SIGNED MEASURES IN EXCHANGEABILITY AND INFINITE DIVISIBILITY

Gary J. Kerns

A Dissertation

Submitted to the Graduate College of Bowling Green
State University in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2004

Committee:

Gabor Szekely, Advisor

Timothy S. Fuerst
Graduate Faculty Representative

Hanfeng Chen

Truc Nguyen

Craig Zirbel

ABSTRACT

Gábor J. Székely, Advisor

The focus of this research was to explore the mathematical uses of signed measures in Probability Theory. Two distinct areas were investigated.

The first involved exchangeability and the representation theorem of de Finetti, which says that an infinite sequence of binary valued exchangeable random variables are mixtures of independent, identically distributed (i.i.d.) random variables. It is well known that the theorem does not hold in general for finite sequences, in particular, it works only for sequences that are nonnegatively correlated. This research proved that while for infinite sequences the classical (that is, nonnegative) mixtures of i.i.d. random variables are sufficient, with some finite sequences a signed "mixture" is needed to retain de Finetti's convenient representation.

Two applications of this idea were examined. One concerned Bayesian consistency, in which it was established that a sequence of posterior distributions continues to converge to the true value of a parameter θ under much wider assumptions than are ordinarily supposed. The next pertained to Statistical Physics, and it was demonstrated that the quantum statistics of Fermi-Dirac may be derived from the statistics of classical (*i.e.* independent) particles by means of a signed mixture of multinomial distributions.

The second area of this research concerned infinitely divisible (ID) random variables. The class of **generalized infinitely divisible** (GID) random variables was defined, and its properties were investigated. It was submitted that under broad conditions the class ID is significantly extended, admitting nonincreasing discrete distributions with bounded

support and continuous candidates such as the uniform distribution on $[0,1]$; all of which are known to be not divisible in the classical sense.

Additionally the class of **strongly infinitely divisible** (SID) distributions was introduced and under a technical assumption of asymptotic negligibility it was shown that many classical traits may be salvaged, such as nonvanishing characteristic functions with a modified Lévy-Khintchine representation. On the other hand, the symmetry is not exact, as displayed by a lack of closure of the SID class.

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to my advisor, Dr. Gábor J. Székely, for his patience, encouragement, and inspiration throughout the entirety of this research. I thank all of the members of my committee, Dr. Craig Zirbel, Dr. Hanfeng Chen, Dr. Truc Nguyen, and Dr. Timothy Fuerst for their advice and helpful comments. A wealth of appreciation is extended to the faculty of the Department of Mathematics and Statistics, particularly the chair Dr. Neal Carothers, for superb instruction, counsel, and support during my short time at Bowling Green State University. I am deeply grateful to Cyndi Patterson, Marcia Seubert, and Mary Busdeker for their frequent assistance and persistent professionalism.

I am wholly indebted to my family, for their love.

Finally, I would like to thank Xiaojing Lu, to whom I dedicate this dissertation.

TABLE OF CONTENTS

CHAPTER 1
PRELIMINARIES

1.1 Random Variables and Expectation

Let $(\Omega, \mathcal{A}, \nu)$ be a complex (respectively signed) measure space. If ν is normalized so that $\nu(\Omega) = 1$, then we call $(\Omega, \mathcal{A}, \nu)$ a **complex (respectively signed) probability space**. A **random variable** is an extended real-valued $\mathcal{A}$-measurable function on Ω.

Every random variable X has associated with it a complex measure ν_X on the Borel subsets of the real line defined by

$$\nu_X(B) = \nu \circ X^{-1}(B) = \nu\left(X^{-1}(B)\right),$$

for all $B \in \mathcal{B}$. We say that X has the **distribution** ν_X or that X is **distributed** according to the complex measure ν_X, and we write $X \sim \nu_X$. When the random variable is understood we will just write ν. Similarly, there exists a function of bounded variation F_X, defined for $x \in \mathbb{R}$ by $F_X(x) = \nu_X\left((-\infty, x]\right)$.

Of course, the random variable X has two other measures associated with it, namely the variation measure $|\nu|$ and the normalized variation measure $|\nu|/\|\nu\|$. Sometimes it will be necessary to refer to them explicitly. We will do so with the notation $X \overset{\text{v}}{\sim} |\nu|$ and $X \overset{\text{nv}}{\sim} |\nu|/\|\nu\|$. Note that if $X \overset{\text{nv}}{\sim} \mu$ then μ is a classical probability distribution. Also, if $X \sim \nu$, then we denote $\mathrm{TVar}(X) = |\nu|(\mathbb{R})$.

There is some notation to be introduced to help us with the integration operations we will be using. For measurable f we denote the **expectation** of $f(X)$ by

$$\mathbb{E}_\nu f(X) = \int_{\mathbb{R}} f(x)\, \mathrm{d}\nu(x),$$

where the right hand side of the expression is the Lebesgue-Stieltjes integral. When the distribution ν is clear from the context we will abbreviate to $\mathbb{E}\, f(X)$. We notate

$$|\mathbb{E}|\, f(X) = \int f \,\mathrm{d}|\nu|, \quad \text{and} \quad \|\mathbb{E}\|\, f(X) = \int f \,\mathrm{d}\left(\frac{|\nu|}{\|\nu\|}\right);$$

that is, $|\mathbb{E}_\nu| = \mathbb{E}_{|\nu|}$ and $\|\mathbb{E}_\nu\| = \mathbb{E}_{|\nu|/\|\nu\|}$. From this notation naturally follows

$$\mathbb{P}(X \in B) = \mathbb{E}1_B = \nu(B),\ |\mathbb{P}|(X \in B) = |\mathbb{E}|1_B,\ \text{and}\ \|\mathbb{P}\|(X \in B) = \|\mathbb{E}\|1_B.$$

Notice that if ν is a classical probability distribution then $\mathbb{E} = |\mathbb{E}| = \|\mathbb{E}\|$ and $\mathbb{P} = |\mathbb{P}| = \|\mathbb{P}\|$.

It is appropriate to mention that many relevant results concerning signed measures are proven in the appendices. A portion of them are standard and the remaining are extensions of known properties of measures to the signed case. Where possible, the results are stated in a form that holds for the more general complex measure; however, for this dissertation only signed measures are needed.

Example 1.1.1 (The Generalized Bernoulli Distribution).

Perhaps the most simple example of a signed distribution is the signed analog of the discrete Bernoulli distribution. A random variable X has a *Generalized Bernoulli distribution* if

$$\mathbb{P}(X = 0) = 1 - p, \quad \mathbb{P}(X = 1) = p,$$

where p is a specified real number. We will denote this by $X \sim GBern(p)$. There are three cases to consider, depending on the value of p.

Case 1. $0 \le p \le 1$. This is the classical and well known $Bern(p)$ case.

Case 2. $p > 1$. The variation measure is

$$|\mathbb{P}|(X = 0) = p - 1, \quad |\mathbb{P}|(X = 1) = p$$

with total variation $2p-1$, and the normalized variation measure is

$$\|\mathbb{P}\|(X=0)=\frac{p-1}{2p-1},\quad \|\mathbb{P}\|(X=1)=\frac{p}{2p-1}.$$

Case 3. $p<0$. Here, the variation measure is

$$|\mathbb{P}|(X=0)=1+|p|,\quad |\mathbb{P}|(X=1)=|p|$$

with total variation $2|p|+1$, and the normalized variation measure is

$$\|\mathbb{P}\|(X=0)=\frac{1+|p|}{2|p|+1},\quad \|\mathbb{P}\|(X=1)=\frac{|p|}{2|p|+1}.$$

We may economize notation by denoting $GBern_{\pm}(p)$ according to whether $p>1$ or $p<0$, respectively. Then we may say

$$\text{If } X\sim GBern_{\pm}(p),\quad \text{then}\quad X\overset{\text{nv}}{\sim} Bern\left(\frac{|p|}{2|p|\mp 1}\right).$$

Notice that as p ranges from $-\infty$ to 0, the quantity $|p|/(2|p|+1)$ decreases from 1/2 to 0, and as p ranges from 1 to ∞, the probability $|p|/(2|p|-1)$ decreases from 1 to 1/2.

Example 1.1.2 (The Generalized Poisson Distribution).

A random variable X has a *Generalized Poisson distribution*, denoted $X\sim GPoi(\lambda)$, if

$$\mathbb{P}(X=k)=\frac{\lambda^k\,e^{-\lambda}}{k!},\quad k=0,1,2,\ldots$$

for some specified $\lambda\in\mathbb{R}$. Of course $\sum_k \mathbb{P}(X=k)=1$, and for $\lambda<0$ we have

$$\mathrm{TVar}(X)=\sum_{k=0}^{\infty}|\mathbb{P}(X=k)|=\sum_{k=0}^{\infty}\left|\frac{\lambda^k\,e^{-\lambda}}{k!}\right|=e^{-\lambda}\sum_{k=0}^{\infty}\frac{|\lambda|^k}{k!}=e^{2|\lambda|}.$$

But then this implies that

$$\|\mathbb{P}\|(X=k)=e^{-2|\lambda|}\frac{|\lambda|^k\,e^{-\lambda}}{k!}=\frac{|\lambda|^k\,e^{-|\lambda|}}{k!},$$

in other words, if $X\sim GPoi(\lambda)$ with $\lambda<0$, then $X\overset{\text{nv}}{\sim} Poi(|\lambda|)$.

Example 1.1.3 (The Generalized Geometric Distribution).

A random variable X has a *Generalized Geometric distribution*, denoted $X \sim GGeo(p)$, if

$$\mathbb{P}(X = k) = p(1-p)^k, \quad k = 0, 1, 2, \ldots$$

for some specified number p, $0 < p < 2$. Note that this distribution collapses to the classical $Geo(p)$ distribution when $0 < p < 1$. The fact that $\sum_k \mathbb{P}(X = k) = 1$ follows from the fact that the geometric series $\sum \alpha^k$ converges absolutely for $|\alpha| < 1$. And for $1 \leq p < 2$, we have

$$\mathrm{TVar}(X) = \sum_{k=0}^{\infty} |\mathbb{P}(X = k)| = p \sum_{k=0}^{\infty} (p-1)^k = \frac{p}{1-(p-1)} = \frac{p}{2-p}.$$

This of course implies that if $X \sim GGeo(p)$ for $1 \leq p < 2$, then $X \overset{\mathrm{nv}}{\sim} Geo(2-p)$.

Remark. *The radius of convergence being 1 for the geometric series explains why we only defined $GGeo(p)$ for $0 < p < 2$. The series simply diverges for all other p. This fact will become significant later.*

1.2 Fourier Transforms and Convolution

Definition 1.2.1. *We begin by defining the* ***Fourier-Stieltjes Transform*** *of $\nu \in M(\mathbb{R})$ to be*

$$\mathcal{F}\nu(t) = \widehat{\nu}(t) = \int_{\mathbb{R}} e^{-2\pi i t x} \, \mathrm{d}\nu(x).$$

Note that if $\mathrm{d}\nu = f \mathrm{d}m$ for some $f \in L^1(m)$, then the above becomes

$$\mathcal{F}f(t) = \widehat{f}(t) = \int_{\mathbb{R}} e^{-2\pi i t x} f(x) \, \mathrm{d}m(x),$$

and $\widehat{f}$ is called the Fourier transform of $f \in L^1(m)$.

Since e^{itx} is uniformly continuous in x it is clear that $\widehat{\nu}$ is a bounded continuous function and that $\|\widehat{\nu}\|_u \leq \|\nu\|$.

Definition 1.2.2. *If* $\mu,\ \nu \in M(\mathbb{R})$, *we define their* ***convolution*** $\mu * \nu \in M(\mathbb{R})$ *by* $\mu * \nu(E) = \mu \times \nu(\alpha^{-1}(E))$, *where* $\alpha(x,y) = x + y$. *In other words,*

$$\mu * \nu(E) = \iint 1_E(x+y)\,\mathrm{d}\mu(x)\mathrm{d}\nu(y).$$

The operation of convolution of measures corresponds to the addition of the respective random variables. We summarize some important properties of convolution in the next proposition.

Proposition 1.2.1 (Folland). *Let* $\mu,\ \nu \in M(\mathbb{R})$.

1. *For every bounded and measurable* h,

$$\int h\,\mathrm{d}(\mu * \nu) = \iint h(x+y)\,\mathrm{d}\mu(x)\mathrm{d}\nu(y).$$

2. *Convolution is commutative and associative.*

3. $\|\mu * \nu\| \leq \|\mu\|\,\|\nu\|$.

4. *If* $\mathrm{d}\mu = f\mathrm{d}m$ *and* $\mathrm{d}\nu = g\mathrm{d}m$, *then* $\mathrm{d}(\mu * \nu) = (f * g)\mathrm{d}m$.

5. $\widehat{\mu * \nu} = \widehat{\mu} \cdot \widehat{\nu}$.

In terms of random variables, for $\nu \in M(\mathbb{R})$ we have the **characteristic function** of $X \sim \nu$, denoted φ_X, defined to be $\varphi_X(t) = \mathbb{E}_\nu\, e^{itX}$ for $t \in \mathbb{R}$. Clearly $\widehat{\nu}(t) = \varphi_X(-2\pi t)$.

Proposition 1.2.2. *If* $f(t) = \mathbb{E}_\nu e^{itX}$ *is the characteristic function of* $X \sim \nu \in M$, *then* $|f|^2$ *is also a characteristic function of* $\mu \in M$ *which is real and nonnegative. If* $\nu \in M_\pm$, *then so is* μ.

Proof. We expand f into its real and imaginary parts:

$$\begin{aligned} f(t) &= \int (\cos tx + i \sin tx)\, \mathrm{d}\left(\nu_r(x) + i\nu_i(x)\right) \\ &= \int \cos tx\, \mathrm{d}\nu_r - \int \sin tx\, \mathrm{d}\nu_i + i \left(\int \cos tx\, \mathrm{d}\nu_i + \int \sin tx\, \mathrm{d}\nu_r \right), \end{aligned}$$

in which case

$$\begin{aligned} \overline{f(t)} &= \int \cos tx\, \mathrm{d}\nu_r - \int \sin tx\, \mathrm{d}\nu_i - i \left(\int \cos tx\, \mathrm{d}\nu_i + \int \sin tx\, \mathrm{d}\nu_r \right) \\ &= \int (\cos tx - i \sin tx)\, \mathrm{d}\left(\nu_r(x) - i\nu_i(x)\right) \\ &= \int e^{-itx}\, \mathrm{d}\overline{\nu}(x), \end{aligned}$$

where $\overline{\nu} = \nu_r - i\nu_i \in M$ satisfies

$$\int \mathrm{d}\overline{\nu} = \int \mathrm{d}\nu_r - i \int \mathrm{d}\nu_i = 1.$$

Now, let $X \sim \nu$ and let $Y \sim \overline{\nu}$ be independent of X. Then

$$|f|^2(t) = f(t) \cdot \overline{f(t)} = \mathbb{E}e^{itX} \cdot \mathbb{E}e^{it(-Y)} = \mathbb{E}e^{it(X-Y)}$$

and we see that $|f|^2$ is the characteristic function of $Z = X - Y \sim \mu \in M$. Obviously $|f|^2$ is real and nonnegative. It is equally clear that if $\nu \in M_\pm$ then $\nu = \overline{\nu}$; consequently the distribution of $-Y$ is also in $M_\pm$ which implies that the distribution of Z is ultimately in $M_\pm$, as claimed. □

Notation. *For a complex measure ν we will write ν^{*n} to denote the n-fold convolution $\nu * \nu * \cdots * \nu$. In this notation,*

$$|\nu^{*n}| = |\nu * \nu * \cdots * \nu| \quad \textit{and} \quad |\nu|^{*n} = |\nu| * |\nu| * \cdots * |\nu|.$$

Similarly, we denote

$$\mathbb{E}_\nu^{*n} = \mathbb{E}_{\nu^{*n}}, \quad |\mathbb{E}_\nu^{*n}| = \mathbb{E}_{|\nu^{*n}|}, \quad \textit{and} \quad |\mathbb{E}_\nu|^{*n} = \mathbb{E}_{|\nu|^{*n}}.$$

In terms of random variables, if $X \sim \nu^{*n}$, then we may think of X as having the form $X = X_1 + X_2 + \cdots + X_n$, where the X_i are independent and identically distributed according to the complex measure ν.

