[\frac{\Gamma(\epsilon(1-s))}{\Gamma((1-s)/2)}\right] \\ &= \epsilon \left[\frac{2^{s-1}}{\sqrt{\pi}} \Gamma\left(\frac{s}{2}\right)\right] \left[\frac{\Gamma(\epsilon(1-s))}{\Gamma((1-s)/2)}\right]. \end{aligned} \tag{4.11}$$

The first bracket is the Mellin transform of a normal distribution (2.9) with a scale. The second bracket is $\overline{\chi}^{0}_{\alpha,1}{}^*(2-s)$ from above.

Hence, the PDF of SaS is

$$L^0_\alpha(x) = \int_0^\infty \mathcal{N}(xs)\, \overline{\chi}^{0}_{\alpha,1}(s)\, s\, ds. \tag{4.12}$$

This is one of the foundations of GAS-SN in (12.1).

4.4.1. Method of Normal Mixture. SaS in (4.12) will be generalized to GSaS in (12.3) in Chapter 12. Both integrals are in the normal mixture structure (9.1) that enjoys several nice properties described in Chapter 9.

The classic exponential power distribution (Section 3.11.1 of [**21**]) is the characteristic function transform in Lemma 9.2.

CHAPTER 5

Fractional Hypergeometric Functions

In this chapter, we extend both the confluent hypergeometric function ${}_1F_1(a, b; x)$ or $M(a, b; x)$ (Chapter 13, DLMF[5]); and the Gauss hypergeometric function ${}_2F_1(a, b, c; x)$ (Chapter 15 of DLMF).

The reader who is not interested in the hypergeometric functions can safely skip this chapter without losing direction.

The former occurs when dealing with the CDF of the GSC and FCM distributions. The latter occurs when handling the CDF of the GSaS and F distributions.

To clear up the situation, we first convert the DLMF formulas to our convention according to (2.2).

From DLMF 13.2.4 and 13.4.16, the Mellin transform of the Kummer function is

$$M(a, b; -x) = \frac{\Gamma(b)}{2\pi i\,\Gamma(a)} \int_{C-i\infty}^{C+i\infty} \frac{\Gamma(a-s)\Gamma(s)}{\Gamma(b-s)}\, x^{-s}\, ds,$$

where $a \neq 0, -1, -2, ...$

From DLMF 15.1.2 and 15.6.6, the Mellin transform of the Kummer function is

$${}_2F_1(a, b, c; -x) = \frac{\Gamma(c)}{2\pi i\,\Gamma(a)\Gamma(b)} \int_{C-i\infty}^{C+i\infty} \frac{\Gamma(a-s)\Gamma(b-s)\Gamma(s)}{\Gamma(c-s)}\, x^{-s}\, ds,$$

where $a, b \neq 0, -1, -2, ...$

Use our Mellin transform notation, they become

$$M(a, b; -x) \overset{\mathcal{M}}{\longleftrightarrow} M^*(a, b; s) = \frac{\Gamma(b)}{\Gamma(a)} \frac{\Gamma(a-s)\Gamma(s)}{\Gamma(b-s)}, \tag{5.1}$$

$${}_2F_1(a, b, c; -x) \overset{\mathcal{M}}{\longleftrightarrow} {}_2F_1^*(a, b, c; s) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a-s)\Gamma(b-s)\Gamma(s)}{\Gamma(c-s)}. \tag{5.2}$$

Now let us add the fractional components to them!

5.1. Fractional Confluent Hypergeometric Function

The fractional confluent hypergeometric function (FCHF) is the merger of the Kummer function and the Wright function. it allows us to extend many classic functions in their fractional forms.

We start with its Mellin transform. And derive the integral and series representations from it.

Definition 5.1. The Mellin transform of the FCHF is

$$M_{\lambda,\delta}(a, b; -x) \overset{\mathcal{M}}{\longleftrightarrow} M^*_{\lambda,\delta}(a, b; s) = \frac{\Gamma(b)}{\Gamma(a)} \frac{\Gamma(a-s)\Gamma(s)}{\Gamma(\delta-\lambda s)\Gamma(b-s)} \tag{5.3}$$

where the $\Gamma(\delta - \lambda s)$ term is from the Wright function (3.5).

Lemma 5.2. The integral representation from DLMF 13.4.1 is extended to

$$M_{\lambda,\delta}(a, b; z) = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} \int_0^1 W_{\lambda,\delta}(zt)\, t^{a-1}(1-t)^{b-a-1}\, dt \tag{5.4}$$

The obvious limit $W_{0,1}(zt) = e^{zt}$ restores it to the classic DLMF formula.

PROOF. Replace the Wright function in (5.4) with its Hankel integral (3.6),

$$M_{\lambda,\delta}(a,b;z) = \frac{\Gamma(b)}{2\pi i\,\Gamma(a)} \int_0^1 \int_{Ha} \left(\frac{e^{s+zt\,s^{-\lambda}}}{s^{\delta}} ds \right) t^{a-1}(1-t)^{b-a-1}\, dt$$

which can be simplified to

$$M_{\lambda,\delta}(a,b;z) = \frac{1}{2\pi i} \int_{Ha} \left(s^{-\delta}\, e^s\, ds \right) M(a,b;-z\,s^{-\lambda})$$

Substitute the Mellin integral from (5.1) to it,

$$\begin{aligned} M_{\lambda,\delta}(b,c;-z) &= \frac{1}{2\pi i} \int_{Ha} \left(s^{-\delta}\, e^s\, ds \right) \frac{\Gamma(b)}{2\pi i \Gamma(a)} \int_{C-i\infty}^{C+i\infty} \frac{\Gamma(b-t)\Gamma(t)}{\Gamma(c-t)} (z\,s^{-\lambda})^{-t}\, dt \\ &= \frac{\Gamma(b)}{2\pi i \Gamma(a)} \int_{C-i\infty}^{C+i\infty} \left[\frac{1}{2\pi i} \int_{Ha} s^{\lambda t-\delta}\, e^s\, ds \right] \frac{\Gamma(b-t)\Gamma(t)}{\Gamma(c-t)} z^{-t}\, dt \\ &= \frac{\Gamma(b)}{2\pi i \Gamma(a)} \int_{C-i\infty}^{C+i\infty} \left[\frac{1}{\Gamma(\delta-\lambda t)} \right] \frac{\Gamma(b-t)\Gamma(t)}{\Gamma(c-t)} z^{-t}\, dt \end{aligned}$$

which is the Mellin transform in (5.3).

From the second line to the third line, we use the well-known Hankel integral of the reciprocal gamma function:

$$\frac{1}{2\pi i} \int_{Ha} s^{-z}\, e^s\, ds = \frac{1}{\Gamma(z)}$$

□

LEMMA 5.3. The series representation is

$$M_{\lambda,\delta}(a,b;z) := \sum_{n=0}^{\infty} \left[\frac{(a)_n}{(b)_n\, \Gamma(\lambda n+\delta)} \right] \frac{z^n}{n!} \tag{5.5}$$

where $(a)_n, (b)_n$ are Pochhammer symbols.

PROOF. Take (5.3) and apply Ramanujan's Master Theorem from Section 2.2. This produces $(M^*_{\lambda,\delta}(a,b;s)/\Gamma(s))|_{s=-n}$, which is equal to the bracket term, since $(x)_n = \Gamma(x+n)/\Gamma(x)$. □

5.1.1. FCHF Subsumes the Kummer Function. It is obvious that $M_{0,1}(a,b;x) = M(a,b;x)$.

5.1.2. FCHF Subsumes the M-Wright Function. By using the same setting from (3.9), we get

$$M_\alpha(z) = M_{-\alpha,1-\alpha}(c,c;-z) \qquad (c \neq 0)$$

5.1.3. FCHF Subsumes Fractional Gamma-Star Function. An important variant of FCHF is the fractionalization of the incomplete gamma function. The reader is referred to Section 8 and Section 13 of DLMF[5] and Wikipedia for the background knowledge.

We are mainly concerned with the following setup:

$$M_{-\alpha,1-\alpha}(c,c+1;-x) = c \int_0^1 M_\alpha(xt)\, t^{c-1}\, dt$$

This integral is found in (3.18). Hence, we obtain -

LEMMA 5.4. The fractional γ^* function (3.18) has the following FCHF representation:

$$\gamma^*_\alpha(s,x) = \frac{\Gamma(\alpha s-\alpha+1)}{\Gamma(s+1)} M_{-\alpha,1-\alpha}(s,s+1;-x) \tag{5.6}$$

The fractional γ^* function is the foundation to express the CDF of GSC in Section 6.5. This was the original motivation to enrich the classic confluent hypergeometric function.

5.2. Fractional Gauss Hypergeometric Function

The Fractional Gauss Hypergeometric Function (FGHF) arises from the ratio distribution between an elementary function and FCM2 ($\widehat{\chi}^2_{\alpha,k}$) in Section 7.5. When $\alpha = 1$, the Mellin transform of FCM2 is reduced from a fractional form to a classic form in (7.25). The ratio distribution is reduced to a Gauss hypergeometric function ${}_2F_1$. Hence, we consider the general form of such ratio distribution as the fractional ${}_2F_1$.

We start by modifying the Mellin transform from (5.2) (DLMF 15.6.6). Then derive the integral and series representations from it.

Definition 5.5. The Mellin transform of the fractional Gauss hypergeometric function is

$$\begin{aligned} {}_2F_1(a,b,c,\epsilon;-x) &\overset{\mathcal{M}}{\longleftrightarrow} {}_2F_1^*(a,b,c,\epsilon;s) \\ &= \Big[M^*(a,c;s)\Big] \left[\frac{B(k/2,1/2)}{\Gamma(1/2)}\,\widehat{\chi}^{2*}_{\alpha,k}(3/2-s)\right] \end{aligned} \tag{5.7}$$

where $\epsilon = 1/\alpha$ is the convention from (4.1), and $b = (k+1)/2$. $M^*(a,c;s)$ is from (5.1), and $\widehat{\chi}^{2*}_{\alpha,k}(s)$ is from (7.24) (We jump ahead). And $B(x,y)$ is the beta function.

This structure is a fractional form of the generalized hypergeometric function ${}_3F_2$ (DLMF 16.5.1, replace s with $-s$). To see this, expand (5.7) and we get

$$ {}_2F_1^*(a,b,c,\epsilon;s) = \left[\frac{\Gamma(c)}{\Gamma(a)}\frac{\Gamma(a-s)\Gamma(s)}{\Gamma(c-s)}\right]\left[2^{2s-1}\,\frac{B(k/2,1/2)}{\Gamma(1/2)}\,\frac{\Gamma((k-1)/2)}{\Gamma(\epsilon(k-1))}\,\frac{\Gamma(2\epsilon\,(k/2-s))}{\Gamma(k/2-s)}\right]. \tag{5.8}$$

There are five gamma functions containing s: three in the numerator, two in the denominator. The $\Gamma(2\epsilon\,(k/2-s))$ term is fractional.

5.2.1. FGHF Subsumes the Gauss Hypergeometric Function.

Lemma 5.6. When $\epsilon = 1$,

$$ {}_2F_1^*(a,b,c,\epsilon=1;s) = {}_2F_1^*(a,b,c;s)$$

Proof. Let $\epsilon = 1$, the second bracket becomes

$$\frac{B(k/2,1/2)}{\Gamma(1/2)}\frac{\Gamma(k/2+1/2-s)}{\Gamma(k/2)} = \frac{\Gamma(k/2+1/2-s)}{\Gamma(k/2+1/2)} = \frac{\Gamma(b-s)}{\Gamma(b)}. \tag{5.9}$$

Hence, (5.7) is reduced to the classic limit of ${}_2F_1^*(a,b,c;s)$ in (5.2). □

5.2.2. The Integral Form.

Lemma 5.7. The integral form of FGHF is

$$ {}_2F_1(a,b,c,\epsilon;-x) := \frac{B(k/2,1/2)}{\Gamma(1/2)}\int_0^\infty M(a,c;-x\nu)\,\widehat{\chi}^2_{\alpha,k}(\nu)\,\sqrt{\nu}d\nu \tag{5.10}$$

where $\epsilon = 1/\alpha$ and $b = (k+1)/2$. $M(a,c;x)$ is the Kummer function (Chapter 13, DLMF). $\widehat{\chi}^2_{\alpha,k}(x)$ is from (7.16).

PROOF. We use the generalized convolution formula:

$$h(x) = \int_0^\infty f(xs)g(s)\, s^p\, ds \overset{\mathcal{M}}{\longleftrightarrow} h^*(s) = f^*(s)g^*(1+p-s),$$

Clearly f is M, and g is $\widehat{\chi}^2_{\alpha,k}$. Substitute $p = 1/2$ due to the $\sqrt{\nu}$ term. The Mellin transform of (5.10) is

$${}_2F_1(a,b,c,\epsilon;-x) \overset{\mathcal{M}}{\longleftrightarrow} \frac{B(k/2,1/2)}{\Gamma(1/2)} M^*(a,c;s)\, \widehat{\chi}^{2*}_{\alpha,k}(3/2-s)$$

This is exactly (5.7).

□

5.2.3. Relation between FGHF and Real-World Usage. This section addresses a broader issue: How does FGHF relate to FCM and GAS (and GAS-SN) in general? The reader can skip this section and come back later after she read the later chapters.

This topic is important. In an abstract sense, most of the univariate PDFs in their ratio distribution forms can be understood by the integral form of FGHF.

Let's make (5.10) more abstract, by ignoring some cumbersome parameters. Assume $F(-x) := {}_2F_1(a,b,c,\epsilon;-x)$ and $M(-x) := M^*(a,c;-x)$ $(x \geq 0)$, then (5.10) becomes

$$F(-x) := B \int_0^\infty M(-x\nu)\, \overline{\chi}^2_{\alpha,k}(\nu;\sigma=1/4)\, \sqrt{\nu} d\nu \tag{5.11}$$

where we employ the notation $\widehat{\chi}^2_{\alpha,k}(x) = \overline{\chi}^2_{\alpha,k}(x;\sigma=\frac{1}{4})$ from (7.16), and $B := B(\frac{k}{2},\frac{1}{2})/\Gamma(\frac{1}{2})$.

LEMMA 5.8. Let $F'(-x)$ be the scaled FGHF, which is more closely related to the real-world use cases. The following ratio-distribution integrals can be converted to F' such as

$$\left.\begin{matrix} \int_0^\infty M(-xs)\, \overline{\chi}^2_{\alpha,k}(s)\, \sqrt{s} ds \\ \int_0^\infty M(-xs^2)\, \overline{\chi}_{\alpha,k}(s)\, s\, ds \end{matrix}\right\} = F'(-x) := \frac{2\sqrt{\pi}\, \sigma_{\alpha,k}}{B(\frac{k}{2},\frac{1}{2})} F\left(-4\sigma^2_{\alpha,k}\, x\right) \tag{5.12}$$

Or use the full FGHF notation explicitly:

$$\left.\begin{matrix} \int_0^\infty M(a,c;-xs)\, \overline{\chi}^2_{\alpha,k}(s)\, \sqrt{s} ds \\ \int_0^\infty M(a,c;-xs^2)\, \overline{\chi}_{\alpha,k}(s)\, s\, ds \end{matrix}\right\} = F'_{\alpha,k}(a,c;-x) := \frac{2\sqrt{\pi}\, \sigma_{\alpha,k}}{B(\frac{k}{2},\frac{1}{2})}\, {}_2F_1\left(a,b,c,\epsilon;-4\sigma^2_{\alpha,k}\, x\right) \tag{5.13}$$

where $\epsilon = 1/\alpha$ and $b = (k+1)/2$ on the RHS.

PROOF. Let Q be the scale that we aim to solve. (5.11) is rewritten to $F'(-x)$ such that

$$F'(-x) := \frac{\sqrt{Q}}{B} F(-Qx) = \sqrt{Q} \int_0^\infty M(-Qx\nu)\, \overline{\chi}^2_{\alpha,k}(\nu;\sigma=1/4)\, \sqrt{\nu} d\nu.$$

Let $s = Q\nu$,

$$\begin{aligned} F'(-x) &= \int_0^\infty M(-xs)\, \overline{\chi}^2_{\alpha,k}(s/Q;\sigma=1/4)/Q\, \sqrt{s} ds \\ &= \int_0^\infty M(-xs)\, \overline{\chi}^2_{\alpha,k}(s;\sigma=Q/4)\, \sqrt{s} ds \end{aligned}$$

Let $Q = 4\sigma^2_{\alpha,k}$, we obtain the integral form in terms of FCM2,

$$F'(-x) = \int_0^\infty M(-xs)\, \overline{\chi}^2_{\alpha,k}(s)\, \sqrt{s} ds$$

This is the first line of (5.12). Then apply (7.18) and (7.19) to get the second line. And on the FGHF side, we have

$$F'(-x) = \frac{\sqrt{Q}}{B} F\left(-Qx\right) = \frac{2\sqrt{\pi}\,\sigma_{\alpha,k}}{B(\frac{k}{2},\frac{1}{2})}\,F(-4\sigma_{\alpha,k}^2\,x)$$

□

5.2.4. Example 1: GSaS. In Lemma 8.3 of [**13**], a fractional extension was explored for the CDF of GSaS. We formalized it further here. However, we note that the $M(-x)$ function needed to describe GAS-SN is more complicated than a Kummer function. See (10.2) and (10.3).

LEMMA 5.9. Assume $\Phi[L_{\alpha,k}](x)$ is the CDF of a GSaS, which is (12.2) with $\beta = 0$. It can be expressed by the scaled FGHF via

$$\Phi[L_{\alpha,k}](x) = \frac{1}{2} + \frac{x}{\sqrt{2\pi}}\,F'_{\alpha,k}\left(\frac{1}{2},\frac{3}{2};-\frac{x^2}{2}\right). \tag{5.14}$$

PROOF. From Lemma 8.3 of [**13**],

$$\Phi[L_{\alpha,k}](x) = \frac{1}{2} + \frac{x}{\sqrt{k}} M_{\alpha,k}\left(a,c;-\frac{x^2}{k}\right),$$

where $a = \frac{1}{2}, c = \frac{3}{2}$ and

$$M_{\alpha,k}(a,c;x) := \sqrt{\frac{k}{2\pi}}\int_0^\infty s\,ds\,M\left(a,c;\frac{xks^2}{2}\right)\,\overline{\chi}_{\alpha,k}(s).$$

This pattern fits right in with the second line of (5.13). It is immediately clear that its $M_{\alpha,k}(a,c;x)$ is our $\sqrt{k/2\pi}F'_{\alpha,k}\left(a,c;kx/2\right)$. Therefore,

$$\Phi[L_{\alpha,k}](x) = \frac{1}{2} + \frac{x}{\sqrt{2\pi}}\,F'_{\alpha,k}\left(a,c;-\frac{x^2}{2}\right),$$

where $a = \frac{1}{2}, c = \frac{3}{2}$.

□

Notice that this formula is much cleaner, without the cluttering of k in the previous attempt in [**13**].

5.2.5. Example 2: Fractional F.

LEMMA 5.10. From (8.2), the standard CDF of a fractional F distribution $F_{\alpha,d,k}$ is

$$\Phi[F_{\alpha,d,k}](x) = \frac{1}{\Gamma\left(\frac{d}{2}\right)}\int_0^\infty ds\,\gamma\left(\frac{d}{2},\frac{dxs}{2}\right)\,\overline{\chi}^2_{\alpha,k}(s).$$

It can be expressed by the scaled FGHF via

$$\Phi[F_{\alpha,d,k}](x) = \left[C_{\alpha,d,k}\,\frac{(dx/2)^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)}\right]\,F'_{\alpha,k+d-1}\left(\frac{d}{2},\frac{d}{2}+1;-\frac{dx}{2\Sigma}\right). \tag{5.15}$$

where $C_{\alpha,d,k}$ is defined in (5.16) and $\Sigma := \sigma^2_{\alpha,k+d-1}/\sigma^2_{\alpha,k}$.

PROOF. Note that

$$\frac{1}{\Gamma\left(\frac{d}{2}\right)}\gamma\left(\frac{d}{2},\frac{x}{2}\right) = \frac{(x/2)^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)}M\left(\frac{d}{2},\frac{d}{2}+1;-\frac{x}{2}\right).$$

Then

$$\Phi[F_{\alpha,d,k}](x) = \int_0^\infty \left[\frac{(dxs/2)^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)} M\left(\frac{d}{2}, \frac{d}{2}+1; -\frac{dxs}{2}\right) \right] \overline{\chi}^2_{\alpha,k}(s)\, ds$$
$$= \frac{(dx/2)^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)} \int_0^\infty M\left(\frac{d}{2}, \frac{d}{2}+1; -\frac{dxs}{2}\right) s^{(d-1)/2}\, \overline{\chi}^2_{\alpha,k}(s) \sqrt{s}\, ds.$$

When $d = 1$, it fits right in with FGHF. When $d > 1$, it needs more work.

From (7.5), let $m = (d-1)/2$, then $k + 2m = k + d - 1$ and

$$\Phi[F_{\alpha,d,k}](x) = C_{\alpha,d,k} \frac{(dx/2)^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)} \int_0^\infty M\left(\frac{d}{2}, \frac{d}{2}+1; -\frac{dxy}{2\Sigma}\right) \overline{\chi}^2_{\alpha,k+d-1}(y) \sqrt{y}\, dy,$$

where $\Sigma := \sigma^2_{\alpha,k+d-1}/\sigma^2_{\alpha,k}$ and $y = \Sigma\, s$, and

$$C_{\alpha,d,k} := \frac{\sigma^{d-1}_{\alpha,k}}{\sqrt{\Sigma}} \frac{C_{\alpha,k}}{C_{\alpha,k+d-1}} = \frac{\sigma^{d}_{\alpha,k}}{\sigma_{\alpha,k+d-1}} \frac{C_{\alpha,k}}{C_{\alpha,k+d-1}}. \tag{5.16}$$

The integral matches the FGHF pattern in Lemma 5.12, and we get (5.15).

□

One final note. There is a connection between (5.14) and (5.15). When $d = 1$, $\Sigma = 1$ and $C_{\alpha,d,k} = 1$. Then

$$\Phi[F_{\alpha,1,k}](x^2) = \frac{2x}{\sqrt{2\pi}} F'_{\alpha,k}\left(\frac{1}{2}, \frac{3}{2}; -\frac{x^2}{2}\right) \tag{5.17}$$

which is $2\,\Phi[L_{\alpha,k}](x) - 1$ in (5.14).

This reflects Lemma 8.3 that, if variable X distributes like a GSaS $L_{\alpha,k}$, then X^2 distributes like a one-dimensional F, aka $F_{\alpha,1,k}$. It is particularly easy to see this relation in the FGHF form.

Part 2

One-Sided Distributions

CHAPTER 6

GSC: Generalized Stable Count Distribution

GSC is the backbone that allows many features in this book. In particular, FCM is a member of GSC. The name "stable count distribution" came from my 2020 work[**12**]. If I could forget about the history and name it again, I would call it *fractional gamma distribution*. It is the fractional version of the generalized gamma distribution, as would become clear to the reader in this chapter.