Proposition 1.2.3. *Suppose $\nu \in M$ and $X \sim \nu^{*n}$. Then for any $1 \leq p < \infty$,*

$$|\mathbb{E}|^{*n}|X|^p \leq n^p \, \|\nu\|^{n-1} \, |\mathbb{E}||X_1|^p. \tag{1.2.1}$$

On the other hand, if $X_1 \geq 0$ then for any $0 < \delta < 1$,

$$n^\delta \, \|\nu\|^{n-1} \, |\mathbb{E}|X_1^\delta \leq |\mathbb{E}|^{*n} X^\delta. \tag{1.2.2}$$

Proof. To prove equation (1.2.1), Minkowski's Inequality A.2.2 shows that

$$\|X\|_p \leq \sum_{k=1}^{n} \|X_k\|_p,$$

where we integrate with respect to the n-fold product measure $|\mu| \times \cdots \times |\mu|$. In that case the above becomes

$$\begin{aligned} \left(|\mathbb{E}|^{*n}|X|^p\right)^{1/p} &\leq \sum_{k=1}^{n} \left(\|\nu\|^{(n-1)}|\mathbb{E}||X_k|^p\right)^{1/p} \\ &= \|\nu\|^{(n-1)/p} \cdot n \cdot \left(|\mathbb{E}||X_1|^p\right)^{1/p}. \end{aligned}$$

Raising both sides to the p^{th} power shows the claim.

To prove equation (1.2.2), the function $f(x) = x^\delta$ is concave and for fixed ω Jensen's Inequality yields

$$\frac{X_1^\delta(\omega) + X_2^\delta(\omega) + \cdots + X_m^\delta(\omega)}{m} \leq \left(\frac{X_1(\omega) + X_2(\omega) + \cdots + X_m(\omega)}{m}\right)^\delta.$$

Taking expectations on both sides with respect to $|\nu| \times \cdots \times |\nu|$ gives

$$\frac{1}{m}\sum_{i=1}^{m} |\mathbb{E}|X_i^\delta \cdot \|\nu\|^{n-1} = |\mathbb{E}|X_1^\delta \cdot \|\nu\|^{n-1} \leq \frac{|\mathbb{E}|^{*n}X^\delta}{m^\delta}.$$

$\square$

Proposition 1.2.4. *If $\nu \in M_{\pm}$ and $X \sim \nu^{*n}$ then for measurable $f \geq 0$:*

$$|\mathbb{E}_{\nu}^{*n}|f(X) \leq |\mathbb{E}_{\nu}|^{*n} f(X).$$

Proof. We write

$$\begin{aligned}
\nu^{*n} &= \nu * \nu * \cdots * \nu \\
&= (\nu^{+} - \nu^{-}) * (\nu^{+} - \nu^{-}) * \cdots * (\nu^{+} - \nu^{-}) \\
&= \sum_{k=0}^{n} \binom{n}{k} (-1)^{k} (\nu^{+})^{n-k} * (\nu^{-})^{k} \\
&\leq \sum_{k=0}^{n} \binom{n}{k} (\nu^{+})^{n-k} * (\nu^{-})^{k} \\
&= |\nu| * |\nu| * \cdots * |\nu| \\
&= |\nu|^{*n},
\end{aligned}$$

where the (in)equalities are meant to hold setwise. That is, for any $E \in \mathcal{B}$ we have $\nu^{*n}(E) \leq |\nu|^{*n}(E)$ which, by virtue of Proposition A.1.1 shows that $|\nu^{*n}|(E) \leq |\nu|^{*n}(E)$ for all $E \in \mathcal{B}$. Thus the proposition is true when f is a nonnegative simple function, and then for $f \geq 0$ by the Monotone Convergence Theorem, since in that case we may take a sequence of nonnegative simple functions increasing to f. □

We continue the discussion of convolution with the extension of two families of random variables that will be very important in the next chapter.

Example 1.2.1 (The Generalized Negative Binomial Distribution).

A random variable X has a *Generalized Negative Binomial distribution* (denoted $GNegB(r,p)$) if

$$\mathbb{P}(X = k) = \binom{-r}{k} (-q)^{k} p^{r}, \quad k = 0, 1, 2, \ldots,$$

for some specified $r > 0$ and $0 < p < 2$. Here, $q = 1 - p$ and

$$\binom{-r}{k} = \frac{-r(-r-1)\cdots(-r-k+1)}{k!}.$$

Remark. *The above binomial coefficient simplifies to* $\binom{-r}{k}(-1)^k = \binom{r+k-1}{k}$*, and our definition matches the classical Negative Binomial distribution.*

The fact that $\sum_k \mathbb{P}(X = k) = 1$ follows from the binomial series and the observation $|-q| < 1$:

$$\sum_{k=0}^{\infty} \binom{-r}{k} (-q)^k p^r = p^r (1-q)^{-r} = 1.$$

When $1 \le p < 2$ we get a distribution whose total variation is

$$\sum_{k=0}^{\infty} |\mathbb{P}(X = k)| = p^r \sum_{k=0}^{\infty} \binom{-r}{k} (-|q|)^k = \left(\frac{p}{1-(p-1)}\right)^r = \left(\frac{p}{2-p}\right)^r.$$

This calculation now shows that

if $X \sim GNegB(r, p)$ with $r > 0$ and $1 \le p < 2$, then $X \overset{\text{nv}}{\sim} NegB(r, 2-p)$.

The characteristic function of $X \sim GNegB(r, p)$ is

$$\varphi_X(t) = \left(\frac{p}{1 - qe^{it}}\right)^r.$$

A moment's reflection will convince us that $\varphi_X = (\varphi_Y)^r$, where $Y \sim GGeo(p)$; that is, X is the convolution of "r" $GGeo(p)$ random variables. Now, we know from Proposition 1.2.1 that when $r \in \mathbb{N}$ we should have $\mathrm{TVar}(X) \le (\mathrm{TVar}(Y))^r$, but we already calculated $\mathrm{TVar}(Y)$ in Section 1.1 and found it to be $\mathrm{TVar}(Y) = p/(2-p)$, therefore we actually get equality in this case.

Example 1.2.2 (The Generalized Binomial Distribution).

The random variable X has the *Generalized Binomial distribution* (denoted $X \sim GBin(n,p)$) if

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n,$$

for some $n \in \mathbb{N}$ and $p \in \mathbb{R}$. The fact that $\sum_k \mathbb{P}(X = k) = 1$ is immediately implied by the Binomial Theorem and a generalized distribution occurs when $p \notin [0,1]$, in which case

$$\mathrm{TVar}(X) = \sum_{k=0}^{n} |\mathbb{P}(X = k)| = \binom{n}{k} (|p|)^k (|p| \mp 1)^{n-k} = (2|p| \mp 1)^n,$$

as $p > 1$ or $p < 0$, respectively. The characteristic function of $X \sim GBin(n,p)$ is $\varphi_X(t) = \left(q + pe^{it}\right)^n$, where $q = 1 - p$. We see that $\varphi_X = (\varphi_Y)^n$, where $Y \sim GBern(p)$; that is, X is the convolution of "n" $GBern(p)$ random variables. Similar to before we should have $\mathrm{TVar}(X) \leq (\mathrm{TVar}(Y))^n$, but in Section 1.1 it was shown that $\mathrm{TVar}(Y) = 2|p| \mp 1$ and again equality unexpectedly holds.

To see why, notice that although $GNegB(r,p)$ was defined for $r > 0$, in fact, $\sum_k \mathbb{P}(X = k) = 1$ for any $r \in \mathbb{R}$ and $0 < p < 2$. Even more is true:

Proposition 1.2.5. *For any $r < 0$ and $0 < p < 2$,*

$$GNegB(r,p) \stackrel{\mathrm{d}}{=} GBin(|r|, 1 - p^{-1}).$$

Proof. Let $X \sim GNegB(r,p)$ with characteristic function

$$\begin{aligned} \varphi_X(t) &= \left(\frac{p}{1 - qe^{it}}\right)^r \\ &= \left[\frac{1}{p} + \left(1 - \frac{1}{p}\right) e^{it}\right]^{|r|}. \end{aligned}$$

The last expression is the characteristic function of a $GBin(|r|, 1-p^{-1})$ random variable. $\square$

We conclude from the last proposition that $GNegB$ and $GBin$ random variables are in essence only different manifestations of the same random phenomena. However, there is a critical difference: for any fixed $r < 0$, as p ranges over the entire set $0 < p < 2$ the collection of $GBin$ random variables generated is not the entire class; only those $GBin(|r|, \alpha)$ for which $\alpha < 1/2$ are included. This will play a **critical** role in Chapter 3.

Example 1.2.3.

If $X \sim GPoi(\lambda)$ then its characteristic function is $\varphi_X(t) = e^{\lambda(e^{it}-1)}$. It is immediate that if X_1 and X_2 are independent $GPoi(\lambda_1)$, $GPoi(\lambda_2)$ random variables, respectively, then

$$\varphi_{X_1+X_2}(t) = \exp\{(\lambda_1 + \lambda_2)(e^{it} - 1)\}$$

showing that $X_1 + X_2$ is consequently distributed as $GPoi(\lambda_1 + \lambda_2)$, just as in the classical case. But this shows something more. It provides us with an example that strict inequality in Proposition 1.2.1 may occur; by taking $\lambda_1 < 0$ and $\lambda_2 > 0$ with $|\lambda_1| > \lambda_2$ one obtains

$$\begin{aligned} \mathrm{TVar}(X_1 + X_2) &= e^{2|\lambda_1+\lambda_2|} = e^{2(|\lambda_1|-\lambda_2)} \\ &< e^{2|\lambda_1|} \cdot 1 = \mathrm{TVar}(X_1) \cdot \mathrm{TVar}(X_2). \end{aligned}$$

Two common features of the previous examples are that all random variables so far have been discrete and their probabilities have been signed. It is natural to go further and look for generalized continuous distributions; at the same time another natural question concerns the existence of probability distributions taking complex values. In a final example we find both: a continuous random variable whose probability density is complex.

Example 1.2.4 (The Complex Gamma Distribution).

A random variable X has a *Complex Gamma distribution*, denoted $X \sim CGamma(\alpha, \beta)$, if its probability density is of the form

$$f_X(x) = \frac{1}{\Gamma(\alpha)\,\beta^\alpha} x^{\alpha-1}\, e^{-\frac{x}{\beta}}, \quad x \in \mathbb{R},$$

for some specified complex constant $\alpha \in \mathbb{C}$ with $|\alpha| > 0$ and some real $\beta > 0$. Note that this distribution collapses to the classical $Gamma(\alpha, \beta)$ distribution when α is real.

1.3 Relative Compactness and Helly's Selection Principle

This section is devoted entirely to the modification of a classical result and its detailed proof. The classical Helly-Bray Theorem asserts that for a given class of uniformly bounded increasing functions $\{F_n\}$ on $\mathbb{R}$ there exists an increasing F such that $F_n \to F$. In Appendix B we see that Prohorov's Theorem gives the appropriate analogue for complex measures. The trouble is that for our purposes we need to know *more* about F than just *existence* in order to be able to say anything useful.

As it turns out, on the real line, by modifying Helly's Selection Principle we may select a specific F with attractive properties and we may select in such a way to ensure that the convergence is sufficiently strong. In some aspects it is exactly this section that provides the underpinning for the next two chapters that follow.

Recall from Appendix B, page 98 that μ_α converges **vaguely** to μ, denoted $\mu_\alpha \xrightarrow{v} \mu$, if $\int f \mathrm{d}\mu_\alpha \to \int f \mathrm{d}\mu$ for all $f \in C_0(\Omega)$.

Theorem 1.3.1. *If $\{\mu_n\} \subset M(\mathbb{R})$ is bounded and tight then there exists a complex measure $\mu \in M(\mathbb{R})$ and a subsequence (n_j) such that $\mu_{n_j} \xrightarrow{v} \mu$. Further, if $\{\mu_n\}$ are signed then by choosing wisely we may select the subsequence so that* $\limsup_j \|\mu_{n_j}\| \leq \|\mu\|$.

This is a modification of a classical result, and its proof is long. We begin by proving the following Lemma.

Lemma 1.3.1 (Widder [48]). *Let the complex constants $(a_{m,n})$ and the positive number A be such that*

$$|a_{m,n}| \leq A, \quad m, n \in \mathbb{N}.$$

Then there exist a sequence of integers $n_1 < n_2 < \cdots$ and a sequence of numbers $a_1, a_2, \ldots$ such that

$$\lim_{k\to\infty} a_{m,n_k} = a_m = \limsup_{k\to\infty} a_{m,n_k}, \quad m \in \mathbb{N}.$$

Proof. Since the sequence $(a_{1,n})_{n=1}^{\infty}$ is bounded, its limit superior, which we denote by a_1, is finite and there exists a sequence

$$n_1^1 < n_2^1 < \cdots \tag{1.3.1}$$

such that

$$\lim_{k\to\infty} a_{1,n_k^1} = a_1 = \limsup_{k\to\infty} a_{1,n_k^1}.$$

Likewise, the sequence $(a_{2,n_k^1})_{k=1}^{\infty}$ has a limit superior so we may find integers $n_1^2 < n_2^2 < \cdots$ all contained in (1.3.1) and a number a_2 such that

$$\lim_{k\to\infty} a_{2,n_k^2} = a_2 = \limsup_{k\to\infty} a_{2,n_k^2}.$$

Proceeding in this way we find for any positive integer m a set of integers $n_1^m < n_2^m < \cdots$ all included in the set $n_1^{m-1} < n_2^{m-1} < \cdots$ and a number a_m such that

$$\lim_{k\to\infty} a_{m,n_k^m} = a_m = \limsup_{k\to\infty} a_{m,n_k^m}.$$

If now we set

$$n_k = n_k^k, \quad k \in \mathbb{N},$$

it is clear that the sequence $(n_k)_{k=1}^{k\to\infty}$ is increasing and that

$$\lim_{k\to\infty} a_{m,n_k^m} = a_m = \limsup_{k\to\infty} a_{m,n_k^m}, \quad \text{for all } m \in \mathbb{N}.$$

□

Let now $\{F_n\}$ be the respective cumulative distribution functions of $\{\mu_n\}$. The complete statement of the theorem is proved only for signed measures, although it turns out even for sequences of complex measures that there exists a subsequence and a complex measure μ such that $\mu_{n_k} \xrightarrow{v} \mu$. For, if α_n and β_n are the real and imaginary parts of F_n, respectively, the sequences (α_n) and (β_n) both satisfy the hypotheses of the theorem. For a suitable sequence of integers $(n_k^{(1)})_{k=1}^{\infty}$ and a function α of bounded variation we would have

$$\limsup_{k\to\infty} \alpha_{n_k^{(1)}}(x) = \alpha(x), \quad x \in \mathbb{R}.$$

From the sequence $(n_k^{(1)})$ we could pick another $(n_k^{(2)})$, and we could determine a function β of bounded variation such that

$$\limsup_{k\to\infty} \beta_{n_k^{(2)}}(x) = \beta(x), \quad x \in \mathbb{R}.$$

Then the complex function F with real and imaginary parts α and β would be the function sought, as would be the subsequence $(n_k^{(2)})_{k=1}^{\infty}$, because the corresponding cumulative distribution functions converge pointwise (Proposition B.1).

However, it is only in the signed case that we are able to exploit the nice properties of the positive and negative variation, so that the resulting total variations behave appropriately. Due to the fact that the real and imaginary parts of a complex measure are not mutually singular, it is not clear which choice of a subsequence would result in a conveniently supported measure.

Proof of Theorem 1.3.1. To prove the theorem for the signed case we write $F_n = F_n^+ - F_n^-$, where F_n^+ and F_n^- are the positive and negative variation cumulative distribution functions. Then $|F_n| = F_n^+ + F_n^-$, and by hypothesis there exists a constant A such that

$$0 \leq |F_n|(x) \leq A, \quad \text{for all } x \text{ and } n.$$

Arrange all of the rational numbers of $\mathbb{R}$ in a sequence $(r_m)_{m=1}^{\infty}$ and then apply Lemma 1.3.1 to the set of numbers

$$a_{m,n} = F_n^+(r_m), \quad m, n \in \mathbb{N}.$$

We thus obtain a function F^+ defined at all rational points and a sequence of integers $(n_k^{(0)})$ such that

$$\lim_{k\to\infty} F^+_{n_k^{(0)}}(r_m) = F^+(r_m) = \limsup_k F^+_{n_k^{(0)}}(r_m), \quad m \in \mathbb{N}.$$

Now set for $x \in \mathbb{R}$

$$\overline{F^+}(x) = \limsup_k F^+_{n_k^{(0)}}(x),$$

and

$$\underline{F^+}(x) = \liminf_k F^+_{n_k^{(0)}}(x).$$

Then of course

$$F^+(r_m) = \overline{F^+}(r_m) = \underline{F^+}(r_m), \quad m \in \mathbb{N}.$$

Moreover, by their definition it is clear that $\overline{F^+}$ and $\underline{F^+}$ are nondecreasing functions. Their common points of continuity are consequently dense in $\mathbb{R}$. Let t be such a point. Then $\overline{F^+}(t) = \underline{F^+}(t)$, since one can find a sequence of rationals converging to t and $\overline{F^+} = \underline{F^+}$ for all those rationals; we therefore define

$$F^+(t) = \overline{F^+}(t) = \underline{F^+}(t), \quad \text{for } t \text{ a continuity pt. of } \overline{F^+} \text{ and } \underline{F^+}.$$

Thus F^+ has been defined at all but a countable set of points $(z_k)_{k=1}^{\infty}$, the points of discontinuity of either $\overline{F^+}$ or $\underline{F^+}$. Apply Lemma 1.3.1 to the set of numbers

$$a_{m,k} = F^+_{n_k^{(0)}}(z_m), \quad m, k \in \mathbb{N}$$

to obtain a set of integers $n_1^{(1)} < n_2^{(1)} < \cdots$ all included in $(n_k^{(0)})_{k=1}^\infty$ and a sequence $(a_m)_{m=1}^\infty$ such that

$$\lim_{k\to\infty} F^+_{n_k^{(1)}}(z_m) = a_m = \limsup_k F^+_{n_k^{(1)}}(z_m), \quad m \in \mathbb{N}.$$

Complete the definition of F^+ by the equations

$$F^+(z_m) = a_m, \quad m \in \mathbb{N}.$$

We then see that

$$\lim_{k\to\infty} F^+_{n_k^{(1)}}(x) = F^+(x) = \limsup_k F^+_{n_k^{(1)}}(x), \quad x \in \mathbb{R},$$

that F^+ is nondecreasing on $\mathbb{R}$, and that $|F^+(x)| \le A$, showing that F^+ is of bounded variation.

We can use an identical argument to find a nondecreasing function F^- of bounded variation satisfying $|F^-(x)| \le A$ and a sub-subsequence (n_k) included within $(n_k^{(1)})$ such that

$$\lim_{k\to\infty} F^-_{n_k}(x) = F^-(x) = \limsup_k F^-_{n_k}(x), \quad x \in \mathbb{R}.$$

Now set $F = F^+ - F^-$. Clearly, F is of bounded variation, satisfies $0 \le |F|(x) \le 2A$ for all x, and also

$$\lim_{k\to\infty} F_{n_k}(x) = F(x) = \limsup_k F_{n_k}(x), \quad x \in \mathbb{R}.$$

Then Proposition B.1 shows that $F_{n_k} \xrightarrow{v} F$, or in other words $\mu_{n_k} \xrightarrow{v} \mu$. We finally go to show

$$\limsup_{k\to\infty} \|\mu_{n_k}\| \le \|\mu\|.$$

Let $\epsilon > 0$. By tightness, there exists a compact set K such that $|\mu_{n_k}|(K^c) < \epsilon$ for all k, in which case $\mu^\pm_{n_k}(K^c) < \epsilon$ as well. Let $r_m \in K^c$ be rational. Then

$$F^+_{n_k}(r_m) \ge F^+_{n_k}(\infty) - \epsilon, \quad \text{for all } k.$$

Now take $\limsup_k$ on both sides to get

$$F^+(r_m) = \limsup_{k\to\infty} F^+_{n_k}(r_m) \geq \limsup_{k\to\infty} \|\mu^+_{n_k}\| - \epsilon,$$

which yields $\|\mu^+\| = F^+(\infty) \geq F^+(r_m) \geq \limsup_{k\to\infty} \|\mu^+_{n_k}\| - \epsilon$, that is,

$$\|\mu^+\| \geq \limsup_{k\to\infty} \|\mu^+_{n_k}\| - \epsilon.$$

Since ϵ can be made arbitrarily small we get

$$\|\mu^+\| \geq \limsup_{k\to\infty} \|\mu^+_{n_k}\|,$$

and by a similar argument we also get

$$\|\mu^-\| \geq \limsup_{k\to\infty} \|\mu^-_{n_k}\|.$$

But then (by boundedness of $\|\mu^\pm_{n_k}\|$)

$$\begin{aligned}
\limsup_{k\to\infty} \|\mu_{n_k}\| &= \limsup_k \left(\|\mu^+_{n_k}\| + \|\mu^-_{n_k}\|\right) \\
&\leq \limsup_k \|\mu^+_{n_k}\| + \limsup_k \|\mu^-_{n_k}\| \\
&\leq \|\mu^+\| + \|\mu^-\| \quad \text{(from above)} \\
&= \|\mu\|.
\end{aligned}$$

□

CHAPTER 2
GENERALIZED INFINITE DIVISIBILITY

2.1 Definitions and Examples

Definition 2.1.1. *A probability distribution ν is said to be **generalized infinitely divisible** (GID) if for each $n \in \mathbb{N}$ there exists a normalized signed distribution $\nu^{(n)} \in M_{\pm}(\mathbb{R})$ such that*

$$\nu = \underbrace{\nu^{(n)} * \nu^{(n)} * \cdots * \nu^{(n)}}_{n\ factors}.$$

Equivalently, $\widehat{\nu}$ is GID if for each $n \in \mathbb{N}$ there exists a normalized signed measure $\nu^{(n)} \in M_{\pm}$ with Fourier-Stieltjes Transform $\widehat{\nu^{(n)}}$ such that

$$\widehat{\nu} = (\widehat{\nu^{(n)}})^n.$$

In terms of random variables this means, for each $n \geq 1$, in a suitable finite signed measure space there exist random variables X and $X_{n,j}$, $1 \leq j \leq n$, such that X has characteristic function $\widehat{\nu}$, $X_{n,j}$ has characteristic function $\widehat{\nu^{(n)}}$, and

$$X = \sum_{j=1}^{n} X_{n,j}.$$

X is thus "divisible" into n independent and identically distributed parts, for each $n \in \mathbb{N}$.

We may speak of a distribution, a Fourier-Stieltjes transform, or a random variable being GID interchangeably. The facet of GID to which we are referring will be stated explicitly in places where it is not immediately clear from the context.

Notation. *If ν is GID and $n \in \mathbb{N}$ we write $X^{(n)}$ for the n^{th} factor of X, i.e. the random variable distributed according to the signed measure $\nu^{(n)}$. In this notation, X is*

GID if and only if for each $n \in \mathbb{N}$ one can write $X = X_1^{(n)} + X_2^{(n)} + \cdots + X_n^{(n)}$, where $X_i^{(n)}$, $i = 1, 2, \ldots, n$ are i.i.d. according to some normalized signed distribution $\nu^{(n)}$.

Example 2.1.1 **(**$GPoi(\lambda)$**).**

We saw in Chapter 1, Example 1.2.3 that for a random variable $X \sim GPoi(\lambda)$ it suffices to take $X^{(n)} \sim GPoi(\lambda/n)$, which satisfies $\|\nu^{(n)}\| = e^{2|\lambda|/n} < \infty$.

Example 2.1.2 **(**$GNegB(r,p)$**).**

Notice for all $r > 0$ and $0 < p < 2$ we may take $X^{(n)} \sim GNegB(r/n, p)$. The classical ID case is $0 < p < 1$. When $1 \le p < 2$, it follows from our calculations in Chapter 1 that $\|\nu^{(n)}\| = [p/(2-p)]^{r/n} < \infty$.

Example 2.1.3 **(**$GBern(q),\ q \le 1/2$**).**

It is easy to see that $X \sim GBern(1/2)$ has characteristic function $\phi(t) = (1 + e^{it})/2$ with n^{th} root

$$[\phi(t)]^{1/n} = 2^{-1/n} \sum_{k=0}^{\infty} \binom{1/n}{k} e^{itk}.$$

The distribution $\{\, 2^{-1/n}\binom{1/n}{k} : k = 0, 1, \ldots \}$ is absolutely summable with sum 1. For the remainder we may use the work that we did in Chapter 1, Section 1.2. Due to Proposition 1.2.5, when $q < 1/2$ this is merely the claim that $GNegB(r,p)$ is GID for $-1 < r < 0$ and $0 < p < 2$. Notice first that $\binom{-r}{0} = 1$ and

$$\binom{-r}{k} \quad \text{is} \quad \begin{cases} > 0, & \text{if } k \text{ is odd,} \\ < 0, & \text{if } k \text{ is even.} \end{cases}$$

Case 1. $0 < p < 1$. Then $q = 1 - p > 0$ and $\mathbb{P}(X = 0) = \binom{-r}{0}(-q)^0 p^r > 0$ with

$$\mathbb{P}(X = k) = \binom{-r}{k} (-q)^k p^r \quad \text{being} \quad \begin{cases} < 0, & \text{if } k \text{ is odd,} \\ < 0, & \text{if } k \text{ is even.} \end{cases}$$

Therefore $|\mathbb{P}(X=k)| = -\mathbb{P}(X=k)$ for $k \geq 1$. We may now say that when $-1 < r < 0$ and $0 < p < 1$:

$$\begin{aligned}
\text{TVar}(GNegB(r,p)) &= \sum_{k=0}^{\infty} |\mathbb{P}(X=k)| \\
&= \mathbb{P}(X=0) - \sum_{k=1}^{\infty} \binom{-r}{k} (-q)^k p^r \\
&= p^r - \left(\sum_{k=0}^{\infty} \binom{-r}{k} (-q)^k p^r - p^r \right) \\
&= 2p^r - 1.
\end{aligned}$$

This shows that $\|\nu^{(n)}\| = 2p^{r/n} - 1 < \infty$.

Case 2. $1 \leq p < 2$. Then $q \leq 0$, $\mathbb{P}(X=0) > 0$, and

$$\mathbb{P}(X=k) = \binom{-r}{k} (-q)^k p^r \quad \text{is} \quad \begin{cases} > 0, & \text{if } k \text{ is odd,} \\ < 0, & \text{if } k \text{ is even.} \end{cases}$$

Therefore

$$|\mathbb{P}(X=k)| = \begin{cases} \mathbb{P}(X=k), & \text{if } k \text{ is odd,} \\ -\mathbb{P}(X=k), & \text{if } k \text{ is even.} \end{cases}$$

We may now say that when $-1 < r < 0$ and $1 \leq p < 2$ we have

$$\begin{aligned}
\text{TVar}(GNegB(r,p)) &= \sum_{k=0}^{\infty} |\mathbb{P}(X=k)| \\
&= \mathbb{P}(X=0) + \sum_{k=1}^{\infty} \mathbb{P}(X=k)(-1)^{k+1} \\
&= p^r - p^r \sum_{k=1}^{\infty} \binom{-r}{k} q^k \\
&= p^r - p^r \left[\frac{1}{(1+q)^r} - 1 \right] \\
&= 2p^r - \left(\frac{p}{2-p} \right)^r
\end{aligned}$$

This shows

$$\|\nu^{(n)}\| = 2p^{r/n} - \left(\frac{p}{2-p} \right)^{r/n} < \infty.$$

We therefore conclude that the class $GBern(q),\ q \leq 1/2$ is GID, as claimed.

2.2 The Uniform(0,1) Distribution

In contrast to the previous section where all GID random variables were discrete, it is in this section that a purely continuous GID random variable is encountered for the first time.

Theorem 2.2.1. *The $Uniform(0,1)$ distribution is GID.*

It is well known (see O'Brien and Steutel [38] or Sato [42]) that the *Unif(0,1)* distribution is not classically ID, a fact perhaps most quickly implied by the random variables bounded support. Thus any factor of the uniform is necessarily signed.

We will motivate its divisibility merely by displaying the characteristic function of its n^{th} factor. Since essentially the same calculation works for the more general α^{th} factor, where $\alpha > 1$ is real, that is what is shown below.

$$\begin{aligned}
\left(\int e^{itx}\, 1_{(0,1)}(x)\,\mathrm{d}x\right)^{1/\alpha} &= \left(\frac{e^{it}-1}{it}\right)^{1/\alpha} \\
&= \left(\frac{1}{-it}\right)^{1/\alpha} \sum_{k=0}^{\infty} \binom{1/\alpha}{k} \left(-e^{it}\right)^k \\
&= \sum_{k=0}^{\infty} \binom{1/\alpha}{k} (-1)^k e^{itk} \int_0^{\infty} \frac{1}{\Gamma(1/\alpha)} u^{1/\alpha-1} e^{\frac{-u}{(1/-it)}}\,\mathrm{d}u \\
&= \sum_{k=0}^{\infty} \binom{1/\alpha}{k} (-1)^k e^{itk} \int_0^{\infty} e^{itu} \frac{u^{1/\alpha-1}}{\Gamma(1/\alpha)}\,\mathrm{d}u \\
&= \sum_{k=0}^{\infty} \binom{1/\alpha}{k} (-1)^k \int_k^{\infty} e^{itx} \frac{(x-k)^{1/\alpha-1}}{\Gamma(1/\alpha)}\,\mathrm{d}x \\
&= \int_0^{\infty} e^{itx} \sum_{k=0}^{\infty} \binom{1/\alpha}{k} (-1)^k \frac{(x-k)^{1/\alpha-1}}{\Gamma(1/\alpha)} 1_{(k,\infty)}(x)\,\mathrm{d}x \\
&= \int_0^{\infty} e^{itx} f_\alpha(x)\,\mathrm{d}x,
\end{aligned}$$

where

$$f_\alpha(x) = \sum_{k=0}^{\infty} \binom{1/\alpha}{k} (-1)^k \frac{1}{\Gamma(1/\alpha)} (x-k)^{1/\alpha - 1} 1_{(k,\infty)}(x), \quad x > 0.$$

We have thus found the p.d.f. of the α^{th} factor of the *Unif(0,1)* distribution. It is not even slightly obvious that $f_\alpha \in L^1$ but, remarkably, it is. This is what is proved next.

Theorem 2.2.2 (G.J.Székely). *For all $a > 0$,*

$$\Gamma(a) u_a(x) = \sum_{k=0}^{\lfloor x \rfloor} (-1)^k \binom{a}{k} (x-k)^{a-1}, \quad x > 0,$$

is in L^1.

We will need several auxiliary lemmas before we are able to prove this theorem. Throughout we will use the following conventions: for $0 < a \leq 1/2$ let

$$a_k = (-1)^{k-1} \binom{a}{k}, \quad b_k = (-1)^k \binom{-a}{k}, \quad k = 0, 1, 2, \ldots$$

The first lemma concerns finding the asymptotic behavior of the above defined constants a_k and b_k.

Lemma 2.2.1.

$$a_k = \frac{a\,\Gamma(k-a)}{k!\,\Gamma(1-a)} = \frac{a}{\Gamma(1-a)} \frac{1}{k^{a+1}} \left(1 + O(1/k)\right),$$

and

$$b_k = \frac{\Gamma(k+a)}{k!\,\Gamma(a)} = \frac{1}{\Gamma(a)} \frac{1}{k^{1-a}} \left(1 + O(1/k)\right).$$

Proof. We prove the lemma for the case of a_k, the case for b_k being remarkably similar. The first equality is true because

$$(-1)^{k-1} \binom{a}{k} = \frac{(-1)^{k-1}\Gamma(a+1)}{k!\Gamma(a-k+1)} = \frac{a}{k!} \left(\frac{(-1)^{k-1}\Gamma(a)}{\Gamma(a-(k-1))} \right)$$

and

$$\begin{aligned}\frac{(-1)^{k-1}\Gamma(a)}{\Gamma(a-(k-1))} &= (-1)^{k-1}(a-1)(a-2)\cdots(a-(k-1)) \\ &= (k-1-a)(k-2-a)\cdots(1-a)\cdot\frac{\Gamma(1-a)}{\Gamma(1-a)} \\ &= \frac{\Gamma(k-a)}{\Gamma(1-a)}.\end{aligned}$$

The second equality is true from the fact that

$$\frac{\Gamma(k-a)}{\Gamma(k+1)} = \frac{1}{k^{a+1}}\left(1+O(1/k)\right),$$

which is evident from the representation

$$\Gamma(n+1) = \left(\frac{n}{e}\right)^n \sqrt{2\pi n}\, e^{\mu(n)}, \quad \text{where } \tfrac{1}{12n+1} < \mu(n) < \tfrac{1}{12n},$$

(see Feller [22], González [28], or Robbins [40]) since

$$\begin{aligned}\frac{\Gamma(k-a)}{\Gamma(k+1)} &= \frac{\left(\frac{k-a-1}{e}\right)^{k-a-1}\sqrt{2\pi(k-a-1)}\, e^{\mu(k-a-1)}}{\left(\frac{k}{e}\right)^k \sqrt{2\pi k}\, e^{\mu(k)}} \\ &\le \frac{(k-a-1)^{k-a-1/2}}{k^{k+1/2}}\ \frac{e^k}{e^{k-a-1}}\ \exp\left\{\frac{1}{12(k-a-1)} - \frac{1}{12k+1}\right\} \\ &= \left(1-\frac{a+1}{k}\right)^k \frac{e^{a+1}}{k^{1/2}\,(k-a-1)^{a+1/2}}\ \exp\left\{\frac{12(a+1)+1}{12^2(k^2-(a+1)k)}\right\} \\ &\le \frac{1}{k^{1/2}\,(k-a-1)^{a+1/2}}\ \exp\left\{\frac{12(a+1)+1}{12^2(k^2-(a+1)k)}\right\}.\end{aligned}$$

The equality now follows because for all $c, d > 0$,

$$\exp\left\{\frac{c}{k^2-dk}\right\} = 1 + o(1/k).$$

□

As a next step we find an identity for the convolution of the a_k and b_k sequences.