6.1. Definition

DEFINITION 6.1 (Generalized stable count distribution (GSC)). GSC is a four-parameter one-sided distribution family, whose PDF is defined as

$$\mathfrak{N}_\alpha(x;\sigma,d,p) := C\left(\frac{x}{\sigma}\right)^{d-1} F_\alpha\left(\left(\frac{x}{\sigma}\right)^p\right) \qquad (x \geq 0) \tag{6.1}$$

where $F_\alpha(x) = W_{-\alpha,0}(-x)$ from (3.8) and $\alpha \in [0,1]$ controls the shape of the Wright function; σ is the scale parameter; p is also the shape parameter controlling the tail behavior ($p \neq 0, dp \geq 0$); d is the *degree of freedom* parameter. When $\alpha \to 1$, the PDF becomes a Dirac delta function: $\delta(x-\sigma)$ assuming σ is finite. When $d \geq 1$, all the moments of the GSC exist and have closed forms.

6.2. Determination of C

The normalization constant C is:

$$C = \begin{cases} \frac{|p|}{\sigma}\frac{\Gamma(\alpha d/p)}{\Gamma(d/p)} & \text{, for } \alpha \neq 0, d \neq 0. \\ \frac{|p|}{\sigma\alpha} & \text{, for } \alpha \neq 0, d = 0. \end{cases} \tag{6.2}$$

It is important to note that d and p are allowed to be negative, as long as $dp \geq 0$.

PROOF. The normalization constant C in (6.1) is obtained from the requirement that the integral of the PDF must be 1:

$$\int_0^\infty \mathfrak{N}_\alpha(x;\sigma,d,p)\,dx = \frac{C\sigma}{|p|}\frac{\Gamma(\frac{d}{p})}{\Gamma(\frac{d}{p}\alpha)} = 1$$

where the integral is carried out by the moment formula of the Wright function.

We typically constrain $dp \geq 0$ and p is typically positive. But it becomes negative in the inverse distribution and/or characteristic distribution types. So we need $|p|$ to ensure C is positive.

For the case of $\alpha \neq 0$ and $d \to 0$, due to (A.3), we have

$$C = \frac{|p|}{\sigma\alpha} \qquad (\alpha \neq 0, d = 0)$$

These two cases are combined to form (6.2). $\square$

6.3. GSC Subsumes Generalized Gamma Distribution

Since the Wright function extends an exponential function to the fractional space, GSC is the fractional extension of the generalized gamma (GG) distribution[25], whose PDF is defined as:

$$f_{\mathrm{GG}}(x; a, d, p) = \frac{|p|}{a\Gamma(d/p)} \left(\frac{x}{a}\right)^{d-1} e^{-(x/a)^p}. \tag{6.3}$$

The parallel use of parameters is obvious, except that a in GG is replaced with σ in GSC to avoid confusion with α.

GG is subsumed to GSC in two ways:

$$f_{\mathrm{GG}}(x; \sigma, d, p) := \begin{cases} \mathfrak{N}_0(x; \sigma, d = d-p, p) & , \text{ at } \alpha = 0. \\ \mathfrak{N}_{\frac{1}{2}}\left(x; \sigma = \frac{\sigma}{2^{2/p}}, d = d - \frac{p}{2}, p = \frac{p}{2}\right) & , \text{ at } \alpha = \frac{1}{2}. \end{cases} \tag{6.4}$$

The first line is treated as the definition of GSC at $\alpha = 0$. The proof is in [13].

Although the first line is more obvious, it is the second line that leads to the fractional extension of the χ distribution.

6.4. Mellin Transform

From Example 2.2, we add σ and C. The Mellin transform of GSC is

$$\begin{aligned} \mathfrak{N}_\alpha(x; \sigma, d, p) &\overset{\mathcal{M}}{\longleftrightarrow} \frac{C\,\sigma^s}{|p|} \frac{\Gamma((s+d-1)/p)}{\Gamma(\alpha(s+d-1)/p)} \\ &= \sigma^{s-1} \frac{\Gamma(\alpha d/p)}{\Gamma(d/p)} \frac{\Gamma((s+d-1)/p)}{\Gamma(\alpha(s+d-1)/p)}, \end{aligned} \tag{6.5}$$

where C is from Section 6.2. The typical limiting case for the gamma functions shall be taken care in each scenario.

GSC is often used in a ratio distribution, such as the role of $g^*(s)$ in (2.6), where $s \to 2-s$. The $s+d-1$ term becomes $d+1-s$. Furthermore, in the FCM case, since $d = k-1$, it becomes the elegant $k-s$ term.

6.5. CDF and Fractional Incomplete Gamma Function

The CDF of GSC is

$$\Phi(x) := \int_0^x \mathfrak{N}_\alpha(s; \sigma, d, p)\, ds \quad (x \geq 0). \tag{6.6}$$

This integral leads to the fractionalization of the incomplete gamma function in Section 3.4.

Lemma 6.2. The CDF of GSC can be represented by γ^*_α in (3.18) as

$$\Phi(x) = z^{d+p}\, \gamma^*_\alpha\left(d/p + 1, z^p\right) \tag{6.7}$$

where $z = x/\sigma$ is the standardized variable.

This could be viewed as one form of fractional extension to the regularized lower incomplete function, $\gamma(s,z)/\Gamma(s)$, which is the CDF of GG mentioned above.

Due to this result, it may even be suitable to call GSC as the *fractional gamma distribution*.

PROOF. The CDF of GSC is

$$\begin{aligned}\Phi(x) &= \int_0^x N_\alpha(s;\sigma,d,p)\,ds \\ &= C\int_0^x ds\left(\frac{s}{\sigma}\right)^{d-1} W_{-\alpha,0}\left(-\left(\frac{s}{\sigma}\right)^p\right). \\ &= C\int_0^x ds\left(\frac{s}{\sigma}\right)^{d-1} F_\alpha\left(\left(\frac{s}{\sigma}\right)^p\right).\end{aligned}$$

Since $F_\alpha(x) = \alpha x\, M_\alpha(x)$ from (3.8), and let $u = s/x$, then

$$\begin{aligned}\Phi(x) &= \alpha C\int_0^x ds\left(\frac{s}{\sigma}\right)^{d+p-1} M_\alpha\left(\left(\frac{s}{\sigma}\right)^p\right) \\ &= \alpha C\,x\int_0^1 du\left(\frac{xu}{\sigma}\right)^{d+p-1} M_\alpha\left(\left(\frac{xu}{\sigma}\right)^p\right)\end{aligned}$$

Recognize that, if $u \in [0,1]$, then $u^p \in [0,1]$. Let $t = u^p$, and $dt/t = p\,du/u$,

$$\Phi(x) = \frac{\alpha\sigma C}{p}\, z^{d+p}\int_0^1 dt\, t^{d/p}\, M_\alpha\left(z^p t\right)$$

Compare the last line with γ^*_α in (3.18), and we get

$$\Phi(x) = \frac{\alpha\sigma C}{p}\,\frac{\Gamma(\frac{d}{p}+1)}{\Gamma((1-\alpha)+\alpha(\frac{d}{p}+1))}\, z^{d+p}\,\gamma^*_\alpha\left(d/p+1, z^p\right)$$

Using the case of $\alpha \neq 0, d \neq 0$ for C, it can be shown that the constant part is just 1. Hence,

$$\Phi(x) = z^{d+p}\,\gamma^*_\alpha\left(d/p+1, z^p\right)$$

$\square$

CHAPTER 7

Fractional Chi Distributions

7.1. Introduction to Fractional Chi Distribution

In Chapter 4, we've discussed the insight that leads to the fractional χ is to interpret the Mellin transform of the PDF of the α-stable distribution as a ratio distribution of two components:

$$L_\alpha^\theta(x) \overset{\mathcal{M}}{\longleftrightarrow} \epsilon \left[\frac{\Gamma(s)}{\Gamma(1-\gamma+\gamma s))}\right] \left[\frac{\Gamma(\epsilon(1-s))}{\Gamma(\gamma(1-s))}\right]$$
$$\text{where } \epsilon = \frac{1}{\alpha}, \gamma = \frac{\alpha-\theta}{2\alpha}.$$

The first bracket is the Mellin transform of the M-Wright function.

The second bracket is interpreted as the Mellin transform of the PDF of the fractional χ-mean distribution (FCM) at $k=1$:

$$\overline{\chi}_{\alpha,1}^\theta(x) \overset{\mathcal{M}}{\longleftrightarrow} \overline{\chi}_{\alpha,1}^{\theta}{}^*(s) \ \propto \frac{\Gamma(\epsilon(s-1))}{\Gamma(\gamma(s-1))},$$

apart from the normalization constant and scale in the PDF.

It becomes obvious after replacing $s \to 2-s$ in $\overline{\chi}_{\alpha,1}^{\theta}{}^*(s)$ in order to comply with the rule of Mellin transform of a ratio distribution.

In this chapter, the "degrees of freedom" parameter k is inserted by replacing $s-1$ with $s+k-2$, such that

$$\overline{\chi}_{\alpha,k}^\theta(x) \overset{\mathcal{M}}{\longleftrightarrow} \overline{\chi}_{\alpha,k}^{\theta}{}^*(s) \ \propto \frac{\Gamma(\epsilon(s+k-1))}{\Gamma(\gamma(s+k-1))}. \tag{7.1}$$

This forms the foundation for more rigorous treatment of FCM.

7.2. FCM: Fractional Chi-Mean Distribution

There are two ways to define FCM. The first approach is to define it via Mellin transform. The second approach is to define the shape of its PDF.

Definition 7.1 (Fractional χ-mean distribution (FCM) via Mellin Transform). The Mellin transform of FCM's PDF is enriched from (7.1) to

$$\begin{aligned} \overline{\chi}_{\alpha,k}^\theta(x) &\overset{\mathcal{M}}{\longleftrightarrow} \overline{\chi}_{\alpha,k}^{\theta}{}^*(s) \\ &= (\sigma_{\alpha,k}^\theta)^{s-1} \frac{\Gamma(\gamma(k-1))}{\Gamma(\epsilon(k-1))} \frac{\Gamma(\epsilon(s+k-2))}{\Gamma(\gamma(s+k-2))}, \\ &\text{where } \sigma_{\alpha,k}^\theta := \gamma^\gamma\, k^{\gamma-\epsilon}. \end{aligned} \tag{7.2}$$

The main differences are (1) to address the normalization of the total density, and (2) to have a proper scale $\sigma_{\alpha,k}^\theta$ such that it is consistent with the classic χ distribution ans α-stable distribution.

For positive k, the PDF of an FCM is

$$\overline{\chi}^{\theta}_{\alpha,k}(x) := \mathfrak{N}_{\gamma\alpha}(x; \sigma = \sigma^{\theta}_{\alpha,k}, d = k-1, p = \alpha) \qquad (x \geq 0)$$
$$= \frac{\Gamma(\gamma(k-1))}{\epsilon\Gamma(\epsilon(k-1))} \left(\sigma^{\theta}_{\alpha,k}\right)^{1-k} x^{k-2} F_{\gamma\alpha}\left(\left(\frac{x}{\sigma^{\theta}_{\alpha,k}}\right)^{\alpha}\right), \tag{7.3}$$

where $\mathfrak{N}_{\lambda}(x; \sigma, d, p)$ is GSC (6.1), and $F_{\lambda}(x) := W_{-\lambda,0}(-x)$ is the Wright function of the second kind (3.8).

Notice the appearances of γ that replaces all the $1/2$ in Section 7.6 of [**13**]. That is is how θ comes into play in the upgraded FCM. This full representation is used in Chapter 11.

However, for GAS-SN in Chapter 12 and beyond, such θ upgrade is unnecessary. The skew-normal framework is based on modulation of normal distributions. It is required to have $\theta = 0$ ($\gamma = 1/2$).

Hence, we recite the original definition of FCM PDF ($k > 0$):

$$\overline{\chi}_{\alpha,k}(x) = \overline{\chi}^{0}_{\alpha,k}(x) := \mathfrak{N}_{\alpha/2}(x; \sigma = \sigma_{\alpha,k}, d = k-1, p = \alpha) \qquad (x \geq 0)$$
$$= \left(C_{\alpha,k}\right)\left(\sigma_{\alpha,k}\right)^{1-k} x^{k-2} F_{\frac{\alpha}{2}}\left(\left(\frac{x}{\sigma_{\alpha,k}}\right)^{\alpha}\right), \tag{7.4}$$

where

$$C_{\alpha,k} := \frac{\alpha\Gamma((k-1)/2)}{\Gamma((k-1)/\alpha)} \tag{7.5}$$

$$\sigma_{\alpha,k} := \frac{|k|^{1/2-1/\alpha}}{\sqrt{2}}. \tag{7.6}$$

Note that the difference between (7.3) and (7.4) is very small: Just replace $\mathfrak{N}_{\gamma\alpha}(...)$ to $\mathfrak{N}_{\alpha/2}(...)$.

7.2.1. FCM CDF. Extending directly from Lemma 6.2, we have

LEMMA 7.2. *The CDF of FCM can be represented by γ^{*}_{α} as*

$$\Phi[\overline{\chi}_{\alpha,k}](x) = z^{k-1+\alpha}\, \gamma^{*}_{\alpha/2}\left(\frac{k-1+\alpha}{\alpha}, z^{\alpha}\right), \qquad (k > 0, \alpha \in [0,2]) \tag{7.7}$$

where $z = x/\sigma_{\alpha,k}$.

7.3. FCM Moments

By letting $s = n+1$ and $\theta = 0$ in (7.2), its n-th moment is

$$\mathbb{E}(X^n|\overline{\chi}_{\alpha,k}) = (\sigma_{\alpha,k})^n \frac{\Gamma((k-1)/2)}{\Gamma((k-1)/\alpha)} \frac{\Gamma((n+k-1)/\alpha)}{\Gamma((n+k-1)/2)}, \qquad (k > 0, \alpha > 0) \tag{7.8}$$

which requires $k > 1$ and $n + k > 1$ to avoid singularity of the gamma functions (See Section 7.6 of [**13**]).

The moment formula of FCM is fundamental to all the fractional distributions built on top of it. But ironically, due to the nature of a ratio distribution, it is often evaluated as negative moments, $n < 0$. Hence, n is confined in the range of $1 - k < n < 0$.

This results in non-existing moments when k is not "large enough", which happens to be a core feature of the α-stable distribution and Student's t distribution. Our two-dimensional parameter space (α, k) adds more complexity to it.

7.3.1. FCM at Infinite Degrees of Freedom. The choice of $\sigma_{\alpha,k}$ is intentional, such that

$$\lim_{k\to\infty} \mathbb{E}(X^n|\overline{\chi}_{\alpha,k}) = \alpha^{-n/\alpha}. \qquad (k>0, \alpha>0) \tag{7.9}$$

Under such condition, its variance is zero. That is, FCM becomes a delta function, $\delta(x-\alpha^{-1/\alpha})$, as $k\to\infty$.

7.4. FCM Reflection Formula

When $k<0$, the PDF of FCM is defined as

$$\overline{\chi}_{\alpha,k}(x) := \mathfrak{N}_{\alpha/2}(x; \sigma = 1/\sigma_{\alpha,k}, d=k, p=-\alpha) \qquad (k<0). \tag{7.10}$$

But it also noted that we might not repeat the $k<0$ scenario everywhere. It is too tedious to the readers. So we choose not to do it for conciseness. The readers interested in full detail are referred to the FCM sections in [**13**].

The $k<0$ case is born out of the properties of the α-stable characteristic function in Chapter 9. It is used to build a generalized two-sided distribution (Section 9 of [**13**]) that subsumes the exponential power distribution (Section 3.11.1 of [**21**]).

Here we quote the FCM reflection formula from Section 7 of [**13**] to summarize the relation:

$$\mathbb{E}(X^n|\overline{\chi}_{\alpha,-k}) = \frac{\mathbb{E}(X^{-n+1}|\overline{\chi}_{\alpha,k})}{\mathbb{E}(X|\overline{\chi}_{\alpha,k})},\ k>0. \tag{7.11}$$

7.5. FCM2: Fractional Chi-Squared-Mean Distribution

If $Z \sim \overline{\chi}_{\alpha,k}$, then $X \sim Z^2$ is FCM2 such as $X \sim \overline{\chi}^2_{\alpha,k}$. This is the fractional extension of the classic χ^2_k/k, which is subsumed by it at $\alpha = 1$.

$\overline{\chi}^2_{\alpha,k}$ is used in the fractional F distribution in the area of the squared variable and the quadratic form in the multivariate elliptical distribution.

Definition 7.3. The PDF of FCM2 is

$$\overline{\chi}^2_{\alpha,k}(x) = \frac{1}{2\sqrt{x}}\,\overline{\chi}_{\alpha,k}(\sqrt{x}) \qquad (x \ge 0, \alpha \in [0,2]) \tag{7.12}$$

Expressed in GSC and (7.4), it is

$$\begin{aligned} \overline{\chi}^2_{\alpha,k}(x) &:= \mathfrak{N}_{\alpha/2}(x; \sigma = \sigma^2_{\alpha,k}, d = (k-1)/2, p = \alpha/2) \qquad (k > 0) \\ &= \frac{C_{\alpha,k}}{2\,\sigma^2_{\alpha,k}} \left(\frac{x}{\sigma^2_{\alpha,k}}\right)^{k/2-3/2} F_{\frac{\alpha}{2}}\left(\left(\frac{x}{\sigma^2_{\alpha,k}}\right)^{\alpha/2}\right). \end{aligned} \tag{7.13}$$

Or for $k < 0$,

$$\overline{\chi}^2_{\alpha,k}(x) := \mathfrak{N}_{\alpha/2}(x; \sigma = \sigma^{-2}_{\alpha,k}, d = k/2, p = -\alpha/2) \qquad (k < 0) \tag{7.14}$$

When dealing with the fractional Gauss hypergeometric function (FGHF) in Section 5.2, we need two more variations from FCM2. The first allows an FCM2 to take a different scale:

$$\overline{\chi}^2_{\alpha,k}(x;\sigma) := \mathfrak{N}_{\alpha/2}(x; \sigma = \sigma, d = (k-1)/2, p = \alpha/2) \qquad (k > 0) \tag{7.15}$$

from which the constant-scale variant is defined by replacing $\sigma_{\alpha,k}$ with $1/2$,

$$\widehat{\chi}^2_{\alpha,k}(x) := \overline{\chi}^2_{\alpha,k}(x; \sigma = 1/4) = \mathfrak{N}_{\alpha/2}(x; \sigma = 1/4, d = (k-1)/2, p = \alpha/2) \qquad (k > 0) \tag{7.16}$$

Notice the hat symbol replaces the bar symbol.

7.5.1. FCM2 CDF.

Extending directly from Lemma 6.2, we have:

Lemma 7.4. The CDF of FCM2 can be represented by γ^*_α as

$$\Phi[\overline{\chi}^2_{\alpha,k}](x) = z^{(k-1+\alpha)/2}\,\gamma^*_{\alpha/2}\left(\frac{k-1+\alpha}{\alpha}, z^{\alpha/2}\right) \qquad (k > 0, \alpha \in [0,2]) \tag{7.17}$$

where $z = x/\sigma^2_{\alpha,k}$.

7.5.2. Representing FCM by FCM2.

In (7.12), let $s = \sqrt{x}$, we get the inverse relation:

$$\overline{\chi}_{\alpha,k}(s) = 2s\,\overline{\chi}^2_{\alpha,k}(s^2) \qquad (s \ge 0) \tag{7.18}$$

Many ratio distribution integrals involving FCM can be rewritten in terms of FCM2, such that

$$\begin{aligned} f(x) &:= \int_0^\infty g(xs)\,\overline{\chi}_{\alpha,k}(s)\,s\,ds \\ &= \int_0^\infty g(x\sqrt{\nu})\,\overline{\chi}^2_{\alpha,k}(\nu)\,\sqrt{\nu}\,d\nu \end{aligned} \tag{7.19}$$

For the CDF case, the incomplete integral can be transformed as

$$\begin{aligned} F(x) := \int_0^x f(x)\,dx &= \int_0^\infty G(xs)\,\overline{\chi}_{\alpha,k}(s)\,ds \\ &= \int_0^\infty G(x\sqrt{\nu})\,\overline{\chi}^2_{\alpha,k}(\nu)\,d\nu \end{aligned} \tag{7.20}$$

where $G(x) := \int_0^x g(x)\,dx$. The lower bound of the incomplete integrals can be $-\infty$ such as $\int_{-\infty}^x dx$ too.

7.5.3. Universal Expression. Assume $x \geq 0$, let $M(x^2) := G(x)/x$ in (7.20) or $g(x)$ in (7.19), we get the universal expression of

$$F(x) = x \int_0^\infty M(x^2\nu)\, \overline{\chi}^2_{\alpha,k}(\nu)\, \sqrt{\nu}\, d\nu \tag{7.21}$$

$$f(x) = \int_0^\infty M(x^2\nu)\, \overline{\chi}^2_{\alpha,k}(\nu)\, \sqrt{\nu}\, d\nu \tag{7.22}$$

Most of the univariate PDFs and CDFs in subsequent chapters can be understood in such framework. It is just a matter of what $M(x)$ is.

When $M(x)$ can be expressed by a Kummer function (apart from a negative sign), these integrals are members of the FGHF in Section 5.2.

7.6. FCM2 Mellin Transform

From (6.5), the Mellin transform of FCM2's PDF is

$$\begin{aligned} \overline{\chi}^2_{\alpha,k}(x) &\overset{\mathcal{M}}{\longleftrightarrow} \overline{\chi}^{2*}_{\alpha,k}(s) \\ &= (\sigma_{\alpha,k})^{2s-2}\, \frac{\Gamma((k-1)/2)}{\Gamma((k-1)/\alpha)}\, \frac{\Gamma((s+k/2-3/2)\,\times 2/\alpha)}{\Gamma(s+k/2-3/2)}. \qquad (k>0) \end{aligned} \tag{7.23}$$

Likewise, for the constant-scale variant, it becomes

$$\begin{aligned} \widehat{\chi}^2_{\alpha,k}(x) &\overset{\mathcal{M}}{\longleftrightarrow} \widehat{\chi}^{2*}_{\alpha,k}(s) \\ &= 2^{2-2s}\, \frac{\Gamma((k-1)/2)}{\Gamma((k-1)/\alpha)}\, \frac{\Gamma((s+k/2-3/2)\,\times 2/\alpha)}{\Gamma(s+k/2-3/2)}, \qquad (k>0) \end{aligned} \tag{7.24}$$

whose most important special case is $\alpha = 1$,

$$\widehat{\chi}^2_{1,k}(x) \overset{\mathcal{M}}{\longleftrightarrow} \widehat{\chi}^{2*}_{1,k}(s) = \frac{\Gamma(s+k/2-1)}{\Gamma(k/2)} \tag{7.25}$$

$\Gamma(s+k/2-1)$ in $\widehat{\chi}^{2*}_{1,k}(s)$ is just an ordinary gamma function without a fractional coefficient in front of s. This property is the basis that connects the fractional Gauss hypergeometric function to its classic form in Section 5.2.