Lemma 2.2.2.

$$\sum_{k=0}^{n} a_k b_{n+1-k} = b_{n+1} - a_{n+1}, \quad n \ge 0.$$

Proof. Consider the functions $G_{\pm a}(z) = (1-z)^{\pm a}$ on $|z| < 1$, which are analytic and have the power series representation

$$G_a(z) = -\sum_{k=0}^{\infty} a_k\, z^k, \quad G_{-a}(z) = \sum_{k=0}^{\infty} b_k\, z^k.$$

Furthermore, notice that there exist coefficients (c_k) such that the power series representation

$$G_a(z) \cdot G_{-a}(z) = \sum_{k=0}^{\infty} c_k\, z^k = -(1-z)^a(1-z)^{-a} = -1$$

holds for any $|z| < 1$. This of course implies that for all $n \geq 0$ the coefficients c_{n+1} are zero, which by the Cauchy rule for multiplying series means

$$c_{n+1} := \sum_{k=0}^{n+1} a_k b_{n+1-k} = \sum_{k=0}^{n} a_k b_{n+1-k} + (-1)b_{n+1} + a_{n+1}(1) = 0,$$

and thus

$$\sum_{k=0}^{n} a_k b_{n+1-k} = b_{n+1} - a_{n+1}, \quad n \geq 0.$$

□

Last we investigate the asymptotic formulas concerning the convolution of the a_k sequence wit the more tractable sequence $(n+1-k)^{a-1}$.

Lemma 2.2.3.

$$\begin{aligned} \sum_{k=1}^{n} a_k(n+1-k)^{a-1} &= \Gamma(a)\sum_{k=1}^{n} a_k b_{n+1-k} + O\left(\sum_{k=1}^{n} \frac{1}{k^{a+1}(n+1-k)^{2-a}}\right) \\ &= \Gamma(a)(b_{n+1} - a_{n+1}) + O\left(n^{-a-1} \vee n^{a-2}\right) \\ &= (n+1)^{a-1} + O\left(n^{-a-1} \vee n^{a-2}\right). \end{aligned}$$

Proof. First we show

$$a_k(n+1-k)^{a-1} = \Gamma(a) a_k b_{n+1-k} + O\left(\frac{1}{k^{a+1}(n+1-k)^{2-a}}\right).$$

This follows from Lemma 2.2.1 because then $a_k[(n+1-k)^{a-1} - \Gamma(a)b_{n+1-k}]$ is

$$\begin{aligned}
&= a_k \left[(n+1-k)^{a-1} - \left(\frac{1}{(n+1-k)^{1-a}} (1 + O(\frac{1}{n+1-k})) \right) \right] \\
&= \frac{1}{k^{1+a}} \left(1 + O(\frac{1}{k}) \right) \left[\; O(\frac{1}{(n+1-k)^{2-a}}) \; \right] \\
&= O\left(\frac{1}{k^{1+a}(n+1-k)^{2-a}} \right) + O\left(\frac{1}{k^{2+a}(n+1-k)^{2-a}} \right) \\
&= O\left(\frac{1}{k^{1+a}(n+1-k)^{2-a}} \right).
\end{aligned}$$

Now sum over k to get

$$\sum_{k=1}^{n} a_k (n+1-k)^{a-1} = \Gamma(a) \sum_{k=1}^{n} a_k b_{n+1-k} + O\left(\sum_{k=1}^{n} \frac{1}{k^{a+1}(n+1-k)^{2-a}} \right).$$

Then Lemma 2.2.2 gives that the right hand side above is just

$$\Gamma(a)(b_{n+1} - a_{n+1}) + O\left(\sum_{k=1}^{n} \frac{1}{k^{a+1}(n+1-k)^{2-a}} \right).$$

But in fact,

$$\sum_{k=1}^{n} \frac{1}{k^{a+1}} \frac{1}{(n+1-k)^{2-a}} = O(n^{-a-1}).$$

To see this consider the related integral

$$\int_1^n \frac{1}{x^{1+a}} \frac{1}{(n+1-x)^{2-a}} \mathrm{d}x$$

which after some algebra and a variable change is

$$\frac{\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} u^{-a-1}(1-u)^{(a-1)-1} \mathrm{d}u}{(n+1)^2}.$$

This last expression is indeterminate of the form ∞/∞ because the numerator is a truncated Beta integral $\int x^{\alpha-1}(1-x)^{\beta-1}\mathrm{d}x$ which diverges if $\alpha < 0,\ \beta < 0$. Then L'Hospital's Rule shows that the limit as $n \to \infty$ of the above is

$$\lim_{n\to\infty} \frac{\frac{\mathrm{d}}{\mathrm{d}n} \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} u^{-a-1}(1-u)^{(a-1)-1} \mathrm{d}u}{2(n+1)},$$

and it suffices to show that the derivative in the numerator is $O(n^{-a})$. This follows perhaps most quickly from Leibnitz' Rule A.3.3, which gives the derivative to be

$$\left[\left(\frac{n}{n+1}\right)^{-(a+1)}\left(1-\frac{n}{n+1}\right)^{a-2}-\left(\frac{1}{n+1}\right)^{-(a+1)}\left(1-\frac{n}{n+1}\right)^{a-2}\cdot(-1)\right]\frac{1}{(n+1)^2}$$

$$=\frac{n+1}{n^{1+a}}-\frac{n+1}{n^{2-a}}=(n+1)\left(\frac{n^{1-2a}-1}{n^{2-a}}\right)=O(n^{-a}).$$

We may therefore write the displayed quantity of interest above as being

$$=\Gamma(a)(b_{n+1}-a_{n+1})+O\left(\,n^{-a-1}\vee n^{a-2}\,\right),$$

and we claim that this last expression is $(n+1)^{a-1}+O(n^{-a-1}\vee n^{a-2})$. This is easily seen by using Lemma 2.2.1 and calculating $\Gamma(a)(b_{n+1}-a_{n+1})$ to be

$$\begin{aligned}
&= \Gamma(a)\left(\frac{1}{\Gamma(a)(n+1)^{1-a}}-\frac{a}{\Gamma(1-a)(n+1)^{1+a}}\right)\left(1+O(\frac{1}{n+1})\right)\\
&\propto \frac{1}{(n+1)^{1-a}}+O\left(\frac{1}{(n+1)^{2-a}}\right)+\frac{1}{(n+1)^{a+1}}+O\left(\frac{1}{(n+1)^{a+2}}\right).
\end{aligned}$$

Looking above we see that the second and fourth terms are $O(n^{a-2})$, and the third is $O(n^{-a-1})$ as claimed; the lemma is proved. □

With all of the technical machinery established in the last three lemmas it is now possible to achieve the ultimate goal of showing that the density of the Uniform(0,1)'s factors is in fact integrable. We now go to prove Theorem 2.2.1.

Proof of the Theorem. It is sufficient to prove the theorem for $0<a\le 1/2$. In this case

$$\Gamma(a)u_a(x)=x^{a-1}-\sum_{k=1}^{\lfloor x\rfloor}a_k(x-k)^{a-1}\,,\quad x>0,$$

where a_k is as defined before Lemma 2.2.1, with

$$\sum_{k=1}^{\infty}a_k=\sum_{k=1}^{\infty}\binom{a}{k}(-1)^{k-1}=-[\,(1-1)^a-1]=1.$$

Although $\Gamma(a)u_a(x)$ was already split into the difference of two positive functions, the terms were not in L^1. We need one more step. Let $n < x < n+1$, i.e. $n = \lfloor x \rfloor$. Then $\Gamma(a)u_a(x) = g(x) - h(x)$, where

$$g(x) = x^{a-1}\left(1 - \sum_{k=1}^{n} a_k \left(\frac{n+1-k}{n+1}\right)^{a-1}\right),$$
$$h(x) = \sum_{k=1}^{n} a_k \left((x-k)^{a-1} - \left(\frac{n+1-k}{n+1}\,x\right)^{a-1}\right).$$

These functions turn out to be in L^1. (They are also nonnegative although this is not quite clear in case of $g(x)$, and we are not going to use this property.)

On the one hand,

$$\begin{aligned}
0 &\le (x-k)^{a-1} - \left(\frac{n+1-k}{n+1}\,x\right)^{a-1} \\
&\le (1-a)\left[\frac{n+1-k}{n+1}\,x - (x-k)\right](x-k)^{a-2} \\
&\le \frac{k(n+1-x)}{n+1}\,(x-k)^{a-2} \le (x-k)^{a-2}\,,
\end{aligned}$$

hence

$$(x-k)^{a-1} - \left(\frac{n+1-k}{n+1}\,x\right)^{a-1} \le (x-k)^{a-1} \wedge (x-k)^{a-2}.$$

Thus

$$\begin{aligned}
\int_0^\infty h(x)\,\mathrm{d}x &= \sum_{n=1}^{\infty}\int_n^{n+1}\sum_{k=1}^{n} a_k\left((x-k)^{a-1} - \left(\frac{n+1-k}{n+1}\,x\right)^{a-1}\right)\mathrm{d}x \\
&= \sum_{k=1}^{\infty} a_k \sum_{n=k}^{\infty}\int_n^{n+1}\left((x-k)^{a-1} - \left(\frac{n+1-k}{n+1}\,x\right)^{a-1}\right)\mathrm{d}x,
\end{aligned}$$

where the interchange of the sums is justified because all of the terms are nonnegative.

Continuing,

$$\begin{aligned}
&= \sum_{k=1}^{n} a_k \int_k^\infty \left((x-k)^{a-1} - \left(\frac{n+1-k}{n+1}\, x \right)^{a-1} \right) \mathrm{d}x \\
&\leq \sum_{k=1}^{n} a_k \int_k^\infty (x-k)^{a-1} \wedge (x-k)^{a-2} \,\mathrm{d}x \\
&= \left(\sum_{k=1}^{n} a_k \right) \int_0^\infty u^{a-1} \wedge u^{a-2} \,\mathrm{d}u \\
&= 1 \cdot \int_0^1 u^{a-1} \,\mathrm{d}u + \int_1^\infty u^{a-2} \,\mathrm{d}u < \infty,
\end{aligned}$$

and therefore $h \in L^1$. On the other hand, we define b_k as above and use Lemmas 2.2.1, 2.2.2, and 2.2.3 to get the equality

$$\sum_{k=1}^{n} a_k (n+1-k)^{a-1} = (n+1)^{a-1} + O\left(n^{-a-1} \vee n^{a-2}\right).$$

This means

$$\int_0^\infty |g(x)| \,\mathrm{d}x = \sum_{n=0}^{\infty} \int_n^{n+1} x^{a-1} \left| 1 - \sum_{k=1}^{n} a_k \left(\frac{n+1-k}{n+1} \right)^{a-1} \right| \mathrm{d}x,$$

and

$$\begin{aligned}
\left| 1 - \sum_{k=1}^{n} a_k \left(\frac{n+1-k}{n+1} \right)^{a-1} \right| &= \left| 1 - \frac{1}{(n+1)^a} \sum_{k=1}^{n} a_k (n+1-k)^{a-1} \right| \\
&= \left| 1 - \frac{1}{(n+1)^a} \left[(n+1)^{a+1} + O(n^{-a-1} \vee n^{a-2}) \right] \right| \\
&= O\left(\frac{n^{-a-1}}{(n+1)^{a-1}} \vee \frac{n^{a-2}}{(n+1)^{a-1}} \right) \\
&= O(n^{-2a} \vee n^{-1}).
\end{aligned}$$

Finally

$$\begin{aligned}
\int_0^\infty |g(x)|\,\mathrm{d}x &\le \sum_{n=0}^{\infty}\int_n^{n+1} x^{a-1}\; O(n^{-2a}\vee n^{-1})\,\mathrm{d}x \\
&= \sum_{n=0}^{\infty} O(n^{-2a}\vee n^{-1})\int_n^{n+1} x^{a-1}\,\mathrm{d}x \\
&\le \sum_{n=0}^{\infty} O(n^{-2a}\vee n^{-1})\cdot n^{a-1} \\
&= \sum_{n=0}^{\infty} O(n^{-(1+a)}\vee n^{-2+a}) < \infty.
\end{aligned}$$

Therefore $u_a = h + g$ is in L^1 as desired. $\square$

As a result of the last theorem we see that for each k there exists an $f_k \in L^1$ such that

$$\left(\int e^{itx} 1_{(0,1)}(x)\,\mathrm{d}x\right)^{1/k} = \int e^{itx} f_k(x)\,\mathrm{d}x$$

and we conclude that the Uniform(0,1) distribution is *GID*.

2.3 Székely's Discrete Convex Theorem

We saw in the last section that the *Unif(0,1)* distribution is *GID*. And this fact is in no way isolated; there are very many *GID* distributions, as the next theorem due to Gábor J. Székely reveals. The interpretation of this theorem is quite startling: in fact, *all* discrete distributions with nonincreasing probabilities on the integers $0, 1, \ldots, N$ for some $N < \infty$ are *GID*!

Theorem 2.3.1 (G. J. Székely[44]). *If $p_0 \ge p_1 \ge \cdots \ge p_N > 0$ satisfy $\sum_{j=0}^{N} p_j = 1$, then for each $k \in \mathbb{N}$ there exists a sequence $(a_j^{(k)})_{j=0}^{\infty}$ of real numbers with*

$$\sum_{j=0}^{\infty} |a_j^{(k)}| < \infty \quad \textit{and} \quad \sum_{j=0}^{\infty} a_j^{(k)} = 1$$

such that

$$\left(\sum_{j=0}^{N} p_j\, e^{itj}\right)^{1/k} = \sum_{j=0}^{\infty} a_j^{(k)}\, e^{itj}, \quad t \in \mathbb{R}.$$

Proof. Let $G(z) = \sum_{j=0}^{N} p_j z^j$. If $|z| \le 1$, then

$$|(1-z)G(z)| = |p_0 - \sum_{j=1}^{N}(p_{j-1} - p_j)z^j - p_N z^{N+1}| \ge$$

$$p_0 - \left|\sum_{j=0}^{N}(p_j - p_{j+1})|z|^j + p_N|z|^{N+1}\right| \ge p_0 - |\sum_{j=0}^{N}(p_j - p_{j+1}) + p_N| = 0,$$

with equality iff $|z| = 1$ and the points $(p_j - p_{j+1})z^j$, $j = 0, 1, \ldots, N-1$, $p_N z^{N+1}$ are on the real line. If the greatest common divisor of the j's for which $p_{j-1} - p_j \ne 0$ is d, then

$$G(z) = (1 + z + \cdots + z^{d-1})G^*(z^d),$$

where every zero of G^* has modulus strictly bigger than 1. Thus the radius of convergence of G^*'s associated power series is strictly bigger than 1, hence $[G^*(z)]^{1/k}$ is absolutely convergent at $z = 1$ for $k = 2, 3, \ldots$. What remains to be proved is that $(1 + z + \cdots + z^{d-1})^{1/k} = (1 - z^d)^{1/k}(1-z)^{-1/k}$ is also absolutely convergent for $z = 1$. This follows from the observation that $\sum_{j=0}^{\infty} |\binom{\alpha}{j}| < \infty$ for all $0 < |\alpha| < 1$, e.g. for $\alpha = \pm 1/2, \pm 1/3, \ldots$. $\square$

We see therefore that every nonincreasing probability distribution with bounded support is GID. However, the assumption of bounded support, while not obviously necessary, turns out to be quite important. In the next chapter this problem will be more closely examined.

2.4 Denseness of GID

According to Theorem 2.3.1, every nonincreasing discrete probability distribution on a finite subset of the nonnegative integers is GID. The next question would be whether there is an analogue for the continuous distributions, and in fact, there is. However, the analogy is not exact, to be made precise below.

We denote by $\mathcal{C}_+$ the set of all probability distributions on $\mathbb{R}_+$ that are **concave**, that is, all cumulative distribution functions F for which

$$F(p\,x + (1-p)\,y) \geq p\,F(x) + (1-p)\,F(y), \quad \text{for all } x, y \geq 0 \text{ and } 0 < p < 1.$$

Proposition 2.4.1. *The class GID is dense in $\mathcal{C}_+$ under the topology of total variation, and hence under the weak topology. That is, given $\epsilon > 0$ and a distribution function $F \in \mathcal{C}_+$ there exists G that is GID such that $\|F - G\| < \epsilon$.*

Proof. Given a distribution function $F \in \mathcal{C}_+$ on $[0, \infty)$, define $f : (0, \infty) \to \mathbb{R}$ by

$$f(x) = \frac{F'_-(x) + F'_+(x)}{2}, \quad x > 0,$$

where $F'_\pm$ is the right (resp. left) limiting difference quotient. Note that from concavity of F follows:

1. F is continuous except possibly at $x = 0$,
2. f is defined and finite at each point,
3. $f \geq 0$ is nonincreasing,
4. f has at most countably many points of discontinuity.

Together these considerations imply that f is Riemann integrable on $[0, M]$ for any M.

Given $\epsilon > 0$ select $M \in \mathbb{N}$ such that $F(M) > 1 - \epsilon/2$, and let $\delta > 0$ be so small that for any partition P of $[0, M]$ with $\mathsf{mesh}(P) < \delta$ we have

$$|S(f, P) - \int_0^M f(x)\,\mathrm{d}x| < \epsilon/2$$

where $S(f, P)$ is a Riemann sum of f with respect to the partition P. Lastly, let $m \in \mathbb{N}$ be such that $1/m < \delta$ and set $N = M \cdot m$.

Define

$$p_k = \begin{cases} f\left(\frac{k+1}{m}\right), & \text{if } k = 0, 1, \ldots, N-1, \\ \\ 0, & \text{if } k = N, N+1, \ldots. \end{cases}$$

and

$$g(x) = \sum_{k=0}^{N-1} p_k \, 1_{(\frac{k}{m}, \frac{k+1}{m}]}(x), \quad x \geq 0.$$

Then $\int |f - g| < \epsilon$. Furthermore,

$$\begin{aligned} \int e^{itx} g(x) \, \mathrm{d}x &= \sum_{k=0}^{N-1} p_k \int_{\frac{k}{m}}^{\frac{k+1}{m}} e^{itx} \, \mathrm{d}x \\ &= \sum_{k=0}^{N-1} p_k \frac{1}{it} \left(e^{it\frac{k+1}{m}} - e^{it\frac{k}{m}} \right) \\ &= \frac{e^{i\frac{t}{m}} - 1}{i(t/m)} \sum_{k=0}^{N-1} \frac{p_k}{m} e^{i\frac{t}{m}k} \\ &= \varphi_U \left(\frac{t}{m} \right) \varphi_D \left(\frac{t}{m} \right), \end{aligned}$$

where φ_U is the characteristic function of a *Unif(0,1)* distribution and φ_D is the characteristic function of a discrete measure placing mass $p_0 \geq p_1 \geq \cdots \geq p_{N-1} > 0$ at $k = 0, 1, \ldots, N-1$, respectively. But we know that φ_U is GID from Theorem 2.2.1 and we know φ_D is GID from Theorem 2.3.1; and g is merely the convolution of these two GID random variables, hence also GID.

The only remaining detail is to note that $\int |f - g| = 2\|F - G\|$ by Scheffé's Lemma A.2.1, and we conclude that GID is dense in $\mathcal{C}_+$. $\square$

CHAPTER 3
STRONG INFINITE DIVISIBILITY

3.1 Definitions, Examples, and First Properties

The definition of the class GID is relaxed enough to allow all sorts of unexpected and pathological behavior from its members. The trouble arises from the lack of structure in a GID random variable's infinitesimal factors; their variation may increase without bound, or mass may "escape to infinity" due to the cancellation of the probabilities involved. One is thus motivated to find mild conditions to impose on its composite pieces that permit new and interesting properties, and yet preserve some of the intuition one already has about classical ID distributions. To that end we define the following subclass of GID:

Definition 3.1.1. *We say that a GID normalized signed distribution, characteristic function, or random variable $X \sim \nu$ is **strongly infinitely divisible** (SID) if there exists $\delta > 0$ such that*

$$\sup_n |\mathbb{E}| \left(1 + |X^{(n)}|^\delta\right) < \infty.$$

*Otherwise we say X is **weakly infinitely divisible** (WID).*

As a first step in trying to qualitatively understand the properties implied by the above condition we look at some examples of SID random variables in the attempt to obtain some intuition.

Example 3.1.1 ($GPoi(\lambda)$).

In Chapter 2 it was shown that a $GPoi(\lambda)$ random variable X is GID, with $X^{(n)} \sim$

$GPoi(\lambda/n)$. That implies

$$|\mathbb{E}|1 = \|\nu^{(n)}\| = e^{2|\lambda|/n} \quad \text{and} \quad |\mathbb{E}||X^{(n)}| = e^{2|\lambda|/n}\,\frac{|\lambda|}{n}$$

whose sum is bounded in n, and thus $GPoi$ is SID.

Example 3.1.2 ($GNegB(r,p)$**).**

It is known from Chapter 2 that for all $r > 0$ and $0 < p < 2$, $X \sim GNegB(r,p)$ is an element of GID, with $X^{(n)} \sim GNegB(r/n, p)$. The classical ID case is $0 < p < 1$, where the moment condition is trivially satisfied because $\mathbb{E}X^{(n)} = rq/np$. And when $1 \le p < 2$, recall that $\|\nu^{(n)}\| = [p/(2-p)]^{r/n}$ and $X^{(n)} \overset{\text{nv}}{\sim} NegB(r/n, 2-p)$, implying

$$|\mathbb{E}||X^{(n)}| = \|\nu^{(n)}\|\,\|\mathbb{E}\||X^{(n)}| = \left(\frac{p}{2-p}\right)^{r/n} \cdot \frac{r(p-1)}{n(2-p)},$$

which is bounded in n.

Example 3.1.3 ($GBern(p),\ p < 1/2$**).**

One may now take full advantage of the work that was done in Chapters 1 and 2. From Proposition 1.2.5 it suffices to demonstrate that $GNegB(r,p)$ is SID for $-1 < r < 0$ and $0 < p < 2$; membership in GID has already been established in Chapter 2. In the spirit of the argument there, two cases will be considered.

Case 1. $0 < p < 1$. Earlier it was seen that $\|\nu^{(n)}\| = 2p^{r/n} - 1$ which is of course bounded in n. To finish, calculate

$$\begin{aligned} |\mathbb{E}||X^{(n)}| &= \sum_{k=0}^{\infty} k\,|\mathbb{P}(X^{(n)} = k)| \\ &= -\sum_{k=1}^{\infty} k \binom{-r/n}{k} (-q)^k p^r \\ &= -\mathbb{E}X^{(n)} = \frac{|r|q}{np}. \end{aligned}$$

This is also bounded in n and this case is thus SID.

Case 2. $1 \leq p < 2$. From before

$$\|\nu^{(n)}\| = 2p^{r/n} - \left(\frac{p}{2-p}\right)^{r/n}$$

which is of course bounded in n. Lastly

$$\begin{aligned}
|\mathbb{E}||X^{(n)}| &= \sum_{k=0}^{\infty} k\,|\mathbb{P}(X^{(n)} = k)| \\
&= -\frac{p^{r/n}}{(2-p)^{r/n}} \sum_{k=0}^{\infty} k \binom{-r/n}{k} (-(p-1))^k (2-p)^{r/n} \\
&= -\left(\frac{p}{2-p}\right)^{r/n} \frac{r(p-1)}{n(2-p)} \\
&= \left(\frac{p}{2-p}\right)^{r/n} \frac{|r|(p-1)}{n(2-p)},
\end{aligned}$$

and because this is bounded in n, the class $GBern(p),\ p < 1/2$ is SID, as claimed.

It might seem that the restrictions necessary to be a member of SID are too stringent and its only constituents are trivial textbook examples; however, as the next theorem shows, strongly infinitely divisible distributions are prevalent and quite easy to construct.

Theorem 3.1.1. *If $\mu \in M_{\pm}(\mathbb{R})$ puts nonnegative mass at 0 and is such that $|\mathbb{E}||X|^p < \infty$ for some $1 \leq p < \infty$, then*

$$\widehat{\gamma}(t) = \exp\left\{\lambda(\widehat{\mu}(t) - 1)\right\}$$

is an SID characteristic function of some $\gamma \in M_{\pm}$ for each $\lambda \in \mathbb{R}$. Furthermore, γ satisfies

$$\begin{aligned}
\|\gamma^{(m)}\| &\leq \exp\left\{\frac{|\lambda|}{m}(\|\mu\| \mp 1)\right\} \quad \textit{and} \\
|\mathbb{E}_{\gamma}||X^{(m)}| &\leq \exp\left\{\frac{|\lambda|}{m}(\|\mu\| \mp 1)\right\} \cdot \frac{|\lambda|}{m} \cdot |\mathbb{E}_{\mu}||X|
\end{aligned}$$

for each $m \in \mathbb{N}$ according to whether λ is positive or negative, respectively.

Remark. *A distribution of the above form is called **Generalized Compound Poisson**, denoted $GCP(\lambda, \mu)$. In one sense this statement extends the classical Lévy's Theorem: the above is an infinitely divisible characteristic function when $\mu \in M_+$ and $\lambda \geq 0$. However, notice that in the more general setting an additional moment restriction is used, so the extension is not entirely direct.*

Proof. In the trivial case $\lambda = 0$ the function f is identically 1 and the corresponding degenerate random variable at 0 is clearly SID, for any μ. Write P and N for the Hahn Decomposition of Ω into μ's positive and negative sets, respectively, and let $\mu = \mu^+ - \mu^-$ be the corresponding Jordan Decomposition of μ.

For $\lambda \neq 0$ define

$$\nu_n = \frac{\lambda}{n} \cdot \mu + \left(1 - \frac{\lambda}{n}\right) \cdot 1_0 = \frac{\lambda}{n}\mu^+ + \left(1 - \frac{\lambda}{n}\right) 1_0 - \frac{\lambda}{n}\mu^-,$$

where 1_0 is the point mass at 0.

Case 1. $\lambda > 0$. For $0 \notin E \subset P$ and $0 \notin F \subset N$,

$$\nu_n(E) = \frac{\lambda}{n}\mu^+(E) > 0, \quad \nu_n(F) = -\frac{\lambda}{n}\mu^-(F) < 0,$$

and

$$\nu_n(\{0\}) = \frac{\lambda}{n}\mu^+(\{0\}) + \left(1 - \frac{\lambda}{n}\right) \geq 0, \quad \text{for } n \geq \lambda.$$

This means that for all n sufficiently large,

$$\begin{aligned}
\nu_n &= \frac{\lambda}{n}\mu^+ + \left(1 - \frac{\lambda}{n}\right) 1_0 - \frac{\lambda}{n}\mu^-, \\
|\nu_n| &= \frac{\lambda}{n}\mu^+ + \left(1 - \frac{\lambda}{n}\right) 1_0 + \frac{\lambda}{n}\mu^- = \frac{\lambda}{n}|\mu| + \left(1 - \frac{\lambda}{n}\right) 1_0, \\
\|\nu_n\| &= \frac{\lambda}{n}\|\mu\| + \left(1 - \frac{\lambda}{n}\right) = \left(1 + \frac{\lambda(\|\mu\| - 1)}{n}\right),
\end{aligned}$$

and

$$|\mathbb{E}_{\nu_n}||X|^p = \frac{\lambda}{n}|\mathbb{E}_\mu||X|^p.$$

Case 2. $\lambda < 0$. One may repeat the above procedure and after similar calculations observe that for all n sufficiently large, that is, for $\lambda < 0$ and $n \geq |\lambda|(\mu^+(\{0\}) - 1)$,

$$\begin{aligned} \nu_n &= \frac{|\lambda|}{n}\mu^- + \left(1 + \frac{|\lambda|}{n}\right) 1_0 - \frac{|\lambda|}{n}\mu^-, \\ |\nu_n| &= \frac{|\lambda|}{n}|\mu| + \left(1 + \frac{|\lambda|}{n}\right) 1_0, \\ \|\nu_n\| &= \left(1 + \frac{|\lambda|(\|\mu\| + 1)}{n}\right), \end{aligned}$$

and

$$|\mathbb{E}_{\nu_n}||X|^p = \frac{|\lambda|}{n}|\mathbb{E}_\mu||X|^p.$$

From the calculations for the two cases it is clear that the characteristic function of ν_n is

$$\widehat{\nu_n} = \frac{\lambda}{n}\widehat{\mu} + \left(1 - \frac{\lambda}{n}\right) = 1 + \frac{\lambda(\widehat{\mu} - 1)}{n}$$

and so its n^{th} power is a characteristic function as well. As $n \to \infty$,

$$\widehat{\nu_n^{*n}} = \left(1 + \frac{\lambda(\widehat{\mu} - 1)}{n}\right)^n \longrightarrow \exp\left\{\lambda(\widehat{\mu} - 1)\right\},$$

thus the characteristic functions converge pointwise; it remains to verify boundedness and tightness. Owing to the above calculations and Proposition 1.2.1 one may say that the variation of ν_n^{*n} when λ is positive (resp. negative) is at most

$$\prod_{i=1}^n \|\nu_n\| = \left(1 + \frac{|\lambda|(\|\mu\| \mp 1)}{n}\right)^n.$$

But this last bound increases as $n \to \infty$ to $\exp\{\lambda(\|\mu\| \mp 1)\}$, and consequently $\{\nu_n^{*n}\}$ is bounded.

One may use Proposition 1.2.4 and equation (1.2.1) in Proposition 1.2.3 to assert

$$|\mathbb{E}_\nu^{*n}||X|^r \leq |\mathbb{E}_\nu|^{*n}|X|^r \leq n^r\|\nu_n\|^{n-1}|\mathbb{E}_{\nu_n}||X|^r,$$

for any $1 \leq r < \infty$. Take $r = 1$ and use the above calculations to see that the right hand side of the last expression is (as λ is pos., neg., respectively)

$$\begin{aligned} &= n \cdot \left(1 + \frac{|\lambda|(\|\mu\| \mp 1)}{n}\right)^{n-1} \frac{|\lambda|}{n} \cdot |\mathbb{E}_\mu||X| \\ &= \left(1 + \frac{|\lambda|(\|\mu\| \mp 1)}{n}\right)^{n-1} |\lambda| \cdot |\mathbb{E}_\mu||X| \\ &\leq \exp\{|\lambda|(\|\mu\| \mp 1)\} \cdot |\lambda| \cdot |\mathbb{E}_\mu||X|. \end{aligned}$$

Therefore by Proposition B.5 it is established that $\{\nu_n^{*n}\}$ is tight and Theorem B.4 guarantees that there exists a signed measure $\gamma \in M_\pm$ such that

$$\widehat{\gamma} = \exp\left\{\lambda(\widehat{\mu} - 1)\right\}, \quad \text{as desired.}$$

It remains to be seen that γ is SID. It is clear that since $\exp\{\lambda(\widehat{\mu} - 1)\}$ is a characteristic function for all $\lambda \in \mathbb{R}$, then

$$(\widehat{\gamma})^{1/k} = \exp\left\{\frac{\lambda}{k}(\widehat{\mu} - 1)\right\}$$

is also a characteristic function for all k, and so γ is GID. And it is known from Proposition B.2 that

$$\|\gamma^{(m)}\| \leq \lim_{n\to\infty} \|\nu_n^{*n}\| \leq \lim_{n\to\infty} \left(1 + \frac{|\lambda|(\|\mu\| \mp 1)}{mn}\right)^n = \exp\left\{\frac{|\lambda|}{m}(\|\mu\| \mp 1)\right\}.$$

Lastly,

$$|\mathbb{E}||X^{(m)}| \leq \exp\left\{\frac{|\lambda|}{m}(\|\mu\| \mp 1)\right\} \cdot \frac{|\lambda|}{m} \cdot |\mathbb{E}_\mu||X|,$$

which shows

$$\sup_m |\mathbb{E}| \left(1 + |X^{(m)}|\right) \leq \exp\{|\lambda|(\|\mu\| \mp 1)\} \left(1 + |\lambda| \cdot |\mathbb{E}_\mu||X|\right) < \infty.$$

Therefore, γ is SID, as desired. $\square$

Example.

The $GPoi(\lambda)$ class is a good example of characteristic functions that have the above mentioned form; indeed, $\varphi(t) = \exp\{\lambda(e^{it}-1)\}$ is such a characteristic function with μ being the measure putting unit mass at the point $x = 1$. Notice in this case that equality holds with $\|\gamma^{(m)}\|$ and $|\mathbb{E}_\gamma||X^{(m)}|$, another way in which the $GPoi(\lambda)$ class is special.

It was suggested at the beginning of the section that the defining property of SID should impose some structure on a GID random variable's infinitesimal summands. This notion is made precise in the following proposition.

Proposition 3.1.1. *If X is SID, then $\{X^{(m)} : m \in \mathbb{N}\}$ is bounded and tight.*

Proof. For X that is SID let

$$C = \sup_n |\mathbb{E}| \left(1 + |X^{(n)}|^\delta\right) < \infty.$$

Then

$$\sup_n |\mathbb{E}| 1 = \sup_n \|\mu^{(n)}\| \le C$$

shows that $\{\mu^{(n)}\}$ is bounded and

$$\sup_n |\mathbb{E}||X^{(n)}|^\delta \le C$$

shows that $\{\mu^{(n)}\}$ is tight by Proposition B.5. □

Now the tools are available to prove a familiar (and useful) property of classical infinitely divisible distributions, which will show that the SID class preserves at least some of the classical intuition. The proof is similar to the proof in the classical case.