7.7. FCM2 Moments

From the Mellin transform by $s = n+1$, its n-th moment is

$$\begin{aligned} \mathbb{E}(X^n|\overline{\chi}^2_{\alpha,k}) &= \mathbb{E}(X^{2n}|\overline{\chi}_{\alpha,k}) \\ &= (\sigma_{\alpha,k})^{2n}\, \frac{\Gamma((k-1)/2)}{\Gamma((k-1)/\alpha)}\, \frac{\Gamma((n+k/2-1/2)\,\times 2/\alpha)}{\Gamma(n+k/2-1/2)}. \qquad (k>0) \end{aligned} \tag{7.26}$$

As mentioned in Section 7.3, due to the nature of a ratio distribution, it is often evaluated as negative moments, $n < 0$. Hence, n is confined in the range of $1/2 - k/2 < n < 0$.

This puts stricter constraint on non-existing moments than FCM when k is not "large enough". For instance, in the case of fractional F distribution in Section 8.4, $k \approx 3$ is in the neighborhood where it second moment barely exists. This makes it rather hard for the statistics of the SPX daily return data set, since its k is just slightly larger than 3 while α is slightly below 1.

7.8. FCM2 Increment of k

LEMMA 7.5. When x^m is multiplied to $\overline{\chi}^2_{\alpha,k}(x)$, it follows a scaling rule where k is incremented to $k+2m$ in the parametrization.

$$x^m\,\overline{\chi}^2_{\alpha,k}(x) = \sigma^{2m}_{\alpha,k}\,Q\,\frac{C_{\alpha,k}}{C_{\alpha,k+2m}}\,\overline{\chi}^2_{\alpha,k+2m}(y). \tag{7.27}$$

where $Q := \sigma^2_{\alpha,k+2m}/\sigma^2_{\alpha,k}$ and $y = Q\,x$.

PROOF. From (7.13),

$$x^m\,\overline{\chi}^2_{\alpha,k}(x) = \sigma^{2m}_{\alpha,k}\,\frac{C_{\alpha,k}}{2\,\sigma^2_{\alpha,k}}\left(\frac{x}{\sigma^2_{\alpha,k}}\right)^{(k+2m)/2-3/2} F_{\frac{\alpha}{2}}\left(\left(\frac{x}{\sigma^2_{\alpha,k}}\right)^{\alpha/2}\right).$$

We see that $\overline{\chi}^2_{\alpha,k}$ should become $\overline{\chi}^2_{\alpha,k+2m}$ according to the power in the $x^{(k+2m)/2-3/2}$ term, but other parts of the formula need to be adjusted too.

Since

$$\overline{\chi}^2_{\alpha,k+2m}(y) = \frac{C_{\alpha,k+2m}}{2\,\sigma^2_{\alpha,k+2m}}\left(\frac{y}{\sigma^2_{\alpha,k+2m}}\right)^{(k+2m)/2-3/2} F_{\frac{\alpha}{2}}\left(\left(\frac{y}{\sigma^2_{\alpha,k+2m}}\right)^{\alpha/2}\right),$$

we obtain $y = x\,\sigma^2_{\alpha,k+2m}/\sigma^2_{\alpha,k}$ in order to match the two structurally.

Then take the ratio of $x^m\,\overline{\chi}^2_{\alpha,k}(x)/\overline{\chi}^2_{\alpha,k+2m}(y)$ to determine the needed constant, we arrive at (7.27).

□

7.9. Sum of Two Chi-Squares with Correlation

The sum of bivariate variables is studied here.

LEMMA 7.6. Let $Z = Z_1/s_1 + Z_2/s_2$ where Z_1, Z_2 are two independent χ^2_1 variables. The PDF of Z is

$$\begin{aligned}\chi^2_{11}(z, s_1, s_2) &= \frac{\sqrt{s_1 s_2}}{2} e^{-s_2 z/2}\,{}_1F_1\left(\frac{1}{2}, 1; \frac{(s_2-s_1)z}{2}\right)\\ &= \frac{\sqrt{s_1 s_2}}{2} e^{-(s_1+s_2)z/4} I_0(|s_2-s_1|z/4)\end{aligned}$$

We apply DLMF 12.6.9 to get the second line, where the symmetry of a, b is explicit since $I_0(x)$ is symmetric. For $x \gg 1$, $I_0(x) \approx e^x/\sqrt{2\pi x}$ (DLMF 10.40.5).

When $Z_1 = U_1^2$, $Z_2 = U_2^2$, and U_1, U_2 has correlation ρ, then s_1, s_2 must be modified by the eigenvalue solution of $\bar{\Omega}^{-1}\mathrm{diag}(\boldsymbol{s})$ such that

$$\begin{aligned}&\chi^2_{11}(z, s_1, s_2, \rho) = \chi^2_{11}(z, s_1', s_2')\\ &\qquad\text{where}\quad (s_1', s_2') = \frac{(s_1+s_2) \pm \sqrt{(s_1-s_2)^2 - 4\rho^2 s_1 s_2}}{2(1-\rho^2)}\end{aligned}$$

CHAPTER 8

Fractional F Distribution

The classic F distribution comes from the ratio of two χ^2 distributions. Assume $U_1 \sim \chi_d^2/d$ and $U_2 \sim \chi_k^2/k$, then $F \sim U_1/U_2$ is an F distribution, $F_{d,k}$.

Two use cases were mentioned in Azzalini (2013)[1]. In its Section 4.3, the squared variable of a univariate skew-t with k degrees of freedom is distributed as $F_{1,k}$.

In its Section 6.2, the quadratic from of a $d \times d$ multivariate skew-t with k degrees of freedom is distributed as $F_{d,k}$. Thus the meaning of d and k is quite clear in such context.

This chapter extends it fractionally.

8.1. Definition

Definition 8.1. Assume $U_1 \sim \chi_d^2/d$ and $U_2 \sim \overline{\chi}_{\alpha,k}$, then $F \sim U_1/U_2$ is a fractional F distribution. We use the notation $F \sim F_{\alpha,d,k}$.

The standard PDF of $F_{\alpha,d,k}$ is

$$F_{\alpha,d,k}(x) = \int_0^\infty s\, ds \left[d\, \chi_d^2(dxs)\right] \overline{\chi}^2_{\alpha,k}(s) \tag{8.1}$$

and note that the classic term in the integrand, $d\,\chi_d^2(d\,z)$, is equivalent to our $\bar{\chi}^2_{1,d}(z)$.

The reader should be aware of the subtlety that "ds" in "$s\,ds$" is the calculus notation, while d elsewhere is just a constant.

The standard CDF of $F_{\alpha,d,k}$ is

$$\Phi[F_{\alpha,d,k}](x) = \int_0^x F_{\alpha,d,k}(x)\, dx \tag{8.2}$$

$$= \int_0^\infty \left[\frac{1}{\Gamma\left(\frac{d}{2}\right)}\, \gamma\left(\frac{d}{2}, \frac{dxs}{2}\right)\right] \overline{\chi}^2_{\alpha,k}(s)\, ds \tag{8.3}$$

since the CDF of a χ_d^2 is the regularized lower incomplete gamma function of $\gamma\left(\frac{d}{2}, \frac{x}{2}\right)/\Gamma\left(\frac{d}{2}\right)$.

It can also be represented by a fractional Gauss hypergeometric function. See Section 5.2.5.

8.1.1. The Origin of Fractional F.

$F_{\alpha,d,k}$ is connected to the quadratic form of a d-dimensional multivariate GAS-SN distribution, $L_{\alpha,k}(0, \bar{\Omega}, \boldsymbol{\beta})$. Indeed, its three parameters, α, d, k, are designated such that the symbols convey the same meanings. However, $\bar{\Omega}$ and β doesn't affect the outcome of $F_{\alpha,d,k}$.

To elaborate from Section 14.6, assume Z is a $d \times d$ multivariate skew-normal (SN) distribution $SN(0, \bar{\Omega}, \beta)$, and $\overline{\chi}_{\alpha,k}$ is a standard FCM. Then $X = Z/\overline{\chi}_{\alpha,k}$ is an $L_{\alpha,k}(0, \bar{\Omega}, \boldsymbol{\beta})$.

The quadratic form of X is $Q = \frac{1}{d}\, X^\intercal \bar{\Omega}^{-1} X$. And $Q \sim F_{\alpha,d,k}$ is a fractional F distribution.

8.1.2. Fractional F Subsumes F.

Lemma 8.2. When $\alpha = 1$, it becomes a classic F. That is, $F_{1,d,k} = F_{d,k}$.

8.1.3. Fractional F Subsumes GSaS-Squared and GAS-SN-Squared. The following cases are for $d = 1$:

LEMMA 8.3. If $X_1 \sim L_{\alpha,k}$, then $X_1^2 \sim F_{\alpha,1,k}$.

LEMMA 8.4. If $X_2 \sim L_{\alpha,k}(\beta)$, then $X_2^2 \sim F_{\alpha,1,k}$, independent of β.

They will be discussed in Chapter 12.

8.2. PDF at Zero

The PDF of an F distribution is singular as $x \to 0$ when $d < 2$. We can see that from

$$\begin{aligned} F_{\alpha,d,k}(x) &\approx \frac{(d/2)^{d/2}}{\Gamma(d/2)} x^{d/2-1} \int_0^\infty s^{d/2}\, ds\, \overline{\chi}^2_{\alpha,k}(s) \\ &= \frac{(d/2)^{d/2}}{\Gamma(d/2)} \mathbb{E}(X^d|\overline{\chi}_{\alpha,k})\, x^{d/2-1} \end{aligned} \tag{8.4}$$

for very small x.

When $d = 1$, the peak is divergent as $F_{\alpha,1,k}(x) \approx \frac{1}{\sqrt{2\pi}}\mathbb{E}(X|\overline{\chi}_{\alpha,k})\sqrt{x}^{-1}$. But its CDF $\propto \sqrt{x}$.

When $d = 2$, this peak is finite. $F_{\alpha,2,k}(0) = \mathbb{E}(X^2|\overline{\chi}_{\alpha,k})$.

When $d > 2$, $F_{\alpha,d,k}(x)$ drops to zero at $x = 0$. This strange phenomenon seems to indicate that the bivariate system is the lowest dimension to have stable quadratic statistics. And a three dimension system is likely more stable. But we only analyze the bivariate case in this book.

8.3. Mellin Transform

From (7.23), and note that $\bar{\chi}^2_d = \bar{\chi}^2_{1,d}$, the Mellin transform of Fractional F's PDF is

$$F_{\alpha,d,k}(x) \overset{\mathcal{M}}{\longleftrightarrow} (\bar{\chi}^2_{1,d})^*(s)\, (\bar{\chi}^2_{\alpha,k})^*(2-s) \qquad (d>0, k>0) \tag{8.5}$$

$$= \left(\sqrt{2d}\,\sigma_{\alpha,k}\right)^{2-2s} \left[\frac{\Gamma((d-1)/2)}{\Gamma(d-1)} \frac{\Gamma((k-1)/2)}{\Gamma((k-1)/\alpha)}\right] \left[\frac{\Gamma(2p(s))}{\Gamma(p(s))} \frac{\Gamma(2q(s)/\alpha)}{\Gamma(q(s))}\right], \tag{8.6}$$

$$\text{where } \; p(s) := s + d/2 - 3/2, \; q(s) := 1/2 + k/2 - s.$$

The number of gamma functions can be reduced via the Legendre duplication formula (A.2).

8.4. Moments

Its n-th moment is

$$\begin{aligned} \mathbb{E}(X^n|F_{\alpha,d,k}) &= d^{-n}\, \mathbb{E}(X^n|\chi^2_d)\, \mathbb{E}(X^{-n}|\overline{\chi}^2_{\alpha,k}) \\ &= \left(\frac{2}{d}\right)^n (d/2)_n\, \mathbb{E}(X^{-n}|\overline{\chi}^2_{\alpha,k}). \end{aligned} \tag{8.7}$$

where $(d/2)_n$ is the Pochhammer symbol, $(a)_n := \Gamma(a+n)/\Gamma(a)$.

Its first moment is $\mathbb{E}(X^{-1}|\overline{\chi}^2_{\alpha,k}) = \mathbb{E}(X^{-2}|\overline{\chi}_{\alpha,k})$, independent of d. This is due to $\mathbb{E}(X|\chi^2_d) = d$.

Note that this first moment is also the second moment of an univariate GAS-SN in (12.9), or simply the variance of the corresponding GSaS.

Its second moment is $(1+2/d)\, \mathbb{E}(X^{-2}|\overline{\chi}^2_{\alpha,k})$. Hence, its variance is

$$\begin{aligned} \mathrm{var}\{F_{\alpha,d,k}\} &= (1+2/d)\, \mathbb{E}(X^{-2}|\overline{\chi}^2_{\alpha,k}) - \mathbb{E}(X^{-1}|\overline{\chi}^2_{\alpha,k})^2 \\ &= (1+2/d)\, \mathbb{E}(X^{-4}|\overline{\chi}_{\alpha,k}) - \mathbb{E}(X^{-2}|\overline{\chi}_{\alpha,k})^2. \end{aligned} \tag{8.8}$$

8.4.1. Stability Issue of the Second Moment. The moment formula appears to be straightforward. But the devil is in the detail.

The stability of moments symbolizes the challenge of stability in the α-stable distribution. Even the second moment has dramatic behaviors when k is smaller than 4.

First, we shall recognize that the first moment of F is actually the second moment of the underlying two-sided distribution, because the variable of F is squared. Having a finite and stable first moment in F is quite meaningful. But it is much harder to make sense of the variance when k is too small.

Notice that, when $d \to \infty$, the variance is independent of d,

$$\mathrm{var}\{F_{\alpha,\infty,k}\} = \mathbb{E}(X^{-2}|\overline{\chi}^2_{\alpha,k}) - \mathbb{E}(X^{-1}|\overline{\chi}^2_{\alpha,k})^2$$

This is the most relevant quantity, if exists, that other variances of finite d are relative to in an inverse d relation, such as

$$\mathrm{var}\{F_{\alpha,d,k}\} - \mathrm{var}\{F_{\alpha,\infty,k}\} = \frac{2}{d}\,\mathbb{E}(X^{-2}|\overline{\chi}^2_{\alpha,k}).$$

8.5. Sum of Two Fractional Chi-Square Mixtures with Correlation

This section addresses a complication that arises from the multivariate adaptive distribution.

TODO need to re-write this. but I may not have enough result to write it though. Alas...

Consider $X_1^2 \sim F_{\alpha_1,1,k_1}$ and $X_2^2 \sim F_{\alpha_2,1,k_2}$. Assume that there is a correlation between X_1 and X_2 as described in Section 7.9. The PDF of the quadratic form $Q = (X_1^2 + X_2^2)/2$ is a convolution that wraps around $Z \sim \chi^2_{11}(\rho)$ such that

$$\begin{aligned}
f_Q(x) &= 2\int_0^{2x} F_{\alpha_1,1,k_1}(w) \cdot F_{\alpha_2,1,k_2}(2x-w)\,dw \\
&= 2\int_0^{\infty} ds_1\,\bar{\chi}^2_{\alpha_1,k_1}(s_1) \int_0^{\infty} ds_2\,\bar{\chi}^2_{\alpha_2,k_2}(s_2)\,\chi^2_{11}(2x, s_1, s_2, \rho)
\end{aligned}$$

This is the PDF of the quadratic form of a standard 2-dimensional adaptive GAS-SN distribution.

TODO When ρ and β mingle together, there are additional complications.

8.6. Fractional Adaptive F Distribution

It should look like this: $\overrightarrow{F}_{\boldsymbol{\alpha},d,\boldsymbol{k}}$, but it is a bit strange, mixing vectors and numbers together...

TODO Ah, this is much harder than I thought !!!

Part 3

Two-Sided Univariate Distributions

CHAPTER 9

Framework of Continuous Gaussian Mixture

The construction of a symmetric two-sided distribution is in the form of a continuous Gaussian mixture. Both the ratio and product distribution methods are used.

In the case of the symmetric α-stable distribution (SaS)[4], the exponential power distribution comes from its characteristic function (CF)[21]. We would like to present a unified framework and familiarize the reader with the notations, which would be otherwise subtle and confusing.

Assume the PDF of a two-sided symmetric distribution is $L(x)$ where $x \in \mathbb{R}$. It has zero mean, $\mathbb{E}(X|L) = 0$. Assume the PDF of a one-sided distribution is $\chi(x)\,(x > 0)$ such that

$$L(x) := \int_0^\infty s\, ds\, \mathcal{N}(xs)\, \chi(s) \tag{9.1}$$

This is nothing new. It is the definition of a ratio distribution with a standard normal variable $\mathcal{N}$. This is the first form of the Gaussian mixture: $L \sim \mathcal{N}/\chi$. A contrive example is that L is a Student's t distribution when χ is $(\chi_k^2)^{1/2}$.

The skewness is added by replacing the normal distribution $\mathcal{N}$ with its skew-normal counterpart $\mathcal{N}(\beta)$. See next chapter for more detail.

It has the equivalent expression in terms of a product distribution by way of *the inverse distribution* $\chi^\dagger$ such that $L \sim \mathcal{N}\chi^\dagger$. This is the second form of the Gaussian mixture.

$\chi^\dagger$ is closer to our typical understanding of the marginal distribution of a volatility process. For example, when the Brownian motion process $dX_t = \sigma_t\, dW_t$ is measured in a particular time interval Δt, we have $\Delta X_t \sim L$ and $\sigma_t \sim \chi^\dagger$.

However, χ in the first form is more natural in the expression of the α-stable distribution. So we are more inclined to use the ratio distribution. The reader should keep this subtlety in mind.

Lemma 9.1. (Inverse distribution) The inverse distribution is defined as[8]

$$\chi^\dagger(s) := s^{-2}\, \chi\left(\frac{1}{s}\right) \tag{9.2}$$

such that

$$\int_0^\infty s\, ds\, \mathcal{N}(xs)\, \chi(s) = \int_0^\infty \frac{ds}{s}\, \mathcal{N}\left(\frac{x}{s}\right)\, \chi^\dagger(s) \qquad (x \in \mathbb{R}) \tag{9.3}$$

$$\int_0^\infty s\, ds\, \mathcal{N}(xs)\, \chi^\dagger(s) = \int_0^\infty \frac{ds}{s}\, \mathcal{N}\left(\frac{x}{s}\right)\, \chi(s) \tag{9.4}$$

The proof is straightforward by a change of variable $t = 1/s$. You can move between LHS and RHS easily.

We use the notation $\mathrm{CF}\{g\}(t) = \mathbb{E}(e^{itX}|g)$ to represent the characteristic function transform of the PDF $g(x)$. Note that $\mathcal{N}$ has a special property that its CF is still itself: $\mathrm{CF}\{\mathcal{N}\}(t) = \sqrt{2\pi}\,\mathcal{N}(t)$.

LEMMA 9.2. (Characteristic function transform of L) Let $\phi(t)$ be the CF of L such that $\phi(t) := \mathrm{CF}\{L\}(t) = \int_{-\infty}^{\infty} dx\, \exp(itx)\, L(x)$. (9.1) is transformed to

$$\phi(t) = \sqrt{2\pi} \int_0^\infty ds\, \mathcal{N}\left(\frac{t}{s}\right) \chi(s) \qquad (t \in \mathbb{R}) \tag{9.5}$$

This allows us to define a new distribution pair: L_ϕ and $\chi_\phi^\dagger$, in terms of a product distribution such that

$$L_\phi(x) := \int_0^\infty \frac{ds}{s}\, \mathcal{N}\left(\frac{x}{s}\right) \chi_\phi^\dagger(s) \qquad (x \in \mathbb{R}) \tag{9.6}$$

$$\chi_\phi^\dagger(s) := \frac{s\,\chi(s)}{\mathbb{E}(X|\chi)} \tag{9.7}$$

where $\mathbb{E}(X|\chi)$ is the first moment of χ. Here $\chi_\phi^\dagger$ is the inverse distribution of χ_ϕ, which can be reverse-engineered according to (9.2),

$$\chi_\phi(s) := \frac{s^{-3}}{\mathbb{E}(X|\chi)}\, \chi\left(\frac{1}{s}\right) \tag{9.8}$$

We are in an interesting place: We start with a one-sided distribution χ, we derive two variants from it: χ_ϕ and $\chi_\phi^\dagger$. We also obtain two two-sided distributions: L and L_ϕ.

We shall call χ_ϕ *the characteristic distribution* of χ since it facilitates the following parallel relation:

$$L \sim \mathcal{N}/\chi$$
$$L_\phi \sim \mathcal{N}/\chi_\phi$$

χ symbolizes the fractional χ distribution we are about to present. The ϕ suffix will be replaced with the *negation* (sign change) of the degree of freedom.

CHAPTER 10

SN: The Skew-Normal Distribution - Review

10.1. Definition

The skew-normal distribution family is well documented in A. Azzalini's 2013 monograph[1]. My contribution is to incorporate the skew methodology to my fractional distributions wherever suitable. This enhances the flexibility of the distributions that can adopt to many different shapes and tails with high skewness and kurtosis.

10.1.1. The Selective Sampling. The *selective sampling* method is used to inject skewness to the stochastic system, that is otherwise symmetric. This mechanism is fairly common in an applied context, e.g. in social sciences, where a variable X_0 is observed only when a correlated variable X_1, which is usually unobserved, fulfills a certain condition (p.128 of [1]).

In quantitative finance, the condition could be market regimes. In a two-regime model, a market index such as the S&P 500 index (SPX) is classified into the growth regime or the crash regime. It is well known that the volatility of the index behaves differently in each regime. In the growth regime, the volatility tends to be low and the market is calm and trending upwards. In the crash regime, the volatility tends to be high, and the market is trending downwards violently.

A univariate random variable $Z \sim SN(0, 1, \beta)$ is a standard skew-normal variable with skew parameter $\beta \in \mathbb{R}$ (Section 2.1 of [1]). The sign of β determines the sign of its skewness (10.18).

One of its stochastic representations is

$$Z = \begin{cases} X_0 & \text{if } X_1 < \beta X_0, \\ -X_0 & \text{otherwise.} \end{cases} \tag{10.1}$$

where X_0, X_1 are independent $N(0, 1)$ variables.

An alternative representation uses filtering, or rejection, such that $Z = (X_0|X_1 < \beta X_0)$.

10.1.2. The PDF and CDF. The standard PDF is

$$\mathcal{N}(x; \beta) := 2\,\mathcal{N}(x)\Phi_{\mathcal{N}}(\beta x), \quad (x \in \mathbb{R}) \tag{10.2}$$

where $\mathcal{N}(x)$ and $\Phi_{\mathcal{N}}(x)$ are the PDF and CDF of $N(0, 1)$.

Its extremal distribution occurs at $\beta \to \infty$, where $\Phi_{\mathcal{N}}(\beta x)$ becomes a step function. The PDF becomes that of a half-normal distribution.

The standard CDF is

$$\Phi_{SN}(x; \beta) := \Phi_{\mathcal{N}}(x) - 2\,T(x, \beta) \tag{10.3}$$

where $T(h, a)$ is called the Owen's T function[22]. Its numerical methods are widely implemented.

Several important properties are quoted from Proposition 2.1 of [1]:

- $\mathcal{N}(0; 0) = 1/\sqrt{2\pi}$. Universal anchor at $x = 0, \beta = 0$.
- $\mathcal{N}(x; 0) = \mathcal{N}(x)$. Continuity at $\beta = 0$.
- $\mathcal{N}(-x; \beta) = \mathcal{N}(x; -\beta)$. This is the reflection rule.
- $Z^2 \sim \chi^2_1$, irrespective of β.

Notice that Z^2 is independent of β. This is an important property. It is due to the fact that, in (10.1), the square of X_0 and $-X_0$ are the same. This property is inherited by the quadratic form of the multivariate elliptical distribution.

10.2. The Location-Scale Family

Its location-scale family is $Y = \xi + \omega Z \sim SN(\xi, \omega^2, \beta)$, where $\xi \in \mathbb{R}$ and $\omega > 0$. Its PDF becomes

$$\frac{1}{\omega} \mathcal{N}\left(\frac{x-\xi}{\omega}; \beta\right). \tag{10.4}$$

10.3. Invariant Quantities

The following quantity plays an important role in the selective sampling concept of SN:

$$\delta = \frac{\beta}{\sqrt{1+\beta^2}}, \quad \delta \in (-1, 1). \tag{10.5}$$

Inversely, given the correlation δ, β can be calculated from

$$\beta = \frac{\delta}{\sqrt{1-\delta^2}}. \tag{10.6}$$

These two quantities will show up in many places in the ensuing chapters. They are invariants in the context of the multivariate elliptical distribution, called the Canonical Form.

One can think of δ as $\sin(\theta)$ of a right triangle, where one leg is 1, the other leg is β, and θ is the angle facing β.

Three representations use δ as the correlation coefficient to generate SN. (Section 2.1.3 of [1]) The correlation matrix is

$$\bar{\Omega} = \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix}.$$

The Cholesky factor of $\bar{\Omega}$ is

$$L = \begin{pmatrix} 1 & 0 \\ \delta & \sqrt{1-\delta^2} \end{pmatrix},$$

so that $L\,L^T = \bar{\Omega}$.

Assume U_0 and U_1 are two independent $N(0,1)$ variates. The first representations of $Z \sim SN(0,1,\beta)$ is

$$Z = \begin{cases} X_0 & \text{if } X_1 > 0, \\ -X_0 & \text{otherwise.} \end{cases} \tag{10.7}$$

where X_0, X_1 are marginals of a standard correlated normal bivariate with $\mathrm{cor}\{X_0, X_1\} = \delta$ such that

$$\begin{pmatrix} X_0 \\ X_1 \end{pmatrix} = L \begin{pmatrix} U_0 \\ U_1 \end{pmatrix}.$$

The second representation is from

$$\begin{pmatrix} - \\ Z \end{pmatrix} = L \begin{pmatrix} U_0 \\ |U_1| \end{pmatrix}$$

such that $Z = \sqrt{1-\rho^2}\, U_0 + \delta\, |U_1| \sim SN(0,1,\beta)$.

The third representation is $Z = \max\{X_0, X_1\} \sim SN(0,1,\beta)$, where X_0, X_1 are marginals of a standard correlated bivariate via $\mathrm{cor}\{X_0, X_1\} = \rho = 1 - 2\delta^2$ such that

$$\begin{pmatrix} X_0 \\ X_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} U_0 \\ U_1 \end{pmatrix}.$$

10.4. Mellin Transform

The following result is elegant, but also peculiar. It is discovered by the author.

LEMMA 10.1. The Mellin transform of the SN PDF is

$$\mathcal{N}(x;\beta) \xleftrightarrow{\mathcal{M}} \mathcal{N}^*(s;\beta) := 2\,\mathcal{N}^*(s)\,\Phi[t_s](\beta\sqrt{s}), \tag{10.8}$$

$$\text{where} \quad \mathcal{N}^*(s) = \frac{1}{2}\frac{2^{s/2}}{\sqrt{2\pi}}\Gamma\left(\frac{s}{2}\right)$$

is the Mellin transform of the PDF of $N(0,1)$ in (2.9). And $\Phi[t_k](x)$ is the CDF of a Student's t distribution with k degrees of freedom. But it is used in a strange way, where s substitutes k and goes into x at the same time.

PROOF. We prove (10.8) via the CDF of GSaS when its $\alpha = 1$. By definition,

$$\mathcal{N}^*(s;\beta) = \int_0^\infty x^{s-1}\,\left[2\mathcal{N}(x)\Phi_{\mathcal{N}}(\beta x)\right] dx.$$

We use the known result from $\bar{\chi}_{1,k}$ where

$$x^{k-1}\mathcal{N}(x) = \frac{2^{k/2-1}\Gamma(k/2)}{\sqrt{2\pi k}}\,\bar{\chi}_{1,k}(x/\sqrt{k}) = \frac{1}{\sqrt{k}}\,\mathcal{N}^*(k)\,\bar{\chi}_{1,k}(x/\sqrt{k}).$$

Then

$$\begin{aligned} \mathcal{N}^*(s;\beta) &= \frac{2\,\mathcal{N}^*(s)}{\sqrt{s}} \int_0^\infty \Phi_{\mathcal{N}}(\beta x)\,\bar{\chi}_{1,s}(x/\sqrt{s})\,dx \\ &= 2\,\mathcal{N}^*(s) \int_0^\infty \Phi_{\mathcal{N}}(\beta\sqrt{s}t)\,\bar{\chi}_{1,s}(t)\,dt \quad \text{via} \quad t = x/\sqrt{s}. \end{aligned}$$

The integral is exactly the CDF of a GSaS, $L_{1,s}$, with the argument $\beta\sqrt{s}$. That is, $\mathcal{N}^*(s;\beta) = 2\,\mathcal{N}^*(s)\,\Phi[L_{1,s}](\beta\sqrt{s})$.

When $\alpha = 1$, that $L_{1,s}$ becomes t_s. Therefore, $\mathcal{N}^*(s;\beta) = 2\,\mathcal{N}^*(s)\,\Phi[t_s](\beta\sqrt{s})$.

□

The beauty of this lemma is that $\mathcal{N}^*(s;\beta)$ is the multiplication of a symmetric component and a skew component, just like its PDF counterpart.

From (2.12), we also obtain that

$$\Phi_{SN}(0;\beta) = 1 - \mathcal{N}^*(1;\beta) = \frac{1}{2} - \frac{1}{\pi}\arctan(\beta), \tag{10.9}$$

stated in Proposition 2.7 of [1]. This is due to $\mathcal{N}^*(1) = \frac{1}{2}$ and $\Phi[t_1](\beta) = \frac{1}{2} + \frac{1}{\pi}\arctan(\beta)$.

10.4.1. Mellin Transform of Owen's T Function.

Define the upper incomplete integral as

$$\Gamma_f(x) := \int_x^\infty \mathcal{N}(x;\beta)\,dx = 1 - \Phi_{SN}(x;\beta) \tag{10.10}$$

$$= 1 - \Phi_{\mathcal{N}}(x) + 2\,T(x,\beta) \tag{10.11}$$

Its Mellin transform is

$$\Gamma_f(x) \overset{\mathcal{M}}{\longleftrightarrow} s^{-1}\,\mathcal{N}^*(s+1;\beta) \tag{10.12}$$

$$= 2s^{-1}\,\mathcal{N}^*(s+1)\,\Phi[t_{s+1}](\beta\sqrt{s+1}) \tag{10.13}$$

Therefore,

$$T(x,\beta) = \frac{\Gamma_f(x) - (1-\Phi_{\mathcal{N}}(x))}{2} \overset{\mathcal{M}}{\longleftrightarrow} s^{-1}\,\mathcal{N}^*(s+1)\,\left[\Phi[t_{s+1}](\beta\sqrt{s+1}) - \frac{1}{2}\right] \tag{10.14}$$

where $1-\Phi_{\mathcal{N}}(x) \overset{\mathcal{M}}{\longleftrightarrow} s^{-1}\,\mathcal{N}^*(s+1)$.

TODO prove this in python

10.5. Moments

By assigning $s = n+1$, the Mellin transform is converted to the moment formula. It is easy to show that the n-th moment of Z is

$$\begin{aligned} \mathbb{E}(Z^n) = \mathbb{E}(X^n|\mathcal{N}(\beta)) &= \mathcal{N}^*(n+1;\beta) + (-1)^n\mathcal{N}^*(n+1;-\beta) \qquad (10.15)\\ &= 2\mathcal{N}^*(n+1) \times \begin{cases} 1, & \text{when } n \text{ is even,} \\ 1-2\Phi[t_{n+1}](-\beta\sqrt{n+1}), & \text{when } n \text{ is odd.} \end{cases} \end{aligned}$$

The even moments are identical to those of $N(0,1)$. It is the odd moments that make the difference when $\beta \neq 0$.

Z's first four moments have simple analytic forms. Its first moment is

$$\mu_z = b\,\delta, \quad \text{where } b = \sqrt{2/\pi}. \tag{10.16}$$

The second moment is simply 1. Its variance is

$$\sigma_z^2 = 1 - (b\,\delta)^2. \tag{10.17}$$

The third moment is $b\,\delta(3-\delta^2)$. Its skewness is

$$\gamma_1\{Z\} = \frac{4-\pi}{2}\,\frac{\mu_z^3}{\sigma_z^3}. \tag{10.18}$$

The fourth moment is 3. Its kurtosis is

$$\gamma_2\{Z\} = 2(\pi-3)\,\frac{\mu_z^4}{\sigma_z^4}. \tag{10.19}$$

The maximum skewness of SN is about 0.9953, and maximum kurtosis is 0.8692. They are not very interesting, since the extremal distribution is just a half-normal distribution.

However, these analytic forms are useful when SN is extended to GAS-SN. Both skewness and kurtosis are extended to much wider ranges, or even infinity!

CHAPTER 11

GAS: Generalized Alpha-Stable Distribution (Experimental)

In this chapter, we show how the *degrees of freedom* k is added to the α-stable distribution L_α^θ through the Mellin transform approach. It is my belief that this is one of the cleanest approaches to understand it.

A new distribution comes out of it, which is called the generalized α-stable distribution (GAS), with the notation of $L_{\alpha,k}^\theta$. The distribution is structurally elegant and capable of generating skewness properly. However, there are discontinuity issues resulted from the reflection rule.

A method to remediate it is proposed, which is documented in this chapter. However, I do not consider it ready to be used for real-world applications.

The value of this chapter is to understand the origin of the fractional χ distribution and GSaS.

The skew-normal approach is much better in terms of generating skewness. There is no problem in the continuity of the PDF. And it is also theoretically elegant. All the subsequent chapters are extended from the skew-normal approach.

11.1. Definition

In Section 4.3, it is shown that the Mellin transform in (4.4)

$$L_\alpha^\theta(x) \overset{\mathcal{M}}{\longleftrightarrow} \epsilon \left[\frac{\Gamma(s)}{\Gamma(1-\gamma+\gamma s)}\right] \left[\frac{\Gamma(\epsilon(1-s))}{\Gamma(\gamma(1-s))}\right].$$

is interpreted in Lemma 4.2 as

$$L_\alpha^\theta(x) \overset{\mathcal{M}}{\longleftrightarrow} \tilde{M}_\gamma^*(s)\, {\overline{\chi}_{\alpha,1}^\theta}^*(2-s),$$

where $\overline{\chi}_{\alpha,1}$ is defined as

$$\overline{\chi}_{\alpha,1}^\theta(x) \overset{\mathcal{M}}{\longleftrightarrow} {\overline{\chi}_{\alpha,1}^\theta}^*(s) \;\propto \frac{\Gamma(\epsilon(s-1))}{\Gamma(\gamma(s-1))},$$

apart from the normalization constant and scale in the PDF.

In Section 7.1, it is shown that the "degrees of freedom" parameter k is inserted by replacing $s-1$ with $s+k-2$ to form the FCM, such that

$$\overline{\chi}_{\alpha,k}^\theta(x) \overset{\mathcal{M}}{\longleftrightarrow} {\overline{\chi}_{\alpha,k}^\theta}^*(s) \;\propto \frac{\Gamma(\epsilon(s+k-1))}{\Gamma(\gamma(s+k-1))}.$$

Then it is natural to combine them and propose -

Definition 11.1 (The ratio-distribution representation of (unadjusted) GAS). The Mellin transform of the PDF of (unadjusted) GAS is defined as

$$\tilde{L}_{\alpha,k}^\theta(x) \overset{\mathcal{M}}{\longleftrightarrow} \tilde{M}_\gamma^*(s)\, {\overline{\chi}_{\alpha,k}^\theta}^*(2-s) \tag{11.1}$$

Based on the Mellin transform, its PDF can be written in a ratio distribution form of

$$\tilde{L}_{\alpha,k}^\theta(x) := \int_0^\infty \tilde{M}_\gamma(xs)\, \overline{\chi}_{\alpha,k}^\theta(s)\, s\, ds \quad (x \geq 0) \tag{11.2}$$

Since the Mellin integral is only valid for $x \geq 0$, it is supplemented with *the reflection rule*:

$$\tilde{L}^{\theta}_{\alpha}(-x) := \tilde{L}^{-\theta}_{\alpha}(x) \tag{11.3}$$

Thus, we've constructed a version of GAS for $x \in \mathbb{R}$, which produces fat tails and skewness -

(1) $\tilde{L}^{\theta}_{\alpha,k}$ subsumes the α-stable distribution L^{θ}_{α}.
(2) $\tilde{L}^{\theta}_{\alpha,k}$ subsumes Student's t distribution t_k.
(3) $\tilde{L}^{\theta}_{\alpha,k}$ subsumes the power-exponential distribution, with proper definition of negative k in FCM.

So what's wrong with it? The problem is that the PDF and its derivatives are discontinuous at $x = \pm 0$ when $k \neq 1$.

The remaining sections of this chapter will explain this problem and provide a remediation. The reader who just wants to explore the skew-normal implementation can safely skip the rest of this chapter. The conclusion is that such discontinuity makes the PDF far from mathematical elegance, which motivates the author to explore other alternatives. The answer is to abandon the M-Wright kernel for skewness ($\tilde{M}_{\gamma}(xs)$ in (11.2)), and integrate with the skew-normal distribution, outlined in the next chapter.

11.2. Limitation

The issue of discontinuity of the PDF $\tilde{L}^{\theta}_{\alpha,k}(x)$ at $x = 0$ is encountered when $k \neq 1$. We lay out a generic framework to understand and address it.

Assume that the unadjusted two-sided density function is $\tilde{f}(x) := \tilde{L}^{\theta}_{\alpha,k}(x)$, which is discontinuous at $x = 0$. It also must satisfy the reflection rule, where, for $x > 0$, $\tilde{f}(x) := \tilde{f}^{+}(x)$ and $\tilde{f}(-x) := \tilde{f}^{-}(-x)$.

$\tilde{f}(x)$ can be expanded at $x = 0$ in terms of x by

$$\tilde{f}^{\pm}(x) := \tilde{L}^{\pm\theta}_{\alpha,k}(x) = \tilde{f}^{\pm}_{0} + \tilde{f}^{\pm}_{1}\, x + ... \tag{11.4}$$

where $\tilde{f}^{\pm}_{0}$ are the densities at $x = 0$, and $\tilde{f}^{\pm}_{1}$ are the respective slopes (aka the first derivatives).

The series expansion can be achieved via either (11.2), or (11.1) in conjuction with Ramanujan's master theorem in Section 2.2, such that

$$\tilde{f}^{+}_{0} = \frac{\gamma^{1-\gamma}}{\Gamma(1-\gamma)}\, E(X|\bar{\chi}^{\theta}_{\alpha,k}), \tag{11.5}$$

$$\tilde{f}^{+}_{1} = \frac{-\gamma^{1-2\gamma}}{\Gamma(1-2\gamma)}\, E(X^2|\bar{\chi}^{\theta}_{\alpha,k}). \tag{11.6}$$

Notice that they are based on the first and second moments of $\bar{\chi}^{\theta}_{\alpha,k}$. $(\tilde{f}^{-}_{0}, \tilde{f}^{-}_{1})$ are obtained by applying the reflection rule from $(\tilde{f}^{+}_{0}, \tilde{f}^{+}_{1})$. That is, θ is replaced with $-\theta$, and γ with $1-\gamma$ in every occurrence of the formula.

Furthermore, it is known that

$$\int_0^\infty \tilde{f}^{+}(x)\, dx = \gamma, \qquad \int_0^\infty \tilde{f}^{-}(x)\, dx = 1-\gamma. \tag{11.7}$$

These two are the only conditions required for $\tilde{f}^{\pm}(x)$.

The discontinuity occurs because $\tilde{f}^{+}_{0} \neq \tilde{f}^{-}_{0}$ and $\tilde{f}^{+}_{1} \neq \tilde{f}^{-}_{1}$ when $k \neq 1$ and $\theta \neq 0$. In fact, this is true for all orders of derivatives $\tilde{f}^{+}_{n} \neq \tilde{f}^{-}_{n}$ in the n-th term, $\tilde{f}^{\pm}_{n}\, x^n$.

Obviously, when $\theta = 0$, the density function is symmetric by definition: $\tilde{f}^{+}(x) = \tilde{f}^{-}(x)$. There is no issue here. So the issue is specific to the injection of skewness from $\theta \neq 0$.

On the other hand, when $k = 1$, the density function is continuous under the reflection rule, regardless the value of θ. This is the original α-stable distribution. It is perfectly fine. So the issue is specific to our attempt of adding degrees of freedom $k \neq 1$.

Either one of θ or k are fine, but when we try to do both, the distribution is broken, so to speak. That is the limitation. The dilemma is that adding θ and k is exactly what we try to achieve.

11.3. Workaround

An adjustment algorithm is proposed such that the PDF and its first derivative are continuous.

Definition 11.2 (The adjusted GAS). The PDF of the adjusted GAS is defined as

$$L^{\pm\theta}_{\alpha,k}(x) := \frac{1}{A^{\pm}\sigma^{\pm}}\, \tilde{f}^{\pm}(x) \left(\frac{x}{\sigma^{\pm}}\right) \quad (x \geq 0) \tag{11.8}$$

It is required that (a) the new density function satisfies the reflection rule of $L^{\theta}_{\alpha,k}(-x) := L^{-\theta}_{\alpha,k}(x)$; (b) $A^{\pm}, \sigma^{\pm}$ are constrained by the continuity conditions that, at $x = 0$, both its density is continuous: $L^{\theta}_{\alpha,k}(0) = L^{-\theta}_{\alpha,k}(0)$; and its slope is continuous: $\frac{d}{dx}\, L^{\theta}_{\alpha,k}(0) = -\frac{d}{dx}\, L^{-\theta}_{\alpha,k}(0)$.

With such definition, we proceed to find the solutions of $A^{\pm}, \sigma^{\pm}$. The solutions form a distribution family. There is a canonical solution, simple and elegant, from which all other solutions are derived as a member of the location-scale family.

A member in the location-scale family shares the same "shapes" such as the skewness and kurtosis. Apart from the location and scale, it brings nothing new to the table. Hence, we can focus on analyzing the canonical distribution.

Definition 11.3 (Two essential quantities for the canonical distribution). We define two essential quantities:

$$\Sigma := -\frac{\tilde{f}^{+}_0}{\tilde{f}^{-}_0}\,\frac{\tilde{f}^{-}_1}{\tilde{f}^{+}_1} \tag{11.9}$$

$$\Psi := \Sigma\,\frac{\tilde{f}^{+}_0}{\tilde{f}^{-}_0} \quad = -\left(\frac{\tilde{f}^{+}_0}{\tilde{f}^{-}_0}\right)^2 \frac{\tilde{f}^{-}_1}{\tilde{f}^{+}_1} \tag{11.10}$$

Notice that $\tilde{f}^{+}_0/\tilde{f}^{-}_0$ is the ratio of the original densities from two sides of $x = 0$. And $\tilde{f}^{-}_1/\tilde{f}^{+}_1$ is the ratio of the slopes of the two sides. Since $\tilde{f}^{-}_1, \tilde{f}^{+}_1$ always have the opposite signs, Σ is a positive quantity.

Note that Σ is singular when $\gamma = 1/2$. Both $\tilde{f}^{-}_1, \tilde{f}^{+}_1$ approach zero at the same speed. Hence, $\Sigma \to 1$ and $\Psi \to 1$.

The most important contribution is the discovery of the canonical distribution.

Definition 11.4 (The canonical GAS). The canonical GAS distribution is defined according to $\sigma^{+} = 1$ and $\sigma^{-} = \Sigma$. Hence, its PDF for $x \geq 0$ is (with the hat symbol)

$$\widehat{L}^{\theta}_{\alpha,k}(x) := \frac{1}{A^{+}}\, \tilde{f}^{+}(x) \tag{11.11}$$

$$\widehat{L}^{-\theta}_{\alpha,k}(x) := \frac{1}{A^{-}\Sigma}\, \tilde{f}^{-}\left(\frac{x}{\Sigma}\right) \tag{11.12}$$

where $A^{+} = \gamma + \Psi(1-\gamma)$ and $A^{-} = A^{+}/\Psi$ from Lemma 11.7.

The reflection rule applies: $\widehat{L}^{\theta}_{\alpha,k}(-x) := \widehat{L}^{-\theta}_{\alpha,k}(x)$.

11.3.1. The Location-scale Family. The following lemmas show that all other solutions must obey $\sigma^-/\sigma^+ = \Sigma$. They are just the location-scale family of the canonical distribution.

Briefly, all other solutions are defined by a choice of scale $\sigma^+ > 0$, such that

$$L^{\theta}_{\alpha,k}(x) := \frac{1}{\sigma^+}\,\widehat{L}^{\theta}_{\alpha,k}\left(\frac{x}{\sigma^+}\right) \tag{11.13}$$

For instance, we found that $\sigma^+ = \Sigma^\gamma$ to be a very good alternative. In the remark of Definition 11.9, we show that the n-th moment of $L^{\theta}_{\alpha,k}$ is just that of $\widehat{L}_{\alpha,k}$ multiplied by its scale $(\sigma^+)^n$.