Theorem 3.1.2. *An SID characteristic function never vanishes.*

Proof. Let $f = \widehat{\mu}$ and $f_n = \widehat{\mu^{(n)}}$ be the respective characteristic functions of the SID random variable and its n^{th} factor. Then Proposition 3.1.1 shows that $\{\mu^{(k)} : k \in \mathbb{N}\}$ is bounded (by say, M) and tight. Now write

$$g = |f|^2, \quad g_n = |f_n|^2.$$

These are characteristic functions of signed measures by Proposition 1.2.2, are real and positive, are bounded (by M^2), and tight (Proposition B.6). Notice that for all $t \in \mathbb{R}$,

$$g_n(t) = [g(t)]^{1/n}.$$

But $0 \le g(t) \le M^2$, hence $\lim_{n\to\infty}[g(t)]^{1/n}$ is 0 or 1 according as $g(t) = 0$ or $g(t) \ne 0$. Thus $\lim_{n\to\infty} g_n(t)$ exists for all t, and the limit function, say $h(t)$, can take at most the two possible values 0 and 1. Furthermore, since g is continuous at $t = 0$ with $g(0) = 1$, there exists a $t_0 > 0$ such that $g(t) \ne 0$ for all $|t| \le t_0$, and it follows that $h(t) = 1$ for $|t| \le t_0$.

By boundedness, tightness, and convergence of the characteristic functions we may conclude from Theorem B.4 that there exists a (complex) measure $\nu \in M$ such that $\widehat{\nu} = h$, and so h must in fact be continuous everywhere. Hence h is identically equal to 1, and so for all t,

$$|f(t)|^2 = g(t) \ne 0.$$

The theorem is proved. $\square$

3.2 The Closure of the Class SID

It has been demonstrated that SID shares many of the familiar characteristics of classical infinitely divisible distributions. However, this is where the similarity ends, for one of the most useful properties of the class ID is that it is closed, and the same may not be said of SID distributions.

Proposition 3.2.1. *SID is **not** a closed subset of $M_\pm(\mathbb{R})$ under the weak topology.*

Proof. In Section 3.1, Example 3.1.3, it was proved that $GBern(p)$ is SID for any $p < 1/2$. Let $X_n \sim GBern(p_n)$ for some sequence (p_n) increasing to $p = 1/2$. Clearly, X_n converges weakly as $n \to \infty$ to $X \sim GBern(1/2)$. However, the characteristic function of X is

$$\varphi_X(t) = \frac{1 + e^{it}}{2},$$

which is zero at the points $t = \pm(2k+1)\pi$, $k \in \mathbb{N}$. Then X is not SID, by Theorem 3.1.2.

$\square$

Consequently, SID is not a closed subset of $M_{\pm}(\mathbb{R})$. Refer to the end of the section to see the reason for the failure in this example .

The above proof actually shows something more. Notice that while SID is not closed in M, it is still possible that SID may be closed *when considered as a subspace of* GID. To see that SID is not closed in this less restrictive sense, remember that from Theorem 2.3.1 follows that X above is GID, and (of course) X_n is GID as well, therefore SID is in fact not even closed as a subspace of GID!

This property (or, rather, lack thereof) is unexpected and somewhat unsettling. In the absence of closure, many of the classical results regarding ID distributions collapse. One might wonder if anything at all may be said about SID. Luckily it turns out that SID is *almost* closed, in the following sense:

Theorem 3.2.1. *SID is an F_σ set under the weak topology.*

If one lets

$$A_N = \{\mu : \sup_j \|\mu^{(j)}\| \leq N\},$$

and

$$B(N,k,M) = \{\mu \in A_N : \sup_l |\mathbb{E}||X^{(l)}|^{1/k} \leq M\},$$

then the claim is that

$$SID = \bigcup_{N=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcup_{M=1}^{\infty} B(N, k, M), \tag{3.2.1}$$

where $B(N, k, M)$ is closed under weak convergence for each $N, k, M \in \mathbb{N}$.

Proof. First is shown the equality of the two sets in equation (3.2.1). If μ is an element of the right hand side then $\mu \in B(N, k, M)$ for some N, k, and M; further $\sup_j \|\mu^{(j)}\| \leq N < \infty$ and $1/k > 0$ satisfies $\sup_j |\mathbb{E}|\, |X^{(j)}|^{1/k} \leq M < \infty$, and so

$$\sup_j |\mathbb{E}| \left(1 + |X^{(j)}|^{1/k}\right) \leq N + M < \infty.$$

Therefore μ is SID by definition of SID. On the other hand, if μ is an element of the left hand side then $\sup_j |\mathbb{E}| \left(1 + |X^{(j)}|^{1/k}\right) = C$ for some $\delta > 0$ and $C < \infty$. This means that $\mu \in A_N$ for all $N \geq C$; also for any j and for any $k \in \mathbb{N}$ satisfying $1/k < \delta$ one has from Proposition A.2.1

$$|\mathbb{E}||X^{(j)}|^{1/k} \leq \left(|\mathbb{E}||X^{(j)}|^{\delta}\right)^{1/(k\delta)} \cdot \|\mu\|^{1-1/(k\delta)} \leq C^{1/(k\delta)}\, N^{1-1/(k\delta)} < \infty.$$

Choosing $M \geq C^{1/(k\delta)}\, N^{1-1/(k\delta)}$ allows $\mu \in B(N, k, M)$.

It remains to be seen that $B(N, k, M)$ is closed under weak convergence. To that end let $\{\mu_n\} \subset B(N, k, M)$ satisfy $\mu_n \Rightarrow \mu$ for some $\mu \in M(\mathbb{R})$, and the task is to show that $\mu \in B(N, k, M)$. To do that it is first established that μ is GID.

Since $\mu_n \Rightarrow \mu$ one has $\widehat{\mu_n} \to \widehat{\mu}$ pointwise. Let $m > 1$ be fixed and consider

$$\widehat{\mu_n^{(m)}} = (\widehat{\mu_n})^{\frac{1}{m}}, \quad n = 1, 2, \ldots.$$

These are by assumption Fourier-Stieltjes Transforms for $\{\mu_n^{(m)}\} \subset M_{\pm}(\mathbb{R})$ and they converge pointwise to

$$\lim_{n\to\infty} (\widehat{\mu_n})^{\frac{1}{m}} = (\widehat{\mu})^{\frac{1}{m}} = g_m, \tag{3.2.2}$$

where g_m is the distinguished m^{th} root as in Chung [9]. It is required that there exist $\mu^{(m)} \in M_{\pm}(\mathbb{R})$ such that

$$\widehat{\mu^{(m)}} = g_m.$$

But since $\{\mu_n\} \subset \bigcup_{k\geq 1}\bigcup_{M\geq 1} B(N,k,M)$ the measures $\{\mu_n^{(m)}\}$ are bounded (by N) and also

$$|\mathbb{E}||X_n^{(m)}|^{1/k} \leq M$$

for all n because $\{\mu_n\} \subset \bigcup_{N\geq 1} B(N,k,M)$. Therefore

$$\sup_n |\mathbb{E}||X_n^{(m)}|^{1/k} \leq \frac{M}{m^{1/k}} < \infty,$$

and a quick application of Proposition B.5 gives that $\{\mu_n^{(m)} : n \geq 1\}$ is bounded and tight. By Theorem 1.3.1 there exists a signed measure $\nu_m \in M_{\pm}(\mathbb{R})$ and a subsequence (n_j) such that $\mu_{n_j}^{(m)} \xrightarrow{v} \nu_m$ as $j \to \infty$, and in fact, one may even arrange it so that $\limsup_{j\to\infty} \|\mu_{n_j}^{(m)}\| \leq \|\nu_m\|$; but then (taking $O = \mathbb{R}$) Proposition B.2 yields

$$\lim_{j\to\infty} \|\mu_{n_j}^{(m)}\| = \|\nu_m\|; \tag{3.2.3}$$

(this will be needed later). Then by Proposition B.4:

$$\widehat{\mu_{n_j}^{(m)}} \longrightarrow \widehat{\nu_m} \tag{3.2.4}$$

pointwise, but it is already known from equation (3.2.2) that $\widehat{\mu_{n_j}^{(m)}} \longrightarrow g_m$, hence $g_m = \widehat{\nu_m}$. The signed measure $\mu^{(m)}$ $(= \nu_m)$ has thus been found such that

$$\widehat{\mu^{(m)}} = \widehat{\nu_m} = g_m = \left(\widehat{\mu}\right)^{1/m},$$

and since m was arbitrary, μ is GID.

Using equation (3.2.3) one sees $\|\mu^{(m)}\| = \lim_{j\to\infty} \|\mu_{n_j}^{(m)}\| \leq N$, so $\mu \in A_N$. Also, Equations (3.2.3) and (3.2.4) together with Theorem B.5 show that $|\mu_{n_j}^{(m)}| \Rightarrow |\mu^{(m)}|$; then by boundedness and Theorem B.6 it may be asserted

$$|\mathbb{E}||X^{(m)}|^{1/k} \leq \sup_j |\mathbb{E}||X_{n_j}^{(m)}|^{1/k} \leq M,$$

and therefore $\mu \in B(N, k, M)$, as required. $\square$

Once this fundamental property of SID has been established, it suggests the question, "What went wrong in the $GBern(p_n)$ example at the beginning of the section?"

The answer lies buried in earlier calculations. The p_n converge to $p = 1/2$, which has been seen to be equivalent to a sequence of $GNegB(-1, \alpha_n)$ random variables with $\alpha_n \to 2$ as $n \to \infty$. Referring to the calculations in Section 3.1, for any fixed $m > 1$,

$$\|\nu^{(m)}\| = 2\alpha_n^{-1/m} - \left(\frac{\alpha_n}{2-\alpha_n}\right)^{-1/m},$$

and this converges as $n \to \infty$ to $2^{1-1/m}$; therefore the trouble must not be in the variations. We look a little further and find

$$|\mathbb{E}||X_n^{(m)}| = \left(\frac{\alpha_n}{2-\alpha_n}\right)^{-1/m} \frac{\alpha_n - 1}{m(2-\alpha_n)}.$$

This last quantity explodes as $n \to \infty$. The conclusion is that even when the variations behave in ways that might seem completely reasonable, mass may still "escape to infinity" in the moments, and this behavior can cause problems in unexpected places.

3.3 Asymptotic Negligibility and the Canonical Representation

It was learned in the last section that even when SID random variables converge, and even when their component variations behave in a completely reasonable manner, still there may be a loss of mass at infinity in the moments which can cause problems that are exhibited in troubling ways, such as the lack of closure of the SID class.

As a result, one is led to the following terminology: For $p > 0,\ \alpha \geq 0$ one says that a GID random variable X is **asymptotically negligible of rank p and order $\boldsymbol{\alpha}$**, denoted $AN_p(\alpha)$, if

$$|\mathbb{E}|\left(1 + |X^{(n)}|^p\right) = 1 + O(n^{-\alpha}), \quad \text{as } n \to \infty.$$

Note that every SID random variable is $AN_\delta(0)$ for some $\delta > 0$.

Remark. *The defining quantity* $|\mathbb{E}_\nu|\left(1 + |X^{(n)}|^p\right)$ *simplifies to*

$$\|\nu^{(n)}\| + |\mathbb{E}||X^{(n)}|^p = 1 + 2\|\nu^{(n)-}\| + |\mathbb{E}||X^{(n)}|^p.$$

In this form it is clear that X *is* $AN_p(\alpha)$ *if and only if both* $\|\nu^{(n)-}\|$ *and* $|\mathbb{E}||X^{(n)}|^p$ *are* $O(n^{-\alpha})$*. Of course, Theorem 3.1.2 implies that* $\nu^{(n)}$ *tends weakly to* 1_0*, which suggests that the above quantities should tend to 0; the purpose of* $AN_p(\alpha)$ *is to quantify the* speed *of this tendency, in some sense.*

The reason for the terminology "rank" and "order" will become clear from the following

Proposition 3.3.1.

$$AN_p(\alpha) \supset AN_q(\beta) \quad \textit{provided} \quad 0 < p < q \textit{ and } \alpha/p \leq \beta/q.$$

Proof. Supposing that $X \sim \nu$ is $AN_q(\beta)$, it is shown first that $|\mathbb{E}||X^{(n)}|^p$ is $O(n^{-\alpha})$:

$$\begin{aligned}
n^\alpha|\mathbb{E}||X^{(n)}|^p &\leq n^\alpha \cdot \left(|\mathbb{E}||X^{(n)}|^q\right)^{p/q} \cdot \|\nu\|^{1-p/q} \quad \text{by Proposition A.2.1} \\
&= n^{\alpha - \beta\frac{p}{q}} \cdot \left(n^\beta\, |\mathbb{E}||X^{(n)}|^q\right)^{p/q} \cdot \|\nu\|^{1-p/q} \\
&\leq C^{p/q}\, M^{1-p/q},
\end{aligned}$$

provided that $\alpha - \beta(p/q) \leq 0$, that is, provided $\alpha/p \leq \beta/q$.

Now after referring to the remark above it suffices to show that $\|\nu^{(n)-}\| = O(n^{-\alpha})$. But the conditions $p < q$ and $\alpha/p \leq \beta/q$ imply $\alpha < \beta$, from which follows

$$n^{\alpha}\|\nu^{(n)-}\| = \frac{n^{\beta}}{n^{\beta-\alpha}}\|\nu^{(n)^{-}}\| \leq n^{\beta}\|\nu^{(n)-}\| = O(1).$$

$\square$

It turns out that $GPoi(\lambda)$ and $GNegB(r,p)$ are $AN_1(1)$, as is any GCP random variable in Theorem 3.1.1.

Example 3.3.1 ($GPoi(\lambda)$).

This is easy to see because in Section 3.1 is found

$$|\mathbb{E}|\left(1 + |X^{(n)}|\right) = e^{2\lambda/n} + e^{2\lambda/n}\,\frac{|\lambda|}{n}$$

which is $1 + O(n^{-1})$.

Example 3.3.2 ($GNegB(r,p)$).

If $0 < p < 1$ then we are in the classical world; there $\|\nu^{(n)}\| = 1$ for all n, and $|\mathbb{E}||X^{(n)}| = rq/np$ which is $O(n^{-1})$. On the other hand if $1 \leq p < 2$ then

$$\|\nu^{(n)}\| = \left(\frac{p}{2-p}\right)^{r/n} = 1 + \sum_{k=1}^{\infty}\frac{1}{k!}\left[\frac{r}{n}\,\log\left(\frac{p}{2-p}\right)\right]^{k} = 1 + O(n^{-1}),$$

and

$$|\mathbb{E}||X^{(n)}| = \left(\frac{p}{2-p}\right)^{r/n}\frac{|r|}{n}\,\frac{p-1}{2-p} = O(n^{-1}).$$

Membership in $AN_1(1)$ follows.

Example 3.3.3 ($GBern(p)$, $p < 1/2$).

Recall that this is merely the claim that $GNegB(r,p)$ is $AN_1(1)$ for $-1 < r < 0$ any $0 < p < 2$. For any such r and $0 < p < 1$ one has

$$\begin{aligned}
\|\nu^{(n)}\| &= 2p^{r/n} - 1 = 2e^{\frac{r}{n}\log p} - 1 \\
&= 1 + 2\sum_{k=1}^{\infty} \frac{1}{k!}\left(\frac{r\log p}{n}\right)^k \\
&= 1 + O(n^{-1}),
\end{aligned}$$

and $|\mathbb{E}||X^{(n)}| = |r|q/np$ which is $O(n^{-1})$, while for $1 \le p < 2$

$$\begin{aligned}
\|\nu^{(n)}\| &= 2p^{r/n} - \left(\frac{p}{2-p}\right)^{r/n} \\
&= 2\exp\left\{\frac{r}{n}\log p\right\} - \exp\left\{\frac{r}{n}\log\left(\frac{p}{2-p}\right)\right\} \\
&= 1 + \sum_{k=1}^{\infty} \frac{1}{k!}\left[\frac{r}{n}\left(\log\left(\frac{p}{2-p}\right) - \log p\right)\right]^k = 1 + O(n^{-1})
\end{aligned}$$

and

$$|\mathbb{E}||X^{(n)}| = \left(\frac{p}{2-p}\right)^{r/n} \frac{|r|}{n}\frac{p-1}{2-p} = O(n^{-1});$$

consequently $GBern(p)$ is $AN_1(1)$ for all $p < 1/2$.

Example 3.3.4 ($GCP(\lambda,\mu)$).

From Section 3.1 these signed measures γ satisfy

$$\begin{aligned}
\|\gamma^{(m)}\| &\le \exp\left\{\frac{|\lambda|}{m}(\|\mu\| \mp 1)\right\} \quad \text{and} \\
|\mathbb{E}_\gamma||X^{(m)}| &\le \exp\left\{\frac{|\lambda|}{m}(\|\mu\| \mp 1)\right\} \cdot \frac{|\lambda|}{m} \cdot |\mathbb{E}_\mu||X|
\end{aligned}$$

and the $AN_1(1)$ condition is obviously fulfilled for these random variables.