Lemma 11.5. The requirement that the density and slope of the *adjusted* density function should be smooth at $x = 0$ leads to

$$\frac{1}{A^+\sigma^+}\,\tilde{f}^+_0 = \frac{1}{A^-\sigma^-}\,\tilde{f}^-_0 \tag{11.14}$$

$$\frac{1}{A^+(\sigma^+)^2}\,\tilde{f}^+_1 = -\frac{1}{A^-(\sigma^-)^2}\,\tilde{f}^-_1 \tag{11.15}$$

Proof. To solve $A^\pm$ and $\sigma^\pm$, take (11.8) and carry out the series expansions from (11.4):

$$L^{\pm\theta}_{\alpha,k}(x) = \frac{\tilde{f}^\pm_0}{A^\pm\sigma^\pm} + \frac{\tilde{f}^\pm_1}{A^\pm(\sigma^\pm)^2}\,x + ... \tag{11.16}$$

(11.14) is straightforward from requiring $L^{\theta}_{\alpha,k}(0) = L^{-\theta}_{\alpha,k}(0)$ in (11.16). Likewise, (11.15) is the result of $\frac{d}{dx}L^{\theta}_{\alpha,k}(0) = \frac{d}{dx}L^{-\theta}_{\alpha,k}(0)$ from (11.16). □

Lemma 11.6. The equations in Lemma 11.5 lead to the following invariant:

$$\frac{\sigma^-}{\sigma^+} = \Sigma \tag{11.17}$$

Proof. Divide the LHS and RHS of (11.14) by those of (11.15) respectively,

$$\sigma^+\frac{\tilde{f}^+_0}{\tilde{f}^+_1} = -\sigma^-\frac{\tilde{f}^-_0}{\tilde{f}^-_1}$$

Rearrange the items and we obtain (11.17). □

Lemma 11.7. The solution for $A^\pm$ are

$$A^+ = \gamma + \Psi(1-\gamma) \tag{11.18}$$

$$A^+/A^- = \Psi \tag{11.19}$$

Proof. (11.19) is derived by rearranging the items in (11.14) and following the definition of Ψ.

(11.18) is derived from the fact that the total density of the adjusted distribution should be equal to 1, that is, $\int_{-\infty}^{\infty} f(x)dx = 1$. Hence,

$$\int_0^\infty f^+(x)dx + \int_0^\infty f^-(x)dx = \frac{1}{A^+}\int_0^\infty \tilde{f}^+(x)dx + \frac{1}{A^-}\int_0^\infty \tilde{f}^-(x)dx = 1$$

Apply (11.7), we get $\frac{\gamma}{A^+} + \frac{1-\gamma}{A^-} = 1$. Multiply it by A^+ on both sides, we obtain (11.18). □

We've shown that $A^\pm$ are well-defined constants based on (α, k, θ), while $\sigma^\pm$ is a choice of parametrization, constrained by (11.17).

11.4. Moments

The structure of the *moments* reveals critical information about the adjusted distribution. We show the moment formula of the canonical distribution, and how the location-scale family relates to it.

To simplify the notations below, let

- $f^{\pm} = L^{\pm\theta}_{\alpha,k}$ be the adjusted distribution family,
- $\widehat{f}^{\pm} = \widehat{L}^{\pm\theta}_{\alpha,k}$ be the canonical distribution,
- $\tilde{f}^{\pm} = \tilde{L}^{\pm\theta}_{\alpha,k}$ be the original (unadjusted) distribution.

First, the n-th one-sided moments of the adjusted distribution are $(x > 0)$

$$E(X^n|f^{\pm}) = \frac{1}{A^{\pm}\sigma^{\pm}} \int_0^{\infty} x^n\, \tilde{f}^{\pm}(x/\sigma^{\pm}))dx = \frac{(\sigma^{\pm})^n}{A^{\pm}}\, E(X^n|\tilde{f}^{\pm}) \tag{11.20}$$

where $E(X^n|\tilde{f}^{\pm})$ are the original n-th one-sided moments. They can be obtained from the Mellin transform (11.1).

The n-th total moment, given the notation of m_n, is the sum of $E(X^n|f^+)$ and $(-1)^n E(X^n|f^-)$. We show the following.

Lemma 11.8. The n-th total moment of the adjusted distribution is based on the original one-sided moments such as

$$m_n := E(X^n|f) = \frac{(\sigma^+)^n}{A^+}\, \left[E(X^n|\tilde{f}^+) + \Psi(-\Sigma)^n\, E(X^n|\tilde{f}^-)\right] \tag{11.21}$$

Proof. By definition, we have

$$\begin{aligned} m_n := E(X^n|f) &= \int_{-\infty}^{\infty} x^n\, f(x)dx = \int_0^{\infty} x^n\, f^+(x)dx + (-1)^n \int_0^{\infty} x^n\, f^-(x)dx \\ &= E(X^n|f^+) + (-1)^n E(X^n|f^-) \end{aligned}$$

Apply (11.20), we get

$$m_n = \frac{(\sigma^+)^n}{A^+}\, E(X^n|\tilde{f}^+) + \frac{(-\sigma^-)^n}{A^-}\, E(X^n|\tilde{f}^-)$$

Factor out $\frac{(\sigma^+)^n}{A^+}$, apply $\sigma^-/\sigma^+ = \Sigma$ from Lemma 11.6, and $A^+/A^- = \Psi$ from 11.7, we obtain (11.21). □

Lemma 11.9 (The moments of the canonical distribution). The n-th moment of the canonical distribution is

$$\widehat{m}_n := E(X^n|\widehat{f}) = \frac{1}{A^+}\, \left[E(X^n|\tilde{f}^+) + \Psi(-\Sigma)^n\, E(X^n|\tilde{f}^-)\right] \tag{11.22}$$

Proof. Lemma 11.8 shows that the canonical distribution $\widehat{f}$ is obtained by letting $\sigma^+ = 1$ and $\sigma^- = \Sigma$. Put them to (11.21), we obtain (11.22). □

Lastly, compare (11.21) with (11.22). We reach $m_n = (\sigma^+)^n\, \widehat{m}_n$. That is, all other members in the adjusted distribution family are rescaled canonical distributions.

CHAPTER 12

GAS-SN: Generalized Alpha-Stable Distribution with Skew-Normal

This fractional univariate distribution combines the features from a classic skew-normal distribution that provides skewness and a fractional distribution that provides fatter tails. The resulting distribution is analytically tractable. The PDF and all of its derivatives are continuous everywhere in $\mathbb{R}$.

12.1. Definition

DEFINITION 12.1. Assume $Z_0 \sim SN(0, 1, \beta)$ is a skew-normal variable and $V \sim \overline{\chi}_{\alpha,k}$ is an FCM variable.

Then $Z \sim Z_0/V$ is a variable with a GAS-SN distribution. We use the notation $Z \sim L_{\alpha,k}(\beta)$ for this standard distribution.

Assume $\mathcal{N}(x)$ and $\Phi_{\mathcal{N}}(x)$ are the PDF and CDF of $N(0, 1)$. The PDF of Z is

$$L_{\alpha,k}(x; \beta) = 2 \int_0^\infty \mathcal{N}(xs)\, \Phi_{\mathcal{N}}(\beta x s)\, \overline{\chi}_{\alpha,k}(s)\, s\, ds. \tag{12.1}$$

This is the fractional extension of (10.2).

Its CDF is

$$\begin{aligned} \Phi[L_{\alpha,k}(\beta)](x) &:= \int_0^\infty \Phi_{SN}(xs; \beta)\, \overline{\chi}_{\alpha,k}(s)\, ds. \\ &= \int_0^\infty \left[\Phi_{\mathcal{N}}(xs) - 2\, T(xs, \beta)\right] \overline{\chi}_{\alpha,k}(s)\, ds. \end{aligned} \tag{12.2}$$

where $\Phi_{SN}(xs; \beta)$ is the CDF of $SN(0, 1, \beta)$ in (10.3), and $T(h, a)$ is the Owen's T function.

We can clearly see that the CDF has two components: One from the symmetric part, and the other skew. The second component vanishes due to $T(h, 0) = 0$.

12.1.1. GAS-SN Subsumes GSaS.

LEMMA 12.2. When $\beta = 0$, it becomes a symmetric distribution, previously called GSaS. The notation of $L_{\alpha,k}$ is given in [**13**].

The PDF of a GSaS is

$$L_{\alpha,k}(x) = \int_0^\infty \mathcal{N}(xs)\, \overline{\chi}_{\alpha,k}(s)\, s\, ds. \tag{12.3}$$

When $\alpha \to 2$ or $k \to \infty$, the symmetric distribution approaches a normal distribution $N(0, \alpha^{2/\alpha})$ (Section 8.2 of [**13**]).

This integral is a normal mixture (9.1) that enjoys several nice properties outlined in Chapter 9. In particular, the generalized exponential power distribution can be obtained via the characteristic function transform in Lemma 9.2 (Section 9 of [**13**]). We point out that the skew extension is straightforward, but leave the detailed description to future research.

12.1.2. GAS-SN Subsumes Skew-t Distribution. An important bridge between SN and GAS-SN is the skew-t (ST) distribution. It is documented in Section 4.3 of [**1**].

ST is fully consistent with GAS-SN by setting $\alpha = 1$. That is, in his notation, $T(\beta, k) = L_{1,k}(\beta)$.

12.2. The Location-Scale Family

Its location scale family is $Y = \xi + \omega Z \sim L_{\alpha,k}(\xi, \omega^2, \beta)$. Its PDF becomes

$$\phi(x) = \frac{1}{\omega} L_{\alpha,k}\left(\frac{x-\xi}{\omega}; \beta\right). \quad (x \in \mathbb{R}) \tag{12.4}$$

In real-world application, this PDF is used for optimization, e.g. in the maximum likelihood estimation (MLE). See Section 12.7.

12.3. Mellin Transform

The Mellin transform of the PDF follows the rule of the ratio distribution. From (10.8) and (7.2), we have

$$\begin{aligned} L_{\alpha,k}(\beta)(x) &\overset{\mathcal{M}}{\longleftrightarrow} L_{\alpha,k}(\beta)^*(s) && (12.5)\\ &= \mathcal{N}^*(s;\beta)\, \overline{\chi}^*_{\alpha,k}(2-s) \\ &= \left[2\,\Phi[t_s](\beta\sqrt{s})\right] \times \left[\mathcal{N}^*(s)\, \overline{\chi}^*_{\alpha,k}(2-s)\right] && (12.6) \end{aligned}$$

Notice that the contribution for the skewness is $2\,\Phi[t_s](\beta\sqrt{s})$ in the first bracket, which becomes one if $\beta = 0$.

The second bracket is the Mellin transform of GSaS PDF. From (2.9) and (7.2), it is

$$\begin{aligned} L_{\alpha,k}(x) \overset{\mathcal{M}}{\longleftrightarrow} L^*_{\alpha,k}(s) &= \mathcal{N}^*(s)\, \overline{\chi}^*_{\alpha,k}(2-s) \\ &= \frac{1}{2\sqrt{\pi}} \left(\frac{2}{\sigma}\right)^{s-1} \Gamma\left(\frac{s}{2}\right) \frac{\Gamma((k-1)/2)}{\Gamma((k-1)/\alpha)} \frac{\Gamma((k-s)/\alpha)}{\Gamma((k-s)/2)}, \end{aligned} \tag{12.7}$$

where $\sigma := k^{1/2-1/\alpha}$ and $k > 0$ is assumed.

12.4. Moments

Based on $\mathbb{E}(X^n|\mathcal{N}(\beta))$ from (10.15), the n-th moment of Z is

$$\begin{aligned} \mathbb{E}(X^n|L_{\alpha,k}(\beta)) &:= \mathbb{E}(X^n|\mathcal{N}(\beta))\, \mathbb{E}(X^{-n}|\overline{\chi}_{\alpha,k}) \\ &= 2\,\mathcal{N}^*(n+1)\, \mathbb{E}(X^{-n}|\overline{\chi}_{\alpha,k}) \\ &\quad \times \begin{cases} 1, & \text{when } n \text{ is even,} \\ 1 - 2\Phi[t_{n+1}](-\beta\sqrt{n+1}), & \text{when } n \text{ is odd.} \end{cases} \end{aligned} \tag{12.8}$$

Its first moment is $\mu_z = b\,\delta$, where $b = \sqrt{2/\pi}\, \mathbb{E}(X^{-1}|\overline{\chi}_{\alpha,k})$.

The second moment is $\mathbb{E}(X^{-2}|\overline{\chi}_{\alpha,k})$. Its variance is

$$\sigma_z^2 = \mathbb{E}(X^{-2}|\overline{\chi}_{\alpha,k}) - (b\,\delta)^2. \tag{12.9}$$

To simplify the symbology, let $q_n := \mathbb{E}(X^{-n}|\overline{\chi}_{\alpha,k})$. The third moment is $\delta_3\, q_3$, where $\delta_3 = \sqrt{\frac{2}{\pi}}\delta(3-\delta^2)$. The fourth moment is $3\,q_4$. To carry out the skewness γ_1 and excess kurtosis γ_2,

$$\begin{aligned} \gamma_1 \times \sigma_z^{3/2} &= \delta_3\, q_3 - 3\mu_z q_2 + 2\mu_z^3, \\ \gamma_2 \times \sigma_z^4 &= 3(q_4 - q_2^2) - 4\mu_z(\gamma_1 \times \sigma_z^{3/2}) + 2\mu_z^4. \end{aligned}$$

The maximum skewness and kurtosis can be infinite. Since $\delta = \sin\theta$, where $\beta = \tan\theta$, we have $\delta \in [-1, 1]$. The infinity has to come from q_3 and q_4.

A typical example is the skew-t distribution at $\alpha = 1$. It is well known that the kurtosis approaches infinity when k approaches 4 from above, and the skewness approaches infinity when k approaches 3 from above.

12.5. Tail Behavior

The tail behavior of GAS-SN is a modified GSaS type. Hence, it is well within what was known. Without losing generality, assume $\beta > 0$, the left tail's decay is more pronounced than the right tail. But it still follows the same power law of x^{-k} as in a $L_{\alpha,k}$.

It takes a small tweak to GSaS to capture that behavior.

DEFINITION 12.3. The shifted GSaS is defined as

$$L_{\alpha,k}(x,\mu) = \int_0^\infty \mathcal{N}(xs-\mu)\,\overline{\chi}_{\alpha,k}(s)\,s\,ds \tag{12.10}$$

Note that the shift μ is not a location parameter that shifts x. It is a shift inside the argument of $\mathcal{N}(\,)$. When $\mu = 0$, it is restored to the PDF of GSaS, $L_{\alpha,k}(x)$.

We use the following approximation of the erf function in (12.1)[9]

$$1 - \mathrm{erf}(x) \approx \frac{1}{B\sqrt{\pi}\,x}(1-e^{-Ax})\,e^{-x^2} \quad (x \geq 0) \tag{12.11}$$

where $A = 1.98$ and $B = 1.135$. It is much better than the first term expansion of $e^{-x^2}/(\sqrt{\pi}x)$ for the entire range of $x \in [0,\infty)$.

LEMMA 12.4. The left tail ($x < 0$) of the PDF in (12.1) can be approximated by

$$\widehat{L}_{\alpha,k}(x;\beta) = \frac{G}{\beta x}\left[e^{\mu^2/2}\,L_{\alpha,k-1}\left(qx,\mu\right) - L_{\alpha,k-1}\left(qx\right)\right] \tag{12.12}$$

where

$$\mu = \frac{A\,\delta}{\sqrt{2}}$$
$$q = \sqrt{1+\beta^2}\,\frac{\sigma_{\alpha,k}}{\sigma_{\alpha,k-1}}$$
$$G = \sqrt{\frac{2}{\pi}}\frac{B\,C_{\alpha,k}}{\sigma_{\alpha,k-1}\,C_{\alpha,k-1}}$$

and both $C_{\alpha,k} = \frac{\alpha\Gamma((k-1)/2)}{\Gamma((k-1)/\alpha)}$ and $\sigma_{\alpha,k}$ are according to FCM in (7.4).

The right tail ($x > 0$) is simply

$$L_{\alpha,k}(x) - \widehat{L}_{\alpha,k}(-x;\beta) \tag{12.13}$$

where the second term $\widehat{L}_{\alpha,k}(-x;\beta)$ becomes much smaller than the first term as $x \to \infty$.

PROOF. TODO add more content here.

□

12.6. Quadratic Form

A squared GAS-SN variable Q is distributed as a fractional F distribution with $d = 1$. That is,

$$Q := \left(\frac{Y-\xi}{\omega}\right)^2 = Z^2 \sim F_{\alpha,1,k}, \quad \text{for all } \beta. \tag{12.14}$$

Notice that Q is based on the standard variable Z, which is invariant to the location and scale. See Chapter 8 for more detail.

12.7. Univariate MLE

We document how we fit the one-dimensional data with GAS-SN. The main algorithm is MLE, supplemented with several small components of regularization.

In the univariate case, the hyperparameter space is $\Theta = \{\alpha, k, \beta, \omega, \xi\}$. Assume there are N samples in the data set, $Y = \{y_i, i \in 1, 2, ..., N\}$, the minus log-likelihood (MLLK) is

$$\textsc{mllk}(\Theta) = -\frac{1}{N} \sum_{i=1}^{N} \log(\phi(y_i; \Theta)) \tag{12.15}$$

Additional components of regularization are added to the objective function. Specifically, the L2 distances between the empirical and theoretical statistics are added for the following:

- Skewness: $|\Delta\gamma_1|^2 := |\Delta\text{skewness}(Y)|^2$. Section 12.4.
- Kurtosis: $|\Delta\gamma_2|^2 := |\Delta\text{kurtosis}(Y)|^2$. Section 12.4.
- The mean of the quadratic form: $\Delta\mu_Q^2 := |\Delta\text{mean}(Q)|^2$. Section 12.6.

MLE seeks the optimal Θ that minimizes the objective function:

$$\hat{\Theta} = \text{argmin}\ \ell(\Theta) \tag{12.16}$$

$$\text{where}\ \ \ell(\Theta) = \textsc{mllk}(\Theta) + |\Delta\gamma_1|^2 + |\Delta\gamma_2|^2 + \Delta\mu_Q^2 \tag{12.17}$$

A custom version of stochastic descent (SD) algorithm is developed. Our experience shows that it is better to standardize the data set to one standard deviation, so that all the parameters in Θ are approximately on the same scale.

It is also important to control the learning rate such that it doesn't make too large of a step on α, empirically no more than 0.01 per step. This ensures the SD not walking into the "undefined" regions for $\ell(\Theta)$. This is particularly important for the SPX fit below.

12.8. Examples of Univariate MLE Fits

12.8.1. VIX fit. Figure 12.1 is the result of the MLE fit to VIX daily returns from 1990 to 2025. Data is standardized to one standard deviation. This helps the stochastic descent algorithm to move correctly in all dimensions.

VIX data is right skewed with a positive β. The sample skewness of 2.0 is quite high. There is a very stretched right tail due to several high-profile one-day panic selling events. This tail creates a very high kurtosis of 17.

The PP-plot shows that the fit overall is satisfactory. The 45-degree line is very clear. α is slightly below 0.8 and k is in the neighborhood of 5.

The QQ-plot of the quadratic form is a powerful tool to examine how the tails are doing. The 45-degree line is okay below 20, but as the quantiles get larger, the observed quantiles start to tilt upward. This means the top 0.5 percent of the tail is not properly captured by the distribution.

12.8.2. SPX fit. Figure 12.2 is the result of the MLE fit to SPX daily returns from 1990 to 2025. Data is standardized to one standard deviation too.

SPX data is left skewed with a negative β. The sample skewness of 0.2 is mild. There is a stretched left tail due to several high-profile one-day panic selling events. This tail creates a very high kurtosis of 17.

The PP-plot shows that the fit overall is satisfactory. The 45-degree line is okay. But there is a small bump between 0 and 0.2. α is around 0.9 and k is in the neighborhood of 3. This region is close to t_3, which is quite peculiar, since the theoretical skewness and kurtosis barely exist.

In the QQ-plot of the quadratic form, the 45-degree line is okay below 100, but as the quantiles get larger, the observed quantiles start to tilt downward. This means the top 0.5 percent of the tail is not properly captured by the distribution.

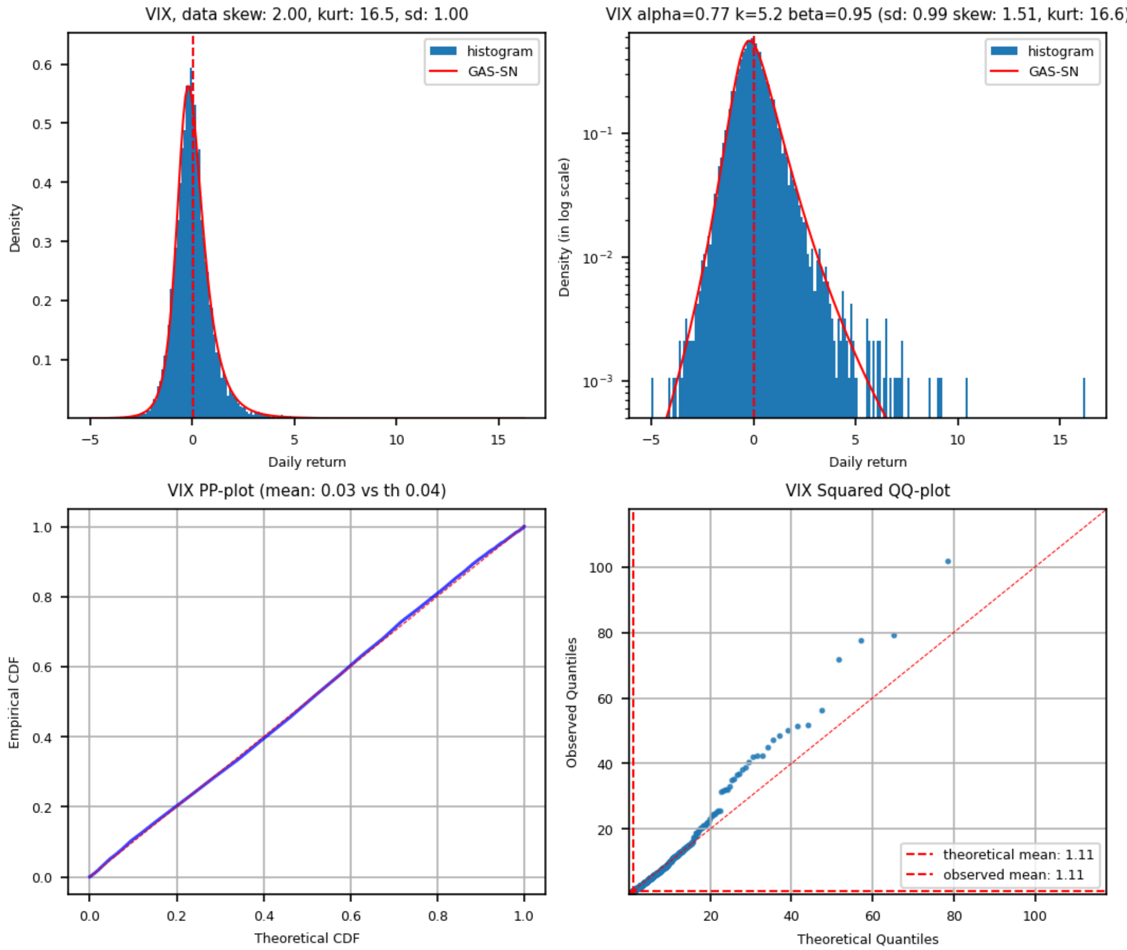

FIGURE 12.1. MLE fit of VIX daily returns from 1990 to 2025. Data is standardized to one standard deviation.

Notice how far the quantiles have stretched. The theoretical mean is 2.8, while the largest point is near 700. It spans almost 3 orders of magnitude.

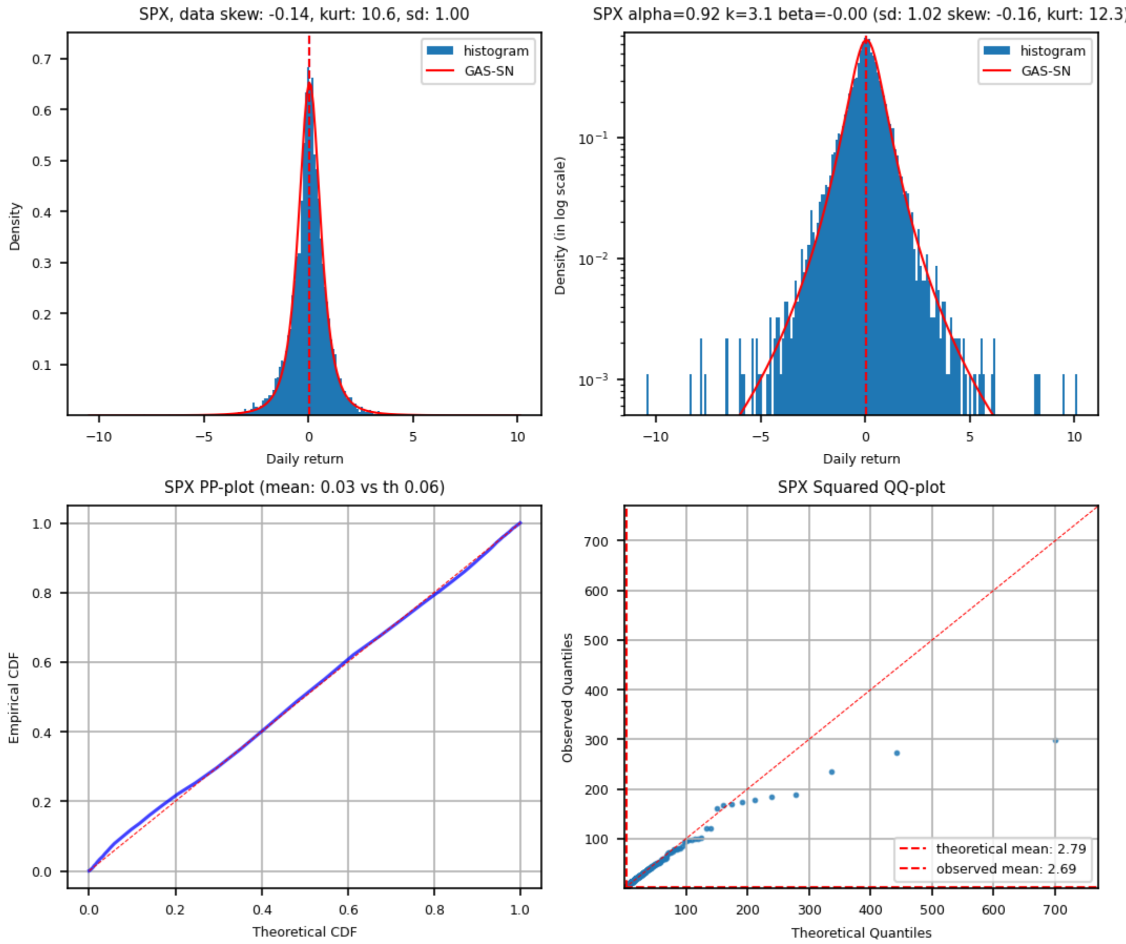

FIGURE 12.2. MLE fit of SPX daily log returns from 1990. Data is standardized to one standard deviation.

Part 4

Multivariate Distributions

CHAPTER 13

Multivariate SN Distribution (Review)

In this chapter, we begin to construct the multivariate distributions. Since the data from the real world is multi-dimensional, a flexible framework of multivariate distribution can be highly useful. That is what we aim to achieve with the next few chapters.

The foundation is the multivariate normal distribution $\mathcal{N}_d(0, \bar{\Omega})$[**26**]. The skew-normal distribution $SN_d(0, \bar{\Omega}, \boldsymbol{\beta})$ adds skewness to it, as described in Chapter 5 of Azzalini[**1**]. In its Chapter 6, the skew-elliptical distributions are discussed. And the multivariate skew-t distribution $ST_d(0, \bar{\Omega}, \boldsymbol{\beta}, k)$ is constructed by combining a multivariate normal distribution with $\chi_k/\sqrt{k}$ in a ratio distribution.

The following work builds on top of the skew-elliptical distribution concept. The denominator $\chi_k/\sqrt{k}$ is expanded to $\overline{\chi}_{\alpha,k}$. The fractional dimension α is added to the shape parameters. This forms a super distribution family called *multivariate GAS-SN distribution* with the notation of $L_{\alpha,k}(0, \bar{\Omega}, \boldsymbol{\beta})$.

Although the skew-elliptical distribution has beautiful properties inherited from the elliptical distribution framework, it deficiency becomes obvious in real-world applications: The structure is multivariate, but the shape parameters $\{\alpha, k\}$ are not.

We propose a more flexible framework called *multivariate adaptive distribution*, in which the shape parameters $\{\boldsymbol{\alpha}, \boldsymbol{k}\}$ are d-dimensional vectors, just like their $\boldsymbol{\beta}$ counterpart. The flexibility in shapes comes with an expensive computational cost. It is analogous to the classic "curse of dimensionality" problem. Therefore, we are forced to study only the bivariate case.

Lastly, we wrap up the book with a real world example. The data is the daily returns of SPX index and VIX index since 1990. To fit this data properly, all the tools mentioned in this book are used, albeit just in the bivariate sense. Two methods of maximum likelihood estimation (MLE) with regularization are put to work.

We are able to fit such bivariate data with high accuracy. The statistics of the two marginal distributions are matched. Their PP-plots are straight lines.

The quadratic forms $Z^T\bar{\Omega}^{-1}Z$ of each multivariate distribution result in the fractional extension of their classic counterparts. The multivariate QQ-plots based on the quadratic forms and these extended fractional distributions are powerful validation to the goodness of the fit. We also manage to build a satisfactory framework as an expansion from Azzalini.

13.1. Definition

We summarize the results from Chapter 6 of Azzalini[**1**]. The basic building block is the standard multivariate normal distribution $\mathcal{N}_d(0, \bar{\Omega})$, where $\bar{\Omega}$ is a $d \times d$ correlation matrix[**26**]. Its PDF is defined as

$$\mathcal{N}(\boldsymbol{x}; \bar{\Omega}) = \frac{1}{(2\pi)^{d/2} \det(\bar{\Omega})^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{x}^\intercal\bar{\Omega}^{-1}\boldsymbol{x}\right), \quad \boldsymbol{x} \in \mathbb{R}^d. \tag{13.1}$$

A standard multivariate skew-normal distribution is $Z \sim SN_d(0,\bar{\Omega}, \boldsymbol{\beta})$, where $\boldsymbol{\beta} \in \mathbb{R}^d$ is the skew parameter (or the slant parameter). Its PDF is

$$\mathcal{N}(\boldsymbol{x}; \bar{\Omega}, \boldsymbol{\beta}) = \mathcal{N}(\boldsymbol{x}; \bar{\Omega})\, \Phi_{\mathcal{N}}(\boldsymbol{\beta}^\intercal\boldsymbol{x}), \tag{13.2}$$

where $\Phi_{\mathcal{N}}(x)$ is the CDF of a standard normal distribution.

13.2. The Location-Scale Family

Its location-scale family is $Y = \xi + \omega Z \sim SN_d(\xi, \Omega, \boldsymbol{\beta})$, where $\xi \in \mathbb{R}^d$ is the location and $\omega = \mathrm{diag}(\omega_1, ..., \omega_d) > 0$ is the diagonal scale matrix. Its PDF becomes

$$\det(\omega)^{-1} \mathcal{N}(\boldsymbol{z}; \bar{\Omega}, \boldsymbol{\beta}). \tag{13.3}$$

where $\boldsymbol{z} = \omega^{-1}(\boldsymbol{x} - \xi)$ and $\Omega = \omega\bar{\Omega}\omega$.

13.3. Quadratic Form

The quadratic form of a multivariate SN distribution is defined as

$$Q := \frac{1}{d}(Y - \xi)^\intercal \Omega^{-1}(Y - \xi) = \frac{1}{d} Z^\intercal \bar{\Omega}^{-1} Z \tag{13.4}$$

Since $Z \sim SN_d(0,\bar{\Omega}, \boldsymbol{\beta})$, $Q \sim \chi^2_d/d = \bar{\chi}^2_{1,d}$ for all β. Notice that our definition of Q is slightly different from that of Azzalini. We prefer to have the distribution of Q tied to the fractional F distribution directly without any constant adjustment. This will make things a lot simpler in Section 14.6.

The distribution of Q is independent of $\boldsymbol{\beta}$. This is an important property due to the rotational invariance of the elliptical distribution.

To prove, we quote Corollary 5.9 of [1] below for a skew-normal distribution with 0 location:

Lemma 13.1. If $Y \sim SN_d(0, \Omega, \boldsymbol{\beta})$ and A is a $d \times d$ symmetric matrix, then

$$Y^\intercal A Y = X^\intercal A X$$

where $X \sim \mathcal{N}_d(0, \Omega)$.

This lemma allows β to be dropped out of the statistics of Q.

13.4. Stochastic Representation

Assume $X_0 \sim \mathcal{N}_d(0, \bar{\Omega})$ and $T \sim N(0, 1)$, then the first representation of $Z \sim SN_d(0,\bar{\Omega}, \boldsymbol{\beta})$ is

$$Z = \begin{cases} X_0 & \text{if } T > \boldsymbol{\beta}^\intercal X_0, \\ -X_0 & \text{otherwise.} \end{cases} \tag{13.5}$$

This form of selective sampling is quite useful in generating random numbers for Z. It is essentially an extension from (10.1).

However, this scheme can be rephrased in a more interesting form. First, define the multivariate version of δ as

$$\boldsymbol{\delta} = (1 + \boldsymbol{\beta}^\intercal \bar{\Omega} \boldsymbol{\beta})^{-1/2}\, \bar{\Omega}\boldsymbol{\beta}, \quad (\boldsymbol{\delta} \in \mathbb{R}^d) \tag{13.6}$$

which is used to construct a $(d + 1) \times (d + 1)$ correlation matrix

$$\Omega^* = \begin{pmatrix} \bar{\Omega} & \boldsymbol{\delta} \\ \boldsymbol{\delta}^\intercal & 1 \end{pmatrix}.$$

It is used to generate the marginals, X_0 and X_1, such that

$$\begin{pmatrix} X_0 \\ X_1 \end{pmatrix} \sim \mathcal{N}_{d+1}(0, \Omega^*).$$

which leads to the second representation

$$Z = \begin{cases} X_0 & \text{if } X_1 > 0, \\ -X_0 & \text{otherwise.} \end{cases} \tag{13.7}$$

This form resembles (10.7). It reveals the role of $\boldsymbol{\delta}$ as a hidden variable that adds correlation to $\bar{\Omega}$ and participates in the selective sampling .

13.5. Moments

Z's first two moments have simple analytic forms. Its first moment is

$$\mu_z = \mathbb{E}(Z) = b\,\boldsymbol{\delta}, \qquad \text{where } b = \sqrt{2/\pi}. \tag{13.8}$$

The second moment is simply $\bar{\Omega}$. Its variance is

$$\Sigma_z = \text{var}\{Z\} = \bar{\Omega} - b^2\,\boldsymbol{\delta}\boldsymbol{\delta}^{\intercal}. \tag{13.9}$$

And $\mathbb{E}\{Y\,Y^{\intercal}\} = \Omega$.

Define the important invariant quantity

$$\beta_* = (\boldsymbol{\beta}^{\intercal}\bar{\Omega}\boldsymbol{\beta})^{1/2} \geq 0, \tag{13.10}$$

from which the quadratic form $\mu_z^{\intercal}\Sigma_z^{-1}\mu_z$ is written as

$$\mu_z^{\intercal}\Sigma_z^{-1}\mu_z = \frac{b^2\beta_*^2}{1+(1-b^2)\beta_*^2}. \tag{13.11}$$

The non-negative scalar quantity β_* encapsulates the departure from normality. A related quantity is

$$\delta_* = (\boldsymbol{\delta}^{\intercal}\bar{\Omega}^{-1}\boldsymbol{\delta})^{1/2} \in [0,1), \tag{13.12}$$

The two are connected by

$$\delta_* = \frac{\beta_*^2}{1+\beta_*^2}, \quad \beta_* = \frac{\delta_*^2}{1-\delta_*^2}.$$

13.6. Canonical Form

The concept of canonical form in SN is very important and fascinating. From Proposition 5.12 of [1], there exists an affine transformation $Z^* = A_*(Y-\xi)$ such that $Z* \sim SN_d(0, \boldsymbol{I}_d, \boldsymbol{\beta}_{Z*})$, where $\boldsymbol{I}_d$ is the $d \times d$ identity matrix, $\boldsymbol{\beta}_{Z*} = (\beta_*, 0, ..., 0)^{\intercal}$ and β_* is defined by (13.10).

The variable Z^*, which is called *canonical variable*, comprises d independent components. Only one of them contains the skew component. All others are standard normal distributions. That is, the PDF of Z^* is

$$\mathcal{N}_*(\boldsymbol{x};\beta_*) = 2\prod_{i=1}^{d}\mathcal{N}(x_i)\Phi_{\mathcal{N}}(\beta_* x_1).$$

This structure helps tremendously for our subsequent development of the elliptical distribution and adaptive distribution.

Proposition 5.13 of [1] describes how to find such A_*. However, due to the rotational symmetry, there are many choices of A_*. This is not a problem, as long as we always look at the system in the quadratic form.

Let $C = \Omega^{1/2}$ be the unique positive definite symmetric square root of Ω. Define $M = C^{-1}\Sigma C^{-1}$, where $\Sigma = \text{var}\{Y\}$. Let $Q\Lambda Q^{\intercal}$ denote a spectral decomposition of M, where we assume that the diagonal eigenvalue matrix Λ are arranged in increasing order.

Let $H = C^{-1}Q$. Then H is the matrix operator to convert Y to Z^*,

$$Z^* = H^\intercal (Y - \xi).$$

Since $\boldsymbol{\delta}_{Z*} = H^\intercal \omega \boldsymbol{\delta}$ and $\boldsymbol{\beta}_{Z*} = \boldsymbol{\delta}_{Z*}/(1 - \delta_*^2)$, the choice of H must make the first element of $\boldsymbol{\delta}_{Z*}$ a non-negative number, that is, $\delta_* \geq 0$. All other elements except the first ones in $\boldsymbol{\delta}_{Z*}$ and $\boldsymbol{\beta}_{Z*}$ must be zero.

The significance of this theorem is that the skew-elliptical distributions derived from the SN framework can only have a single source of skewness. It might be hidden and not easy to observe in the real-world data. But there is only one number. Everything else is from the multivariate normal distribution.

If we want a more "sophisticated" distribution, we have to go beyond the skew-normal elliptical distribution.

13.7. 1D Marginal Distribution

We are particularly interested in the 1D marginal distribution, since we are going to fit a bivariate data set. We can add the likelihood of the two 1D marginal distributions to the objective function, so that each dimension is addressed properly.

In fact, for the adaptive distribution, the full 2D likelihood is so compute-intensive that it is too slow to perform MLE on a desktop. The alternative is to compute the sum of the likelihoods of each 1D marginal distribution plus regularization on the statistical quantities, such as the correlation coefficient.

Assume the marginal is on the first dimension. The correlation matrix is decomposed as

$$\bar{\Omega} = \begin{pmatrix} \bar{\Omega}_{11} & \bar{\Omega}_{12} \\ \bar{\Omega}_{21} & \bar{\Omega}_{22} \end{pmatrix}.$$

The formula can be simplified due to $\bar{\Omega}_{11} = 1$ in the 1D case.

The marginal distribution is $Y_1 \sim SN(\xi_1, \Omega_{11}, \beta_{1(2)})$. Its $\beta_{1(2)}$ is derived as

$$\begin{aligned} \beta_{1(2)} &= \left(1 + \beta_2^\intercal \bar{\Omega}_{22.1} \beta_2\right)^{-1/2} \left(\beta_1 + \bar{\Omega}_{12}\beta_2\right) \\ \text{where} \quad \bar{\Omega}_{22.1} &= \bar{\Omega}_{22} - \bar{\Omega}_{21}\bar{\Omega}_{12}. \end{aligned} \tag{13.13}$$

In the bivariate case, it is simple. Assume we want to get the marginal β of the i-th dimension, $\beta_{i(j)}$, where j is the other dimension. Then

$$\beta_{i(j)} = \frac{\beta_i + \rho\beta_j}{\sqrt{1 + \beta_j^2 \, |\bar{\Omega}|}} \tag{13.14}$$

where $|\bar{\Omega}| = 1 - \rho^2$.

ρ in the numerator describes how β_j is mixed into the marginal. When $\rho = 0$, there is no mixing from the other dimension, only a reduction in total scale through $\beta_{i(j)} = \beta_i / \sqrt{1 + \beta_j^2}$.

CHAPTER 14

Multivariate GAS-SN Elliptical Distribution

14.1. Definition

This chapter follows the structure laid out in Chapter 6 of Azzalini (2013)[1]. We implemented the skew-elliptical distribution by our $\overline{\chi}_{\alpha,k}$, which fully extends his multivariate skew-t distribution.

Definition 14.1. Assume $Z_0 \sim SN_d(0,\bar{\Omega},\boldsymbol{\beta})$ is a $d \times d$ standard multivariate skew-normal (SN) distribution, and $V \sim \overline{\chi}_{\alpha,k}$ is a standard FCM. $\bar{\Omega}$ is a correlation matrix.

Then $Z \sim Z_0/V$ is a $d \times d$ standard multivariate GAS-SN elliptical distribution. It is given the notation of $Z \sim L_{\alpha,k}(0,\bar{\Omega},\boldsymbol{\beta})$.

Equivalently, using the location-scale notation, $Z \sim SN_d(0,\Sigma,\boldsymbol{\beta})$ where $\Sigma = \bar{\Omega}/V^2$.

Assume $\mathcal{N}(x;\bar{\Omega})$ is the PDF of a standard multivariate normal distribution $N(0,\bar{\Omega})$ [26]. $\Phi_{\mathcal{N}}(x)$ is the CDF of a standard normal distribution.

We expand on the construction of multivariate SN distribution in (13.1) and (13.2). And the PDF of $Z \sim L_{\alpha,k}(0,\bar{\Omega},\boldsymbol{\beta})$ is

$$\begin{aligned} L_{\alpha,k}(\boldsymbol{x};\bar{\Omega},\boldsymbol{\beta}) &= \int_0^\infty ds\, \overline{\chi}_{\alpha,k}(s)\, s^d\, \mathcal{N}(\boldsymbol{x}s;\bar{\Omega},\boldsymbol{\beta}) \\ &= 2\int_0^\infty ds\, \overline{\chi}_{\alpha,k}(s)\, s^d\, \mathcal{N}(\boldsymbol{x}s;\,\bar{\Omega})\,\Phi_{\mathcal{N}}(\boldsymbol{\beta}^\intercal \boldsymbol{x}\, s). \end{aligned} \tag{14.1}$$

The s^d term comes from $\det(s\boldsymbol{I}_d)$ where $\boldsymbol{I}_d$ is the $d \times d$ identity matrix. It is easy to see how it is reduced to a univariate GAS-SN distribution when $d = 1$.

14.1.1. Multivariate Skew-t Distribution.

An important bridge between multivariate SN and GAS-SN is the multivariate skew-t distribution. It is documented in Section 6.2 of [1].

It is fully consistent with multivariate GAS-SN by setting $\alpha = 1$. That is, in his notation of skew-t: $ST_d(\Omega,\boldsymbol{\beta},k) \sim L_{1,k}(\Omega,\boldsymbol{\beta})$.

14.2. Location-Scale Family

The location-scale family follows the standard procedure: $Y = \xi + \omega Z$, which is denoted as $Y \sim L_{\alpha,k}(\xi,\Omega,\boldsymbol{\beta})$, where $\Omega := \omega^\intercal \bar{\Omega}\omega$ is the covariance matrix, and ω is a $d \times d$ diagonal scale matrix.

The PDF of Y is

$$L_{\alpha,k}(\boldsymbol{x};\,\xi,\Omega,\boldsymbol{\beta}) := \det(\omega)^{-1} L_{\alpha,k}(\boldsymbol{z};\bar{\Omega},\boldsymbol{\beta}) \tag{14.2}$$

where $\boldsymbol{z} := \omega^{-1}(\boldsymbol{x} - \xi)$. Notice that it has to be computed via the standard PDF.

14.3. Moments

The first moment of Z is $\mu_z := b\,\boldsymbol{\delta}$, where $b := \sqrt{2/\pi}\,\mathbb{E}(X^{-1}|\overline{\chi}_{\alpha,k})$.

The second moment of Z is $m_2\,\bar{\Omega}$, where $m_2 = \mathbb{E}(X^{-2}|\overline{\chi}_{\alpha,k})$. Hence, the covariance is

$$\mathrm{var}\{Z\} := \Sigma_z := m_2\,\bar{\Omega} - b^2\,\boldsymbol{\delta}\boldsymbol{\delta}^\intercal$$

The moments of Y follow the rule of the location-scale family. The first moment of Y is $\xi + \omega\, \mu_z$. The covariance of Y is $\omega\, \Sigma_z\, \omega$.

14.4. Canonical Form

The concept of canonical form in GAS-SN is extended from the multivariate SN in Section 13.6. There exists an affine transformation $Z^* = A_*(Y - \xi)$ such that $Z^* \sim L_{\alpha,k}(0, \boldsymbol{I}_d, \boldsymbol{\beta}_{Z*})$, where $\boldsymbol{\beta}_{Z*} = (\beta_*, 0, ..., 0)^\intercal$ and β_* is defined by (13.10). And the algorithm of finding A_* is exactly the same as in Section 13.6.

The variable Z^*, which is called *canonical variable*, comprises d independent components. Only one of them contains the skew component. All others are standard GSaS distributions. That is, the PDF of Z^* is

$$L_{\alpha,k_*}(\boldsymbol{x}; \beta_*) = 2 \int_0^\infty ds\, \overline{\chi}_{\alpha,k}(s)\, s^d \prod_{i=1}^{d} \mathcal{N}(x_i) \Phi_{\mathcal{N}}(\beta_* x_1). \tag{14.3}$$

It can be further simplified to an elegant univariate-style integral. When $|\boldsymbol{x}| \neq 0$, let $\beta_*(\boldsymbol{x}) := \beta_* x_1 / |\boldsymbol{x}| \in \mathbb{R}$, and

$$L_{\alpha,k_*}(\boldsymbol{x}; \beta_*) = (2\pi)^{-(d-1)/2} \int_0^\infty ds\, \overline{\chi}_{\alpha,k}(s)\, s^d\, \mathcal{N}(|\boldsymbol{x}| s; \beta_*(\boldsymbol{x})). \tag{14.4}$$

When $|\boldsymbol{x}| = 0$, It is simply

$$L_{\alpha,k_*}(0; \beta_*) = (2\pi)^{-d/2}\, \mathbb{E}(X^d | \overline{\chi}_{\alpha,k}), \tag{14.5}$$

independent of β_*.