One of the most famous results for classical ID distributions is the Lévy-Khintchine canonical representation of ID characteristic functions. Now that many properties and examples of SID distributions have been built up, one might wonder whether an appropriate analogue of this result exists in the signed case and, if so, exactly what form it would have.

It turns out that the SID class *does* have a corresponding representation for its characteristic functions, the derivation of which we now undertake.

Theorem 3.3.1 (The Canonical Representation). *In order that the function $f(t)$ be the characteristic function of a strongly infinitely divisible $AN_1(1)$ distribution, it is necessary that its logarithm be representable in the form*

$$\log f(t) = iat + \int \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2}\, \mathrm{d}\theta(x), \tag{3.3.1}$$

where a is a real constant and $\theta(x)$ is a function of bounded variation satisfying the moment conditions $\int |x|^k \,\mathrm{d}|\theta|(x) < \infty,\ k = -1, 0, 1$. The integrand at $x = 0$ is defined by continuity to be

$$\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2}\bigg|_{x=0} = -\frac{t^2}{2}.$$

Conversely, if the logarithm of a function f is representable in the form (3.3.1)*, where θ is as above and satisfies the additional requirements $\int |x|^{-2}\,\mathrm{d}|\theta|(x) < \infty$ and $\int \mathrm{d}\theta \neq -\int x^{-2}\,\mathrm{d}\theta(x)$, then f is the characteristic function of an SID random variable X that is $AN_1(1)$. Moreover, $X \sim \nu$ has*

$$\begin{aligned} \|\nu^{(m)}\| &\leq \exp\left\{\frac{2}{m}\int \left(1 + \frac{1}{x^2}\right)\,\mathrm{d}\theta^{-}(x)\right\} \qquad \textit{and} \\ |\mathbb{E}_\nu||X^{(m)}| &\leq \frac{1}{m}\cdot\int\left(|x| + \frac{1}{|x|}\right)\,\mathrm{d}|\theta|(x). \end{aligned}$$

Lastly, the representation of $\log f$ by (3.3.1) *is unique.*

Remark. *Owing to Proposition 3.3.1, we may say that the above representation in fact holds for all random variables X that are $AN_q(\beta)$ with $0 < q \leq \beta$.*

Before the theorem may be proved some preliminary lemmas will be needed.

Lemma 3.3.1. *Let a be a constant and let θ be a real-valued function of bounded variation satisfying*

$$\int |x|^k \, \mathrm{d}|\theta|(x) < \infty, \;\; k = -2, -1, 0, 1, \quad \textit{with} \quad \int \mathrm{d}\theta \neq -\int x^{-2} \, \mathrm{d}\theta(x).$$

Suppose that a function $f(t)$ of the real variable t admits the representation (3.3.1)*. Then f takes the form*

$$\log f(t) = \lambda(\widehat{\mu} - 1) + ict,$$

and, in particular, f is the characteristic function of an SID random variable X.

Proof. Manipulate f into a more recognizable form:

$$\begin{aligned}
\log f(t) = iat + \int \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2} \, \mathrm{d}\theta(x) \\
= \int e^{itx} \frac{1+x^2}{x^2} \, \mathrm{d}\theta(x) - \int \frac{1+x^2}{x^2} \, \mathrm{d}\theta(x) + iat - it \int \frac{1}{x} \, \mathrm{d}\theta(x) \\
= \left(\int 1 + \frac{1}{u^2} \, \mathrm{d}\theta(u)\right) \left[\frac{1}{\int 1 + \frac{1}{u^2} \, \mathrm{d}\theta(u)} \int e^{itx} \left(1 + \frac{1}{x^2}\right) \mathrm{d}\theta(x) - 1\right] \\
+ it \left(a - \int \frac{1}{x} \, \mathrm{d}\theta(x)\right).
\end{aligned}$$

The above is recognized as having the form $\log f(t) = \lambda(\widehat{\mu}(t) - 1) + ict$, where

1. $\lambda = \int \left(1 + \frac{1}{u^2}\right) \mathrm{d}\theta(u)$,

2. $\mathrm{d}\mu(x) = \frac{1}{\int \left(1+\frac{1}{u^2}\right) \mathrm{d}\theta(u)} \cdot \left(1 + \frac{1}{x^2}\right) \mathrm{d}\theta(x)$,

3. $c = \left(a - \int \frac{1}{x} \, \mathrm{d}\theta(x)\right)$.

The moment condition $\int |x|^{-1} \, \mathrm{d}|\theta|(x) < \infty$ implies that θ (and hence μ) put mass 0 at the origin. Next is verified

$$\begin{aligned}
\int |x| \, \mathrm{d}|\mu|(x) &= \frac{1}{\left|\int \left(1 + \frac{1}{u^2}\right) \mathrm{d}\theta(u)\right|} \int |x| \left(1 + \frac{1}{x^2}\right) \mathrm{d}|\theta|(x) \\
&= \frac{1}{\left|\int \left(1 + \frac{1}{u^2}\right) \mathrm{d}\theta(u)\right|} \int \left(|x| + \frac{1}{|x|}\right) \mathrm{d}|\theta|(x) < \infty,
\end{aligned}$$

and it then follows immediately from Theorem 3.1.1 that $X \sim \nu = f^{\vee}$ is SID.

$\square$

Lemma 3.3.2 (Kolmogorov [26]). *If the logarithm of a characteristic function is representable in the form* (3.3.1)*, where θ is a function of bounded variation, then such a representation is unique.*

Proof. Suppose that f has two representations of the form (3.3.1) with constants $a_1,\ a_2$ and functions $\theta_1,\ \theta_2$, respectively.

Let $\theta_1^{\pm}$ and $\theta_2^{\pm}$ be the positive and negative variations of $\theta_1,\ \theta_2$, and let $h(t,x)$ be the integrand in (3.3.1).

One obtains then

$$\begin{aligned} ia_1t &+ \int h(t,x)\,\mathrm{d}\theta_1(x) \\ &= ia_1t + \int h(t,x)\,\mathrm{d}\theta_1^+(x) - \int h(t,x)\,\mathrm{d}\theta_1^-(x) \\ &= ia_2t + \int h(t,x)\,\mathrm{d}\theta_2(x) \\ &= ia_2t + \int h(t,x)\,\mathrm{d}\theta_2^+(x) - \int h(t,x)\,\mathrm{d}\theta_2^-(x), \end{aligned}$$

whence

$$ia_1t + \int h(t,x)\,\mathrm{d}(\theta_1^+(x) + \theta_2^-(x)) = ia_2t + \int h(t,x)\,\mathrm{d}(\theta_1^-(x) + \theta_2^+(x)).$$

But the functions $\theta_1^+(x) + \theta_2^-(x)$ and $\theta_1^-(x) + \theta_2^+(x)$ are nondecreasing; hence in the equation written above, *representing the logarithm of a classical infinitely divisible law*, one must have

$$a_1 = a_2 \quad \text{and} \quad \theta_1^+(x) + \theta_2^-(x) \equiv \theta_1^-(x) + \theta_2^+(x),$$

that is,

$$\theta_1^+(x) - \theta_1^-(x) \equiv \theta_2^+(x) - \theta_2^-(x).$$

This is equivalent to $\theta_1 \equiv \theta_2$. $\square$

Lemma 3.3.3. *Let $\{\phi_n\}$ be determined according to* (3.3.1) *by constants $\{a_n\}$ and functions $\{\theta_n\}$ which are tight and of uniformly bounded variation. Suppose ϕ_n converges pointwise to a complex valued function ϕ. Then there exist a constant a and a function θ of bounded variation such that*

1. $a_n \longrightarrow a$,
2. $\theta_n \Rightarrow \theta$,
3. $|\theta_n| \Rightarrow |\theta|$.

The function ϕ is determined by a and θ according to (3.3.1).

Proof. Since $\{\theta_n\}$ is bounded and tight, there exists a function θ of bounded variation such that $\theta_n \Rightarrow \theta$ and $|\theta_n| \Rightarrow |\theta|$ due to Theorem 1.3.1 and Theorem B.5. But $\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}$ is bounded and continuous in x; therefore weak convergence implies

$$I_n(t) = \int \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}\,\mathrm{d}\theta_n \quad \text{approaches}$$
$$\int \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}\,\mathrm{d}\theta = I(t).$$

Now $\phi_n(t) \to \phi(t)$ and $I_n(t) \to I(t)$; it then follows that the sequence $\{a_n\}$ converges to a constant a and that ϕ is determined by a and θ according to (3.3.1). $\square$

Lemma 3.3.4. *Let f be the characteristic function of an SID random variable X that is $AN_1(1)$. Then there exists a sequence (ϕ_n) of functions determined according to* (3.3.1) *by constants $\{a_n\}$ and functions $\{\theta_n\}$ which are tight and of uniformly bounded variation, and there exists a complex valued function ϕ such that $\phi_n \longrightarrow \phi$ pointwise.*

Proof. Since f is the characteristic function of an SID law, one has $f \neq 0$ and thus, as $n \to \infty$,

$$\begin{aligned} n\{[f(t)]^{1/n} - 1\} &= n\{e^{(1/n)\log f(t)} - 1\} \\ &= n\{(1/n)\log f(t) + o(1/n)\} = \log f(t) + o(1). \end{aligned}$$

Define $\phi = \log f$ and see that ϕ_n approaches ϕ pointwise.

Now by hypothesis $(f)^{1/n}$ is a characteristic function of a distribution which will be denoted by F_n. Then

$$\begin{aligned} n\{[f(t)]^{1/n} - 1\} &= \int (e^{itx-1}\, \mathrm{d}F_n(x)) \\ &= nit \int \frac{x}{1+x^2}\, \mathrm{d}F_n(x) + n \int \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \mathrm{d}F_n(x). \end{aligned}$$

If one denotes

$$\begin{aligned} a_n &= n \int \frac{x}{1+x^2}\, \mathrm{d}F_n(x), \\ \theta_n(x) &= n \int_{-\infty}^{x} \frac{y^2}{1+y2}\, \mathrm{d}F_n(y), \\ \phi_n(t) &= ia_n t + \int \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2}\, \mathrm{d}\theta_n(x), \end{aligned}$$

then it is clear that ϕ_n has the form (3.3.1), as required.

It remains to show that $\{\theta_n\}$ is bounded and tight. Since for $A \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} \theta_n(A) &= \int_A \frac{nx^2}{1+x^2}\, \mathrm{d}F_n \\ &= \int_A \frac{nx^2}{1+x^2}\, \mathrm{d}F_n^+ - \int_A \frac{nx^2}{1+x^2}\, \mathrm{d}F_n^-, \end{aligned}$$

every positive (resp. neg., null) set of F_n is a positive (resp. neg., null) set of θ_n, implying

$$\mathrm{d}|\theta_n|(x) = \frac{nx^2}{1+x^2}\, \mathrm{d}|F_n|(x).$$

Then

$$\|\theta_n\| = |\theta_n|(\mathbb{R}) = n \int \frac{x^2}{1+x^2}\, \mathrm{d}|F_n|(x).$$

Now remembering that X is $AN_1(1)$ it follows

$$\begin{aligned} \|\theta_n\| &= n\int |x|\frac{|x|}{1+x^2}\,\mathrm{d}|F_n|(x) \\ &\leq n\int |x|\,\mathrm{d}|F_n|(x) = n|\mathbb{E}||X^{(n)}| = O(1), \end{aligned}$$

and so $\{\theta_n\}$ is bounded. Similarly, for any given $-1 \leq \delta \leq 1$

$$\begin{aligned} |\mathbb{E}_{\theta_n}||X|^\delta &= \int |x|^\delta\,\mathrm{d}|\theta_n|(x) \\ &= n\int |x|^\delta \frac{x^2}{1+x^2}\,\mathrm{d}|F_n|(x) \\ &= n\int |x|\left(\frac{|x|^{1+\delta}}{1+x^2}\right)\mathrm{d}|F_n|(x) \\ &\leq n\int |x|\,\mathrm{d}|F_n|(x) = O(1), \end{aligned}$$

thus tightness follows from Proposition B.5. □

Proof of Theorem 3.3.1. (Necessity). Suppose f is the characteristic function of an SID random variable X that is $AN_1(1)$. Then by Lemma 3.3.4 there exists a sequence (ϕ_n) of functions determined according to (3.3.1) by constants $\{a_n\}$ and functions $\{\theta_n\}$ which are bounded and tight, such that $\phi_n \longrightarrow \phi = \log f$ pointwise.

Now Lemma 3.3.3 shows that there exist a constant a and a function of bounded variation θ such that

1. $a_n \longrightarrow a$,

2. $\theta_n \Rightarrow \theta$,

3. $|\theta_n| \Rightarrow |\theta|$.

Further, $\log f = \phi$ is completely determined by the constant a and the function θ according to (3.3.1). The last detail is to check the moment conditions. It has been proved

$\int |x|^k \,\mathrm{d}|\theta_n|(x) = O(1)$ in Lemma 3.3.4 for the cases $k = 0$ (the boundedness calculation) and $k = \pm 1$ (take $\delta = \pm 1$ in the calculation of $|\mathbb{E}_{\theta_n}||X|^\delta$). These facts together with the aid of the weak convergence of the total variation measures and arguments analogous to those in Theorem B.6 yield that $\int |x|^k \,\mathrm{d}|\theta|(x) < \infty$, $k = -1, 0, 1$.

(Sufficiency). Suppose that (3.3.1) holds with θ satisfying $\int |x|^k \,\mathrm{d}|\theta|(x) < \infty$, $k = -2, -1, 0, 1$ and $\int \,\mathrm{d}\theta \neq -\int x^{-2} \,\mathrm{d}\theta(x)$. Then we proved in Lemma 3.3.1 that $f = \widehat{\nu}$ is indeed the characteristic function of a random variable X that is SID. It was shown there that $\widehat{\nu}$ takes the form $\widehat{\nu}(t) = \exp\{\lambda(\widehat{\mu}(t) - 1) + ict\}$, where

1. $\lambda = \int \left(1 + \frac{1}{u^2}\right) \,\mathrm{d}\theta(u)$,
2. $\mathrm{d}\mu(x) = \frac{1}{\int \left(1 + \frac{1}{u^2}\right) \mathrm{d}\theta(u)} \cdot \left(1 + \frac{1}{x^2}\right) \,\mathrm{d}\theta(x)$,
3. $c = \left(a - \int \frac{1}{x} \,\mathrm{d}\theta(x)\right)$,

and

$$|\mathbb{E}_\mu||X| = \int \left(|x| + \frac{1}{|x|}\right) \,\mathrm{d}\frac{|\theta|(x)}{\left|\int \left(1 + \frac{1}{u^2}\right) \,\mathrm{d}\theta(u)\right|}.$$

As in the Examples, note that this implies X is $AN_1(1)$.

Now refer to Theorem 3.1.1, and from there is seen $\|\nu\| \leq \exp\{|\lambda|(\|\mu\| \mp 1)\}$ according to whether $\lambda > 0$ or $\lambda < 0$, respectively. Next is calculated:

$$\begin{aligned}
|\lambda|(\|\mu\| \mp 1) &= \left|\int \left(1 + \frac{1}{u^2}\right) \,\mathrm{d}\theta(u)\right| \left[\int \left(1 + \frac{1}{x^2}\right) \frac{\mathrm{d}|\theta|(x)}{\left|\int 1 + \frac{1}{u^2} \,\mathrm{d}\theta(u)\right|} \mp 1\right] \\
&= \pm \int \left(1 + \frac{1}{u^2}\right) \,\mathrm{d}\theta(u) \left[\int \left(1 + \frac{1}{x^2}\right) \frac{\mathrm{d}|\theta|(x)}{\pm \int 1 + \frac{1}{u^2} \,\mathrm{d}\theta(u)} \mp 1\right] \\
&= \int \left(1 + \frac{1}{x^2}\right) \,\mathrm{d}|\theta|(x) - \int \left(1 + \frac{1}{x^2}\right) \,\mathrm{d}\theta(x).
\end{aligned}$$

Combine the integrals and use $|\theta| - \theta = \theta^+ + \theta^- - (\theta^+ - \theta^-) = 2\theta^-$ to reach

$$\|\nu\| \leq \exp\left\{2 \int \left(1 + \frac{1}{x^2}\right) \,\mathrm{d}\theta^-(x)\right\}.$$

Lastly from Theorem 3.1.1 follows

$$\begin{aligned}
|\mathbb{E}_\nu||X^{(m)}| &\leq \frac{1}{m}\cdot|\lambda||\mathbb{E}_\mu||X| \\
&= \frac{1}{m}\cdot\left|\int\left(1+\frac{1}{u^2}\right)\,\mathrm{d}\theta(u)\right|\int|x|\,\left(1+\frac{1}{x^2}\right)\,\frac{\mathrm{d}|\theta|(x)}{\left|\int 1+\frac{1}{u^2}\,\mathrm{d}\theta(u)\right|} \\
&= \frac{1}{m}\cdot\int\left(|x|+\frac{1}{|x|}\right)\,\mathrm{d}|\theta|(x).
\end{aligned}$$

Finally, uniqueness of the representation was shown in Lemma 3.3.2, and the proof of the theorem is complete. □

As an application of the theorem, we return to an example discussed by Gnedenko and Kolmogorov [26]. This, together with $GBin(p)$, when $0 < p < 1/2$, provides an example of *classical* discrete random variables which are not ID, yet are SID with a corresponding Lévy-Khinchine representation. We may now look back many years later and see this example in an entirely new light.