14.5. Marginal 1D Distribution

The construction of the 1D marginal distribution from Y extends directly from Section 13.7, where $\beta_{1(2)}$ is calculated.

Then the marginal distribution is an univariate GAS-SN: $Y_1 \sim L_{\alpha,k}(\xi_1, \Omega_{11}, \beta_{1(2)})$.

14.6. Quadratic Form

The quadratic form is

$$Q := \frac{1}{d}(Y - \xi)^\intercal \Omega^{-1} (Y - \xi) = \frac{1}{d} Z^\intercal \bar{\Omega}^{-1} Z. \tag{14.6}$$

This leads to the fractional extension of the classic F distribution.

Q distributes like a fractional F distribution, $Q \sim F_{\alpha,d,k}$ for all β. The QQ-plot between the empirical data and theoretical values is used to evaluate the goodness of a fit. A perfect fit should produce a 45-degree line.

To prove, from Section 14.1, we have $Z \sim Z_0/V$, $Z_0 \sim SN_d(0, \bar{\Omega}, \boldsymbol{\beta})$, and $V \sim \overline{\chi}_{\alpha,k}$. Put them together,

$$Q = \frac{1}{d} Z^\intercal \bar{\Omega}^{-1} Z = \frac{Z_0^\intercal \bar{\Omega}^{-1} Z_0}{d\, V^2} \sim \left(\frac{X^2}{d} \right) / V^2$$

where $X \sim \mathcal{N}_d(0, \bar{\Omega})$, according to Lemma 13.1.

Since $X^2 \sim \chi_d^2$ and $V^2 \sim \overline{\chi}_{\alpha,k}^2$, this leads to $Q \sim F_{\alpha,d,k}$, according to Section 8.1.

Azzalini (2013) provided a point of validation from his multivariate skew-t distribution. From Section 6.2 of [1], Q of a skew-t variable distributes like the classic $F(d, k)$. This is a special case of our fractional F distribution at $\alpha = 1$. That is, $Q \sim F_{1,d,k}$.

14.7. Multivariate MLE

TODO write better

A stochastic gradient decent (SGD) method on the maximum likelihood estimation (MLE) can be implemented efficiently. First, we calculate the sum of the minus-log of the PDF evaluated at every data point. This sum is called MLLK. Then we calculate the gradients for each hyperparameter. And make a small move along the direction that is most likely to minimize the MLLK. The scale of the move is based on a learning rate, which can be adjusted dynamically. Some randomness can be added to the small move. This allows the algorithm to explore the nearby regions, and maybe get "unstuck" over unexpected edges.

Use the bivariate optimization as an example. The hyperparameter space is

$$\Theta = \{\rho, \alpha, k, \beta_1, \beta_2, w_1, w_2, \xi_1, \xi_2\}$$

where $\alpha \in (0, 2)$, $k \in (2, \infty)$, $w_1 > 0, w_2 > 0$, and $\beta_1, \beta_2, \xi_1, \xi_2 \in \mathbb{R}$. Since $\rho \in (-1, 1)$, it is preferred to use a transformed parameter $\rho_\theta = \arctan(\rho/\pi) \in \mathbb{R}$.

It is also found that each dimension in the data set should be normalized to one standard deviation. This allows all the gradients to have similar scales. This helps the SGD algorithm.

Let Y represent the data set of size N, and $L(Y_i; \Theta)$ is the PDF, then

$$\begin{aligned}
\mathrm{MLLK}(\Theta; Y) &:= -\sum_{i=1}^{N} \log\, L(Y_i; \Theta) \\
\mathrm{Gradient}(\Theta) &:= \left\{ \frac{\partial\, \mathrm{MLLK}(\Theta; Y)}{\partial \Theta_j}, \forall\, \Theta_j \in \Theta \right\}
\end{aligned}$$

When N is large, it may not be computationally feasible to evaluate every $L(Y_i; \Theta)$. One may use histogram to compress the data into smaller numbers of bins.

Regularization can be added to the MLLK. For instance, we find it makes a lot of sense to add the L2 distances of the skewness and kurtosis on the marginal 1D distributions.

CHAPTER 15

Multivariate GAS-SN Adaptive Distribution

15.1. Definition

The goal of an adaptive distribution is to allow each dimension to have its own shape parameter in α, k. This is the departure from the the elliptical distribution.

Therefore, $\boldsymbol{\alpha} = \{\alpha_i\}$ is a d-dimensional vector, so is $\boldsymbol{k} = \{k_i\}$. We now have a list of standard FCM to work with: $\{\overline{\chi}_{\alpha_i,k_i}, i \in 1, 2..., d\}$.

Definition 15.1. Assume Z_0 is a d-dimensional random variable from a standard $d \times d$ multivariate skew-normal (SN) distribution, $SN_d(0,\bar{\Omega}, \boldsymbol{\beta})$, where $\bar{\Omega}$ is a correlation matrix.

Let Z be a d-dimensional random variable. Each element is a ratio distribution such as $Z_i \sim (Z_0)_i/\overline{\chi}_{\alpha_i,k_i}$. Then $Z \sim \overrightarrow{L}_{\boldsymbol{\alpha},\boldsymbol{k}}(0, \bar{\Omega}, \boldsymbol{\beta})$ is a standard multivariate GAS-SN adaptive distribution. The arrow-over sign is to emphasize the vector nature of $(\boldsymbol{\alpha}, \boldsymbol{k})$.

Assume $\mathcal{N}(x; \bar{\Omega})$ is the PDF of a standard multivariate normal distribution $N(0, \bar{\Omega})$ [**26**]. $\Phi_{\mathcal{N}}(x)$ is the CDF of a standard normal distribution.

The PDF of $Z \sim \overrightarrow{L}_{\boldsymbol{\alpha},\boldsymbol{k}}(0, \bar{\Omega}, \boldsymbol{\beta})$ is

$$\overrightarrow{L}_{\boldsymbol{\alpha},\boldsymbol{k}}(\boldsymbol{x}; \bar{\Omega}, \boldsymbol{\beta}) = 2 \int \cdots \int_0^\infty \mathcal{N}\left(\boldsymbol{s}\,\boldsymbol{x}; \bar{\Omega}\right)\, \Phi_{\mathcal{N}}\Big(\boldsymbol{\beta}^\intercal (\boldsymbol{s}\,\boldsymbol{x})\Big) \prod_{i=1}^{d} s_i ds_i\, \overline{\chi}_{\alpha_i,k_i}(s_i). \tag{15.1}$$

where $\boldsymbol{s} := \mathrm{diag}(s_1, ..., s_d)$ is the $d \times d$ diagonal matrix from the vector $\{s_i\}$. It is easy to see how it is reduced to a univariate GAS-SN distribution when $d = 1$.

Compared to the elliptical PDF (14.1), the major difference is that (15.1) is a d-dimensional integral. This is much more computationally demanding.

15.2. Location-Scale Family

The location-scale family follows the standard procedure: $Y = \xi + \omega Z$, which is denoted as $Y \sim \overrightarrow{L}_{\boldsymbol{\alpha},\boldsymbol{k}}(\xi, \Omega, \boldsymbol{\beta})$. The covariance matrix is $\Omega = \omega^\intercal \bar{\Omega} \omega$, and ω is the $d \times d$ diagonal scale matrix.

The PDF of Y is

$$\overrightarrow{L}_{\boldsymbol{\alpha},\boldsymbol{k}}(\boldsymbol{x};\, \xi, \Omega, \boldsymbol{\beta}) := \det(\omega)^{-1}\, \overrightarrow{L}_{\boldsymbol{\alpha},\boldsymbol{k}}(\boldsymbol{z}; \bar{\Omega}, \boldsymbol{\beta}). \tag{15.2}$$

where $\boldsymbol{z} := \omega^{-1}(\boldsymbol{x} - \xi)$. Notice that it has to be computed via the standard PDF because the mixtures $\{s_i\}$ must work with the standardized variable Z, not the location-scale variable Y.

15.3. Moments

The first moment of Z is $\mu_z := \boldsymbol{b} \odot \boldsymbol{\delta}$, where $b_i := \sqrt{2/\pi}\, \mathbb{E}(X^{-1} | \overline{\chi}_{\alpha_i,k_i})$ and $\odot$ is the Hadamard product.

The (i, j) element of the second moment of Z is

$$\boldsymbol{m_2}(i,j) := \begin{cases} \mathbb{E}(X^{-2} | \overline{\chi}_{\alpha_i,k_i}) & \text{if } i = j, \\ \bar{\Omega}_{i,j} \mathbb{E}(X^{-1} | \overline{\chi}_{\alpha_i,k_i}) \mathbb{E}(X^{-1} | \overline{\chi}_{\alpha_j,k_j}) & \text{if } i \neq j. \end{cases}$$

where $\bar{\Omega}_{i,i} = 1$ is ignored in the first line. Hence, the covariance is

$$\mathrm{var}\{Z\} := \Sigma_z := \boldsymbol{m_2} - \mu_z\, \mu_z^\intercal$$

The moments of Y follow the rule of the location-scale family. The first moment of Y is $\xi + \omega\, \mu_z$. The covariance of Y is $\omega\, \mathrm{var}\{Z\}\, \omega$.

15.4. Canonical Form

The adaptive distribution *doesn't* enjoy the rotational symmetry that an elliptical distribution has. Its canonical form is *not* particularly useful, since it has no connection to other distributions in the family through an affine transformation.

Assume the variable Z^* is a *canonical variable*. Then $Z^* \sim \overrightarrow{L}_{\boldsymbol{\alpha},\boldsymbol{k}}(0, \boldsymbol{I}_d, \boldsymbol{\beta}_{Z*})$, where $\boldsymbol{\beta}_{Z*} = (\beta_*, 0, ..., 0)^\intercal$ and β_* is defined by (13.10).

The PDF of Z^* is

$$\overrightarrow{L}_{\boldsymbol{\alpha},\boldsymbol{k}_*}(\boldsymbol{x}; \beta_*) = L_{\alpha_1,k_1}(x_1; \beta_*) \prod_{j=2}^{d} L_{\alpha_j,k_j}(x_j). \tag{15.3}$$

We can clearly see that only the first component is GAS-SN, all other components are GSaS, each with its own (α, k) shape.

Only the first component of its μ_z is non-zero, which is $\sqrt{2/\pi}\, \delta_*\, \mathbb{E}(X^{-1}|\overline{\chi}_{\alpha_1,k_1})$. Its $\boldsymbol{m_2}$ is a diagonal matrix where $\boldsymbol{m_2}(i,i) = \mathbb{E}(X^{-2}|\overline{\chi}_{\alpha_i,k_i})$.

15.5. Marginal 1D Distribution

The construction of the 1D marginal distribution from Y extends directly from Section 13.7, where $\beta_{1(2)}$ is calculated.

Then the marginal distribution is an univariate GAS-SN: $Y_1 \sim L_{\alpha_1,k_1}(\xi_1, \Omega_{11}, \beta_{1(2)})$.

15.6. Quadratic Form

TODO The corresponding F distribution is very hard. I have not figured this out yet.

15.7. 2D Adaptive MLE

TODO this needs more refinement since a normal 2D MLE doesn't work here.

TODO I am still working on the numerical method.

A stochastic gradient decent (SGD) method on the maximum likelihood estimation (MLE) can be implemented, but some adjustments are needed. Use the bivariate optimization as an example. The hyperparameter space is

$$\Theta = \{\rho, \alpha_1, \alpha_2, k_1, k_2, \beta_1, \beta_2, w_1, w_2, \xi_1, \xi_2\}$$

where $\alpha_1, \alpha_2 \in (0, 2)$, $k_1, k_2 \in (2, \infty)$, $w_1 > 0, w_2 > 0$, and $\beta_1, \beta_2, \xi_1, \xi_2 \in \mathbb{R}$. Since $\rho \in (-1, 1)$, it is preferred to use a transformed parameter $\rho_\theta = \arctan(\rho/\pi) \in \mathbb{R}$.

The computation of the adaptive PDF is very slow on a desktop, even for two dimensions. The MLLK is modified to perform on the two marginal 1D distributions. We supplement it with a regularization on the L2 distance of the correlation coefficient.

It is also found that each dimension in the data set should be normalized to one standard deviation. This allows all the gradients to have similar scales. This helps the SGD algorithm.

Let Y represent the data set of size N, and $L_m(Y_i;\Theta)$ is the marginal 1D PDF at dimension m ($m = 1...d$), then

$$\mathrm{MLLK}(\Theta;Y) := -\sum_{i=1}^{N}\sum_{m=1}^{d} \log\, L_m(Y_i;\Theta)$$

$$\mathrm{Gradient}(\Theta) := \left\{ \frac{\partial\,\mathrm{MLLK}(\Theta;Y)}{\partial\Theta_j}, \forall\, \Theta_j \in \Theta \right\}$$

Once the MLLK and gradients are calculated. The program makes a small move along the direction that is most likely to minimize the MLLK. The scale of the move is based on a learning rate, which can be adjusted dynamically. Some randomness can be added to the small move. This allows the algorithm to explore the nearby regions, and maybe get "unstuck" over unexpected edges.

When N is large, it may not be computationally feasible to evaluate every $L(Y_i;\Theta)$. One may use histogram to compress the data into smaller numbers of bins.

More regularization can be added to the MLLK. For instance, we find it makes a lot of sense to add the L2 distances of the skewness and kurtosis on the marginal 1D distributions.

We also regulate the mean of the quadratic form. But the exact distribution of the quadratic form is still under research.

CHAPTER 16

Fitting SPX-VIX Daily Returns with Bivariate Distributions

Two MLE fits are performed for the VIX/SPX daily log returns from 1990 to 2025. The first fit uses the bivariate elliptical GAS-SN distribution. The second fit uses the bivariate adaptive GAS-SN distribution.

The major difference is that the adaptive distribution allows each dimension to have its own (α, k) shape. However, it is much more compute-intensive, it requires alternative methods to work around. And it breaks the rotational symmetry that the elliptical distribution has. This requires a different approach to evaluate the quadratic form.

16.1. Elliptical Fit

TODO describe the fit outcome in more detail.

Correlation matches nicely. The distribution of quadratic form also matches nicely.

and each 1D marginal

The major issue with the fit is that the peak of the marginal PDF for VIX is higher than the observed peak. On the other hand, the theoretical peak in the SPX marginal is lower than the observed peak.

Strangely, the elliptical MLE finds the best fit at $\alpha = 0.75, k = 4.5$. When projected to 1D marginal distributions, such univariate GAS-SN is near the border of infinite kurtosis. (The reader is reminded that the degrees of freedom need to be higher than 4 to have valid kurtosis in the Student's t distribution. $k = 4.5$ is in the neighborhood of that threshold.)

TODO localize – Figure 4 or 5 of [13].

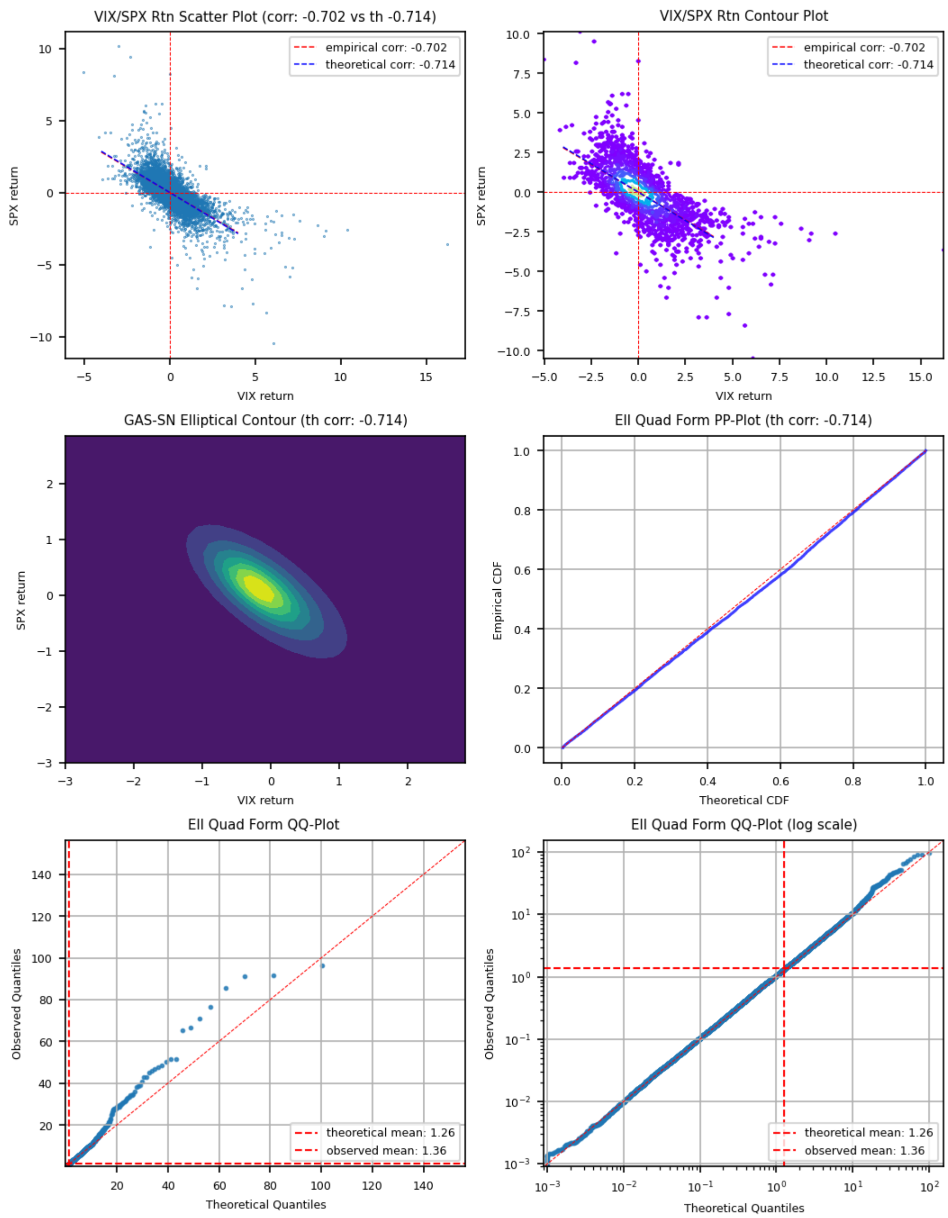

FIGURE 16.1. MLE fit of VIX/SPX daily log returns from 1990 to 2025 with the bivariate elliptical GAS-SN distribution. Data is standardized to one standard deviation on each axis.

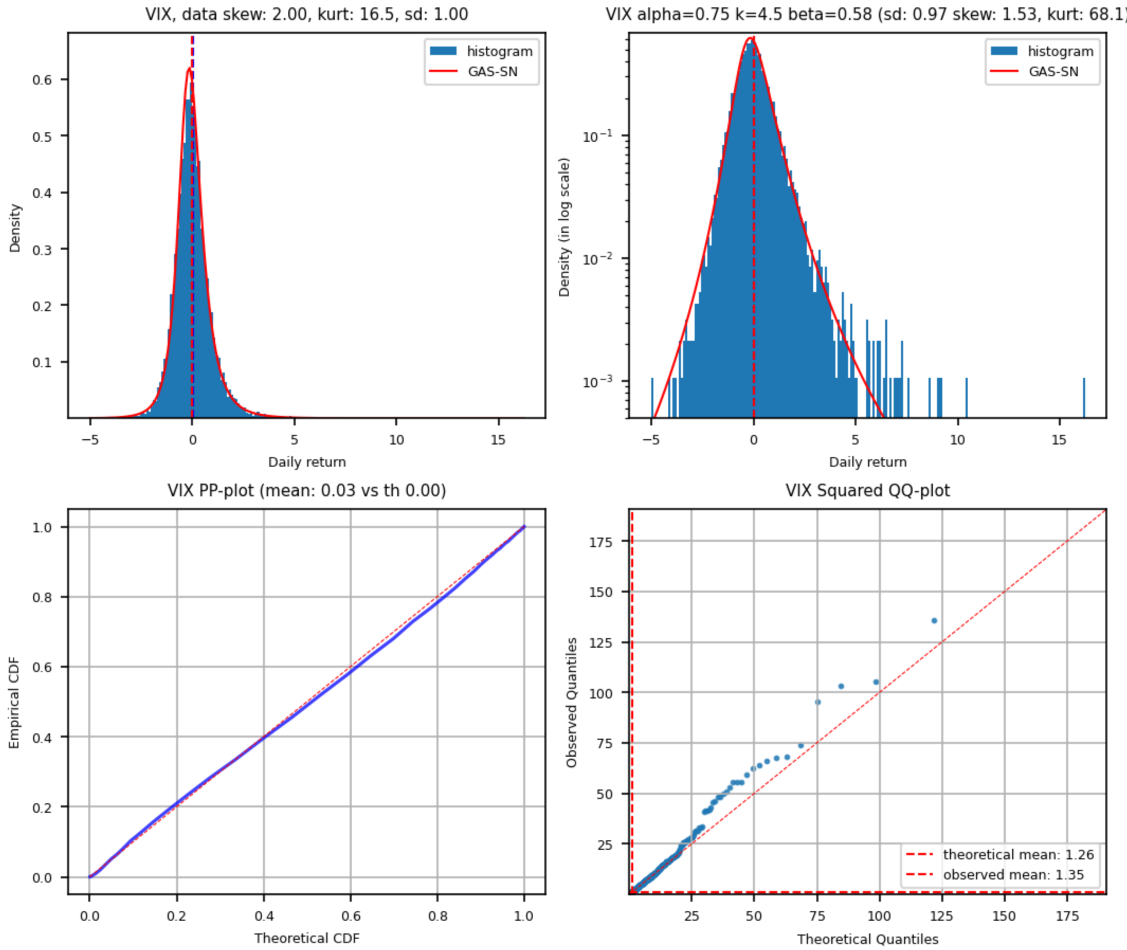

FIGURE 16.2. VIX Marginal from MLE fit of VIX/SPX daily log returns from 1990 to 2025 with the bivariate elliptical GAS-SN distribution. Data is standardized to one standard deviation on each axis.

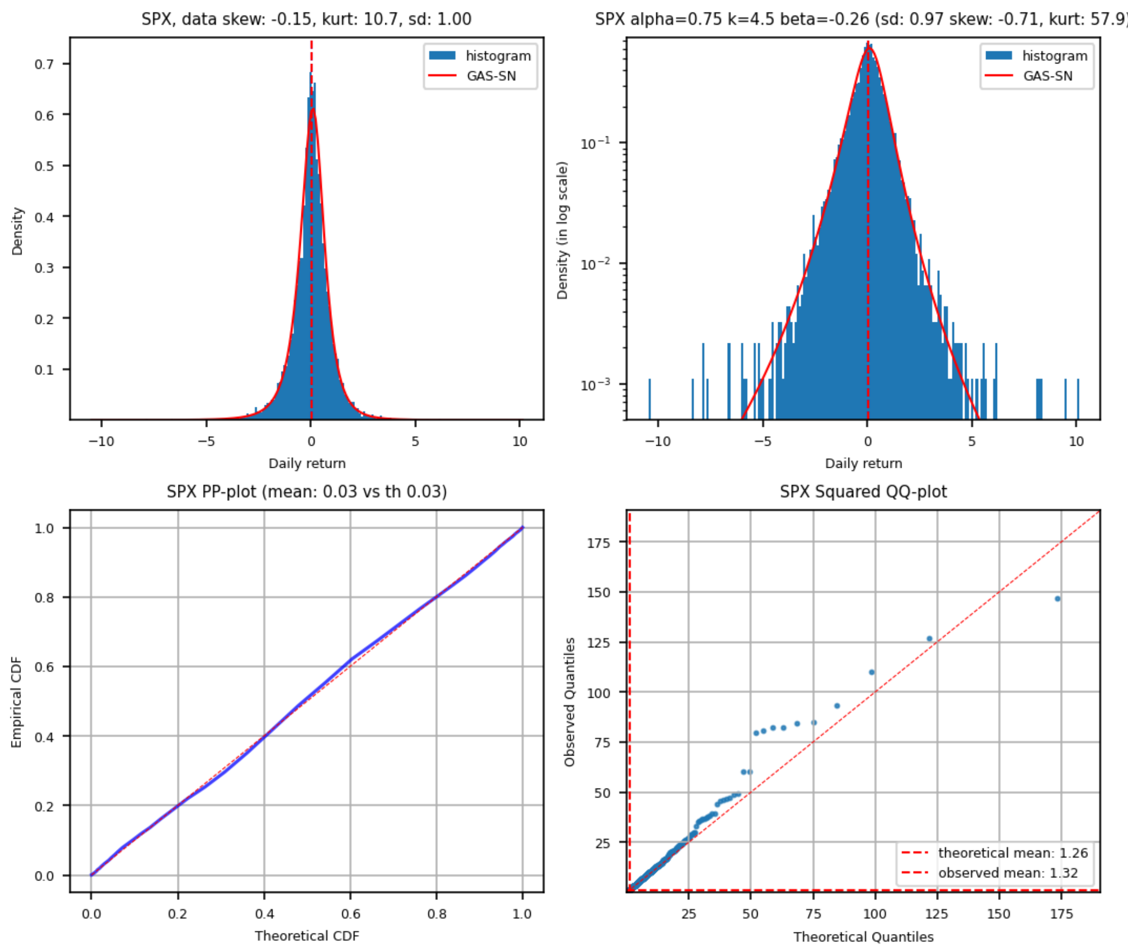

FIGURE 16.3. SPX Marginal from MLE fit of VIX/SPX daily log returns from 1990 to 2025 with the bivariate elliptical GAS-SN distribution. Data is standardized to one standard deviation on each axis.

16.2. Adaptive Fit

The adaptive fit is done by MLE on the two marginal distributions with regularization, e.g. the L2 distance between the empirical and theoretical correlations. This is a hack since a direct bivariate MLE is computationally infeasible on my workstation.

The adaptive fit produces the contour plot with somewhat rectangular shapes. That is quite impressive.

The theoretical correlation gets to -0.5, but unable to be closer to the empirical correlation of -0.7.

One would think the adaptive distribution allows each dimension to express its own shape. It should be much easier to produce a good fit. But the interaction between the correlation parameter and the skew parameters is quite complicated.

It is difficult to get the skewness and kurtosis to match in the SPX marginal. It is very complex to navigate the region near $\alpha \approx 1, k \approx 3$. In the Student's t distribution, the skewness and kurtosis are not defined.

The quadratic form needs a multiplier (scale adjustment) to produce a good fit. The origin of this multiplier requires further study.

In the squared QQ plots of the marginals, the fits don't capture the tails as good as the elliptical fits. This is somewhat disappointing.

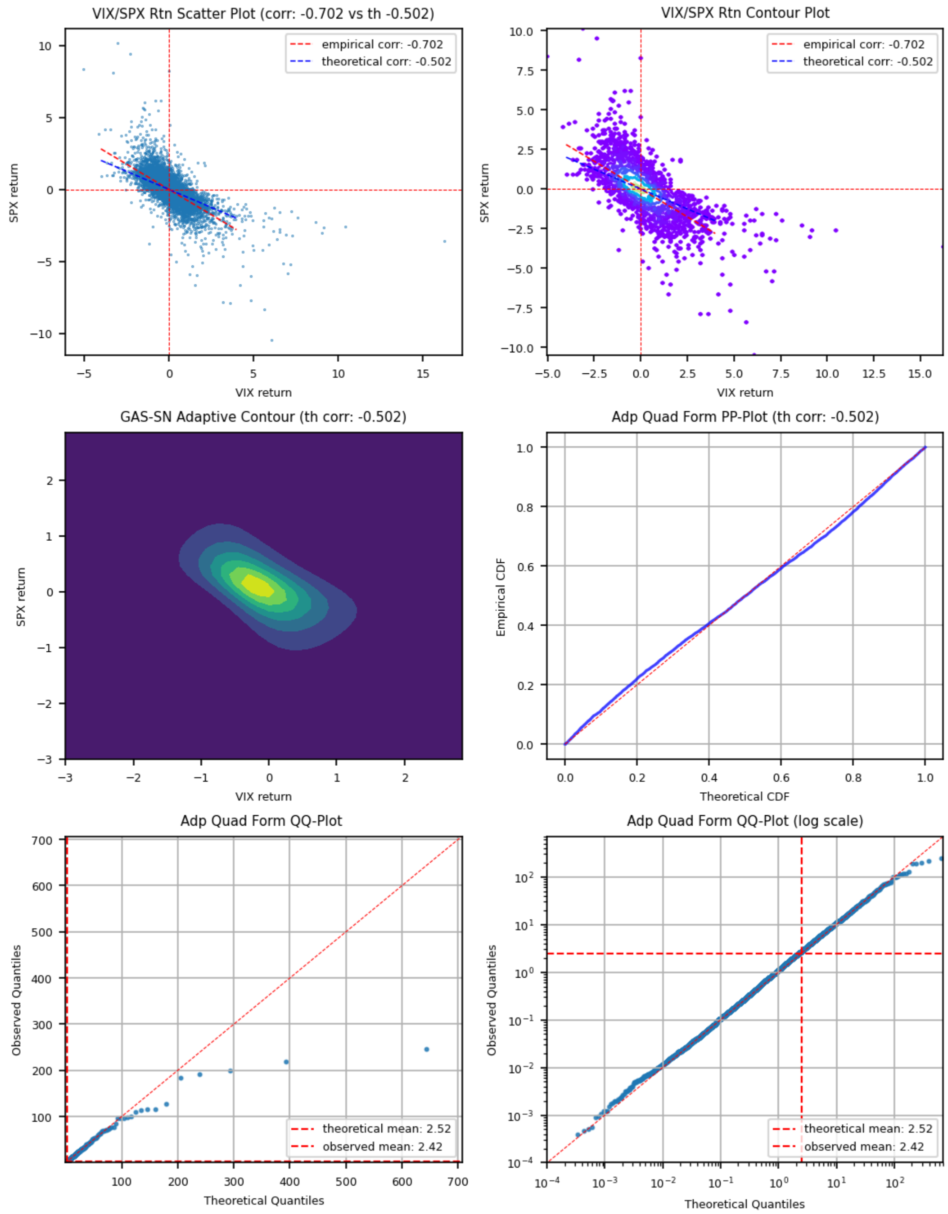

FIGURE 16.4. MLE fit of VIX/SPX daily log returns from 1990 to 2025 with the bivariate adaptive distribution. Data is standardized to one standard deviation on each axis.

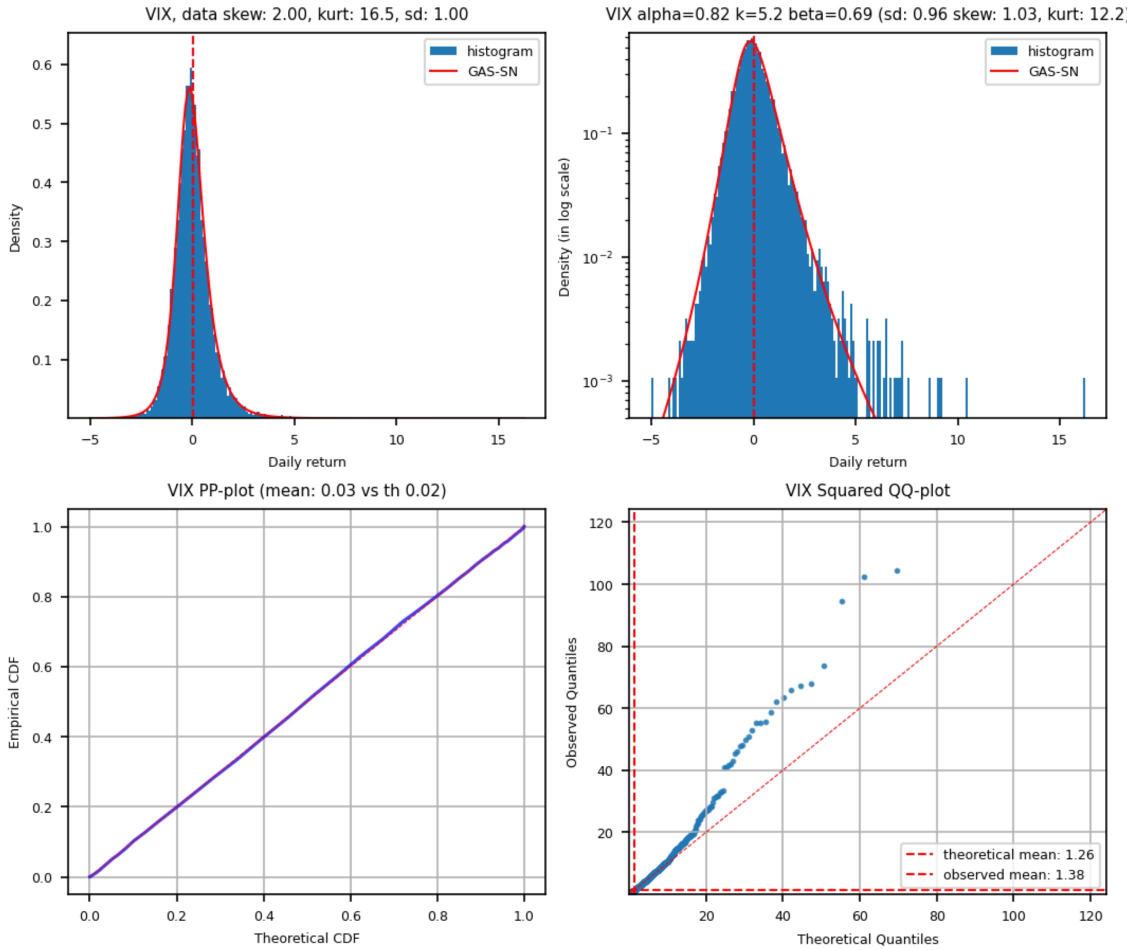

FIGURE 16.5. VIX Marginal from MLE fit of VIX/SPX daily log returns from 1990 to 2025 with the bivariate adaptive GAS-SN distribution. Data is standardized to one standard deviation on each axis.

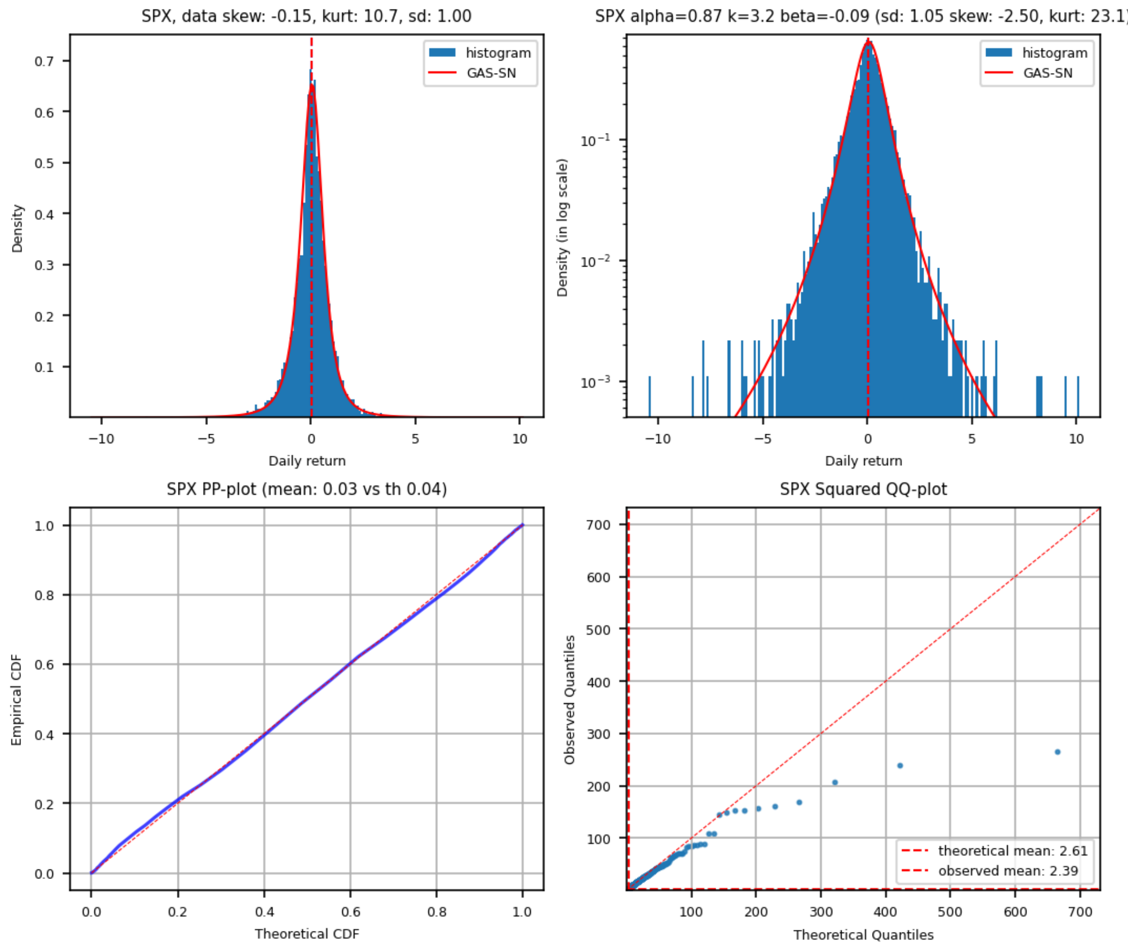

FIGURE 16.6. SPX Marginal from MLE fit of VIX/SPX daily log returns from 1990 to 2025 with the bivariate adaptive GAS-SN distribution. Data is standardized to one standard deviation on each axis.

APPENDIX A

List of Useful Formula

A.1. Gamma Function

Gamma function is used extensively in this paper. First, note that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Its **reflection formula** is

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)} \tag{A.1}$$

And the **Legendre duplication formula** is

$$\Gamma(z)\Gamma(z+1/2) = 2^{1-2z}\sqrt{\pi}\,\Gamma(2z) \tag{A.2}$$

Gamma function Asymptotic: At $x \to 0$, gamma function becomes

$$\lim_{x\to 0} \Gamma(x) \sim \frac{1}{x} \tag{A.3}$$

$$\lim_{x\to 0} \frac{\Gamma(ax)}{\Gamma(bx)} = \frac{b}{a} \qquad (ab \neq 0)$$

For a very large x, assume a, b are finite,

$$\lim_{x\to\infty} \frac{\Gamma(x+a)}{\Gamma(x+b)} \sim x^{a-b} \tag{A.4}$$

Sterling's formula is used to expand the kurtosis formula for a large k, which is:

$$\lim_{x\to\infty} \Gamma(x+1) \sim \sqrt{2\pi x}\left(\frac{x}{e}\right)^x, \tag{A.5}$$

$$\text{or } \lim_{x\to\infty} \Gamma(x) \sim \sqrt{2\pi}\, x^{x-1/2} e^{-x}. \tag{A.6}$$

A.2. Transformation

Laplace transform of cosine is[1]

$$\int_0^\infty dt\, \cos(xt) e^{-t/\nu} = \frac{\nu^{-1}}{x^2+\nu^{-2}} = \frac{\nu}{(\nu x)^2+1} \tag{A.7}$$

Gaussian transform of cosine is[2]

$$\int_0^\infty dt\, \cos(xt)\, e^{-t^2/2} = \sqrt{\frac{\pi}{2}}\, e^{-x^2/2} \tag{A.8}$$

$$\text{Hence } \int_0^\infty dt\, \cos(xt)\, e^{-t^2/2s^2} = \sqrt{\frac{\pi}{2}}\, s\, e^{-(sx)^2/2}$$

[1]See https://proofwiki.org/wiki/Laplace_Transform_of_Cosine

[2]See https://www.wolframalpha.com/input?i=integrate+cos%28a+x%29+e%5E%28-x%5E2%2F2%29+dx+from+0+to+infty

A.3. Half-Normal Distribution

The moments of the half-normal distribution (HN)[3] are used several times. Its PDF is defined as

$$p_{HN}(x;\sigma) := \sqrt{\frac{2}{\pi}}\frac{1}{\sigma}\,e^{-x^2/(2\sigma^2)}\,,\ x>0 \tag{A.9}$$

which is a special case of GG with $d=1, p=2, a=\sqrt{2}\sigma$. Its moments are

$$E_{HN}(T^n) = \sigma^n\frac{2^{n/2}}{\sqrt{\pi}}\,\Gamma\left(\frac{n+1}{2}\right) \tag{A.10}$$

which are the same as those of a normal distribution.

[3]See https://en.wikipedia.org/wiki/Half-normal_distribution

Bibliography

[1] Adelchi Azzalini. *The Skew-Normal and Related Families*. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2013.

[2] H. Bateman and A. Erdélyi. *Higher Transcendental Functions*, volume 3 of *California Institute of Technology Bateman Manuscript Project*. McGraw-Hill Book Company., 1955.

[3] Berndt Bruce C. *Ramanujan's Notebooks, Part I*. Springer New York, NY, 1985.

[4] J.C. Crisanto-Neto, M.G.E. da Luz, E.P. Raposo, and G.M. Viswanathan. An efficient series approximation for the lévy α-stable symmetric distribution. *Physics Letters A*, 382(35):2408–2413, 2018.

[5] *NIST Digital Library of Mathematical Functions*. https://dlmf.nist.gov/, Release 1.2.0 of 2024-03-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[6] William Feller. On a generalization of marcel riesz' potentials and the semi-groups generated by them. *Meddelanden Lunds Universitets Matematiska Seminarium*, pages 73–81, 1952.

[7] William Feller. *An Introduction to Probability Theory and its Applications (2nd Edition)*, volume 2. John Wiley & Sons, Inc., 1971.

[8] Wikimedia Foundation. Inverse distribution, 2023. Retrieved December 1, 2023, from https://en.wikipedia.org/wiki/Inverse_distribution.

[9] George Karagiannidis and Athanasios Lioumpas. An improved approximation for the gaussian q-function. *IEEE Communications Letters*, 11:644–646, 08 2007.

[10] Paul Lévy. *Calcul des Probabilités*. Gauthier-Villars, 1925.

[11] Xiaoyue Li and John M. Mulvey. Optimal portfolio execution in a regime-switching market with non-linear impact costs: Combining dynamic program and neural network. *arXiv preprint*, 6 2023. Available at arXiv: https://arxiv.org/abs/2306.08809.

[12] Stephen H.T. Lihn. Stable count distribution for the volatility indices and space-time generalized stable characteristic function. *SSRN Working Paper*, 7 2020. Available at SSRN: https://ssrn.com/abstract=3659383.

[13] Stephen H.T. Lihn. Generalization of the alpha-stable distribution with the degree of freedom. *arXiv preprint*, 5 2024. Available at arXiv: https://arxiv.org/abs/2405.04693.

[14] F. Mainardi. *Fractional Calculus and Waves in Linear Viscoelasticity: An Introduction to Mathematical Models*. Imperial College Press, 2010. See Appendix F at https://appliedmath.brown.edu/sites/default/files/fractional/36%20TheWrightFunctions.pdf.

[15] Francesco Mainardi and Armando Consiglio. The wright functions of the second kind in mathematical physics. *Mathematics*, 8(6), 2020.

[16] Francesco Mainardi, Yuri Luchko, and Gianni Pagnini. The fundamental solution of the space-time fractional diffusion equation. *Frac. Calc. Appl. Anal.*, 4, 03 2007. See https://arxiv.org/abs/cond-mat/0702419.

[17] Francesco Mainardi and Gianni Pagnini. Mellin-barnes integrals for stable distributions and their convolutions. *Fractional Calculus and Applied Analysis*, 11:443–456, 2008.

[18] Arak Mathai and Hans Haubold. *Fractional and Multivariable Calculus: Model Building and Optimization Problems*. Springer Cham, 12 2017.

[19] Hjalmar Mellin. Zur theorie zweier allgemeinen klassen bestimmter integrale. *Acta Societatis Scientiarum Fennicæ*, page 1–75, 1897.

[20] The mpmath development team. *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 1.3.0)*, 2023. http://mpmath.org/.

[21] John Nolan. *Univariate Stable Distributions, Models for Heavy Tailed Data*. Springer Cham, 01 2020.

[22] Donald B. Owen. Tables for Computing Bivariate Normal Probabilities. *The Annals of Mathematical Statistics*, 27(4):1075 – 1090, 1956.

[23] W. R. Schneider. Stable distributions: Fox function representation and generalization. In S. Albeverio, G. Casati, and D. Merlini, editors, *Stochastic Processes in Classical and Quantum Systems*, pages 497–511, Berlin, Heidelberg, 1986. Springer Berlin Heidelberg.

[24] Y. Shu, C. Yu, and J.M. Mulvey. Downside risk reduction using regime-switching signals: a statistical jump model approach. *Journal of Asset Management*, 25:493–507, 2024.

[25] E. W. Stacy. A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, 33(3):1187 – 1192, 1962. See also https://en.wikipedia.org/wiki/Generalized_gamma_distribution.

[26] Y.L. Tong. *The Multivariate Normal Distribution*. Research in Criminology. Springer New York, 1990.

[27] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[28] E. M. Wright. On the coefficients of power series having exponential singularities. *Journal London Math. Soc.*, s1-8, Issue 1 (1933):71–79, 1933.

[29] E. M. Wright. The asymptotic expansion of the generalized bessel function. *Proc. London Math. Soc.*, s2-38, Issue 1 (1935):257–270, 1935.