Example 3.3.5.

Consider the function

$$f(t) = \frac{1-\beta}{1+\alpha}\cdot\frac{1+\alpha e^{-it}}{1-\beta e^{it}}, \quad 0 < \alpha \leq \beta < 1.$$

The function is continuous, $f(0) = 1$, and

$$f(t) = \frac{1-\beta}{1+\alpha}\left[\alpha e^{-it} + (1+\alpha\beta)\sum_{n=0}^{\infty}\beta^n e^{int}\right].$$

Hence it is the characteristic function of a random variable taking all integral values from -1 to ∞; moreover

$$\mathbb{P}(X=-1) = \frac{1-\beta}{1+\alpha}\alpha, \quad \mathbb{P}(X=n) = \frac{1-\beta}{1+\alpha}\cdot(1+\alpha\beta)\beta^n, \quad n = 0, 1, \ldots$$

Observe that X is an SID random variable which is not classically ID. For,

$$\begin{aligned}\log f(t) &= \sum_{n=1}^{\infty}\left[(-1)^{n-1}\frac{\alpha^n}{n}(e^{-int}-1)+\frac{\beta^n}{n}(e^{int}-1)\right]\\ &= i\gamma t+\int(e^{itu}-1-\frac{itu}{1+u^2})\frac{1+u^2}{u^2}\mathrm{d}\theta(u),\end{aligned}$$

where, as can easily be calculated,

$$\gamma=\sum_{n=1}^{\infty}\frac{\beta^n+(-1^n)\alpha^n}{1+n^2},$$

and $\theta(u)$ is a function of bounded variation, having jumps at the points $u=\pm1,\ \pm2,\ \pm3,\ldots,$ of magnitudes

$$\frac{n\beta^n}{n^2+1}\quad\text{for } u=+n$$

and

$$(-1)^{n-1}\frac{n\alpha^n}{n^2+1}\quad\text{for } u=-n.$$

Thus θ is not monotone, and by the classical Lévy-Khinchine representation X is not ID. It is however immediate that X is SID, in fact $AN_1(1)$, once one quickly verifies the moment conditions

$$\int|x|^k\,\mathrm{d}|\theta|(x)<\infty,\ \ k=-2,-1,0,1,\quad\text{with}\quad\int\mathrm{d}\theta\neq-\int x^{-2}\,\mathrm{d}\theta(x).$$

The section is closed with a table that collects the Lévy-Khinchine representations for the other SID random variables introduced in the text.

3.4 Weak Infinite Divisibility

In this section we examine the more pathological class WID which is, by definition, all GID random variables that are not strongly infinitely divisible. By their very nature examples of this class are not well behaved and difficult to describe; nevertheless some important ones are explained .

Name	Ch. Function	Lévy-Khinchine Representation	
		a	$\theta(x)$
$GPoi(\lambda)$	$\exp\{\lambda(e^{it}-1)\}$	$\lambda/2$	$\frac{\lambda}{2}1_0(x-1)$
$GNegB(r,p)$	$\left(\frac{p}{1-qe^{it}}\right)^r$	$\sum_{k=1}^{\infty}\frac{rq^k}{1+k^2}$	$\sum_{k=0}^{\infty}\frac{rkq^k}{1+k^2}\,1_0(x-k)$
$GBin(n,p)$	$(q+pe^{it})^n$	$\sum_{k=1}^{\infty}\frac{-n}{1+k^2}\left(\frac{p}{p-1}\right)^k$	$\sum_{k=0}^{\infty}\frac{-nk}{1+k^2}\left(\frac{p}{p-1}\right)^k 1_0(x-k)$
$GCP(\lambda,\mu)$	$\exp\{\lambda(\widehat{\mu}-1)\}$	$\lambda\int\frac{u}{1+u^2}\,\mathrm{d}\mu(u)$	$\lambda\int_{-\infty}^{x}\frac{u^2}{1+u^2}\,\mathrm{d}\mu(u)$

Example 3.4.1. *The $Unif(0,1)$ distribution is WID.*

Proof. It is known from Theorem 2.2.1 that $Unif(0,1)$ is GID, yet its characteristic function

$$\varphi(t)=\frac{e^{it}-1}{it}$$

is zero at the points $t=\pm 2\pi k,\ k=1,2,\ldots$; by Theorem 3.1.2 it must not be SID, hence it is in fact WID. □

Example 3.4.2. *The $GBern(1/2)$ distribution is WID.*

This was the example that showed that SID is not closed. It would be nice to be able to say something about $GBern(p)$ for $p>1/2$, but none of those random variables are even GID. However, one can say something more, taking as motivation the $GBern(1/2)$ example. First recall that a random variable X is said to have the *Discrete Uniform distribution* (denoted $DUnif(N)$) if

$$\mathbb{P}(X=k)=\frac{1}{N},\quad k=0,1,\ldots,N-1$$

for some specified positive integer $N\in\mathbb{N}$.

Example 3.4.3. *The $DUnif(N)$ distribution is WID for every $N\geq 2$.*

Proof. It is of course obvious that $X \sim DUnif(N)$ is GID because of Theorem 2.3.1. On the other hand, it is easy to see that the characteristic function of X,

$$\varphi_X(t) = \frac{1}{N} \sum_{k=0}^{N-1} e^{itk},$$

vanishes infinitely often. In particular, $t = 2\pi(k + 1/N)$ is a root for $k \geq 0$. □

One may easily construct numerous other WID distributions by convolving finitely many of the WID distributions already discussed.

CHAPTER 4
DE FINETTI'S THEOREM

4.1 Introduction

The roots of the idea of finite exchangeability of events can be traced back to the new "Algebra" in De Moivre [12] where the game "Rencontre" ("Matches") was analyzed (Problem XXXV. and Problem XXXVI.) In the Preface de Moivre writes enthusiastically, "In the 35th and 36th Problems I explain a new sort of Algebra,... I assure the Reader, that the Method I have followed has a degree of Simplicity, not to say of Generality,...".

Infinite sequences of exchangeable events seem to have first been discussed by Jules Haag [29]. Not long after, de Finetti [13] independently introduced exchangeable random variables and proved his famous representation theorem for the 2-valued case. More precisely: if $X_1, X_2, \ldots$ take values in $\{0,1\}$ and

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n) = \mathbb{P}(X_1 = x_{\pi(1)}, X_2 = x_{\pi(2)}, \ldots X_n = x_{\pi(n)})$$

holds for all $n \in \mathbb{N}$, $x_1, \ldots, x_n \in \{0,1\}$, and all finite permutations π of $\{1, 2, \ldots, n\}$, then for some measure μ on $[0,1]$

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n) = \int_0^1 p^{\sum x_i}(1-p)^{n-\sum x_i}\mathrm{d}\mu(p).$$

In words, $X_1, X_2, \ldots$ are conditionally i.i.d. given the random variable p, which is distributed according to the measure μ.

This theorem has been extended and generalized in various directions. De Finetti himself showed that his representation held for real-valued random variables [14]. Dynkin [21] replaced $\mathbb{R}$ with more general spaces that are separable in some sense, and Hewitt and

Savage [30] extended the result to compact Hausdorff spaces, from which it can readily be extended to, for example, Polish or locally compact spaces. The theorem does not hold without some topological assumptions, however, as was shown by Dubins and Freedman [20].

It is well known that de Finetti's theorem also does not hold in general for finite sequences of exchangeable random variables; for an easy example showing failure in case $n = 2$ see Diaconis [16], [17], or Spizzichino [43]. It is not difficult to see the trouble: suppose that $X_1, X_2, \ldots X_N$ are conditionally i.i.d. given some random variable θ which is distributed according to some probability measure ν. Then for $1 \leq i, j \leq N$,

$$\begin{aligned}
\mathbb{E}X_iX_j &= \mathbb{E}X_1X_2 \quad \text{(by exchangeability)} \\
&= \mathbb{E}_\nu\{\mathbb{E}[X_1X_2|\theta]\} \\
&= \mathbb{E}_\nu\{\mathbb{E}[X_1|\theta]\ \mathbb{E}[X_2|\theta]\} \quad \text{(conditional independence)} \\
&= \mathbb{E}_\nu\{\,(\mathbb{E}[X_1|\theta])^2\} \quad \text{(conditionally identically dist'd)}
\end{aligned}$$

And of course $\mathbb{E}X_i = \mathbb{E}_\nu\{\mathbb{E}[X_1|\theta]\} = \mathbb{E}X_j$, so that

$$\begin{aligned}
\mathrm{Cov}(X_i, X_j) &= \mathbb{E}_\nu\{\,(\mathbb{E}[X_1|\theta])^2\} - (\mathbb{E}_\nu\{\mathbb{E}[X_1|\theta]\})^2 \\
&= \mathrm{Var}_\nu(\mathbb{E}[X_1|\theta]) \\
&\geq 0, \quad \text{when } \nu \geq 0.
\end{aligned}$$

This gives immediately the familiar fact that the classical de Finetti representation works only for sequences that are nonnegatively correlated, and instances where this condition fails are often encountered in practice; take, for example, hypergeometric sequences.

In response to this problem there have been several versions and modifications of the theorem developed for the finite case. Kendall [33] (also de Finetti [15]) showed that every finite system of exchangeable events is equivalent to a random sampling scheme without replacement, where the number of items in the sampling has an arbitrary distribution. Diaconis [16] and Diaconis and Freedman [17] used this to find total variation distances to the closest mixture of i.i.d. random variables, (which turns out to imply de Finetti's theorem in the limit). Gnedin [27] explored conditions on the density of a finite exchangeable sequence and found a criterion for extendability to an infinite sequence. Several other aspects of finite exchangeability are in Diaconis and Freedman [18], von Plato [46], and Diaconis et al. [19]; Kingman [34] and Aldous [1] give useful surveys of exchangeability. See also Ressel [39].

The present Chapter reexamines the failure of de Finetti's theorem and in the process reveals that while for infinite exchangeable sequences which are nonnegatively correlated the classical (that is, nonnegative) mixtures of i.i.d. random variables are sufficient, with some finite sequences, we need to consider an extended notion of "mixture" to retain de Finetti's convenient representation. In particular, we have the

Theorem 4.1.1. *Let $(S,\mathcal{B})$ be an abstract measurable space and write S^* for the set of probability measures on $(S,\mathcal{B})$. Endow S^* with the smallest σ-field $\mathcal{B}^*$ making $p\mapsto p(A)$ measurable for all $A\in\mathcal{B}$. If $p\in S^*$, then p^n is the product measure on $(S^n,\mathcal{B}^n)$.*

If $\mathbb{P}$ is an exchangeable probability on $(S^n,\mathcal{B}^n)$ then there exists a signed measure ν such that

$$\mathbb{P}(A)=\int_{S^*}p^n(A)\,\mathrm{d}\nu(p),\qquad \text{for all } A\in\mathcal{B}^n.$$

Further, ν is of bounded variation and satisfies $\nu(S^*)=1$.

It is noted that Jaynes [31] independently discovered this theorem in the case where the random variables take only the values $\{0,1\}$. His proof is very concise, using Bernstein and Legendre polynomials. In that essay he alludes to the extension of the theorem to the abstract case, saying, "...A more powerful and abstract approach, which does not require us to go into all that detail, was discovered by Dr. Eric Mjolsness while he was a student of [mine]. We hope that, with its publication, the useful results of this representation will become more readily obtainable...". Alas, to the knowledge of the present author that publication never appeared.

In any case, this research is distinct from the above papers in a number of ways. First, this result holds for an arbitrary measurable space without topological assumptions, in symmetry to Diaconis and Freedman [17], but in contrast to Dubins and Freedman [20]. Such luxury is afforded because one deals with finite sequences and avoids the pathologies that arise with infinite product spaces and limits.

Second, the method of proof of the representation, while elementary, is of an entirely different character. For example, de Finetti [15] shows that finite exchangeable sequences are mixtures of hypergeometric processes, and then takes a weak limit as the sample size increases without bound. For us, there is no limit whatever; we have only finitely many exchangeable random variables to use. Similarly, Kendall [33] used the (Reversed) Martingale Convergence Theorem combined with the hypergeometric mixture representation to show the existence of a sigma algebra conditional on which $X_1, X_2, \ldots$ are i.i.d., again a limiting argument. Hewitt and Savage [30] start the analysis with an infinite sequence and derive its mixture representation; we can get no help from there. Ressel [39] uses concepts of Harmonic Analysis on semigroups and positive definite functions to find a unique Radon mixing measure for the infinite sequence. Our paper uses simple measure theory and some linear algebra.

Third, the conclusion of this chapter usefully complements what is currently known regarding finite exchangeable sequences. Indeed, Diaconis and Freedman [17], [25] showed that for a finite exchangeable real-valued sequence of length k that can be embedded in a sequence of length $N \geq k$, a classical mixture of an i.i.d. sequence is at most $2(1 - N_k/N^k)$ away (in terms of total variation distance), where $N_k = N(N-1)\cdots(N-k+1)$. This bound is sharp.

Now, suppose we have an exchangeable sequence that is only *slightly* negatively correlated, say $\rho = -0.01$. Then it is easy to see (in Kingman [34] or Spizzichino [43], for example) that N can be at most 100. One can see the reason why by considering exchangeable $X_1, X_2, \ldots, X_n$ with X_1 having finite nonzero variance σ^2 and correlation coefficient with X_2 being ρ. Then the variance of $\sum_{i=1}^n X_i$ is

$$\sum_{i=1}^{n} \sigma^2 + \sum_{i \neq j} \rho\sigma^2 = n\sigma^2[1 + (n-1)\rho],$$

and since the above quantity is nonnegative it follows

$$\rho \geq -(n-1)^{-1},$$

or in other words, $N \leq 100$.

The bound $2(1-100_k/100^k)$ monotonically increases as a function of k, and by the time $k = 12$, it has already passed the value 1, the maximum total variation distance between any two distributions being of course 2. From a practical standpoint this means that for even moderately sized samples, the present bound gives little insight. It is completely unclear how close the closest classical mixture is, and the situation only worsens as the variables become more correlated. Again, with an infinite sequence the above correlation inequality would hold for all $n \geq 1$, and in such a case ρ would therefore be nonnegative. Thus here is an alternative proof that a necessary condition for finite sequences to have the classical de Finetti representation is that they be nonnegatively correlated.

This chapter guarantees an exact representation, for any finite k, and regardless of the underlying correlation structure. In an application to Bayesian Theory, we show that by allowing the mixture to stand in with a Prior distribution one can formally justify common practical procedures; in addition, the resulting Posteriors continue to converge to degenerate distributions under usual regularity conditions, together with much wider assumptions than are ordinarily supposed.

We present a second application in the area of Statistical Physics. We demonstrate that the quantum statistics of Fermi-Dirac may be derived from the statistics of classical (*i.e.* independent) particles by means of a signed mixture of multinomial distributions, or Maxwell-Boltzmann statistics. This work continues in the vein of work by Bach, Blank and Francke [4], who performed a similar derivation of Bose-Einstein Statistics using the classical de Finetti theorem. In many ways the resulting asymmetry for Fermi-Dirac statistics is eliminated.

As a final remark, it should be pointed out that the use of signed measures in probability theory is by no means new; it has claimed even proponents such as Bartlett[6] and Feynman [23]. The existing literature is voluminous. For the purposes of this dissertation those results are not needed, but the interested reader could begin by consulting the survey by Mückenheim [37].

4.2 Finite Exchangeable Sequences

We start by extending some earlier results of Diaconis [16] from the space $\{0,1\}$ to the more general $\{0,1,\ldots,n\}$. Let $\mathcal{P}_n$ represent all probabilities on $\prod_{i=1}^{n} S_i$, where $S_i = \{s_0, s_1, \ldots, s_{n-1}\}$ for each i. Then $\mathcal{P}_n$ is a $n^n - 1$ dimensional simplex which is naturally embedded in Euclidean n^n space.

$\mathcal{P}_n$ can be coordinatized by writing $\mathbf{p} = (p_0, p_1, \ldots, p_{n^n-1})$ where p_j represents the probability of the outcome j, and $j = 0, 1, \ldots, n^n - 1$ is thought of as having its n-ary representation, written with n n-ary digits. Thus if $n = 3$, then $j = 5$ refers to the point 012. Let $\Omega_n(\mathbf{k}) = \Omega_n(k_0, k_1, \ldots, k_{n-1})$ be the set of j, $0 \le j < n^n$, with exactly k_0 digits 0, k_1 digits 1, $\ldots$, k_{n-1} digits $n-1$. The number of elements in $\Omega_n(\mathbf{k})$ is $n!/(k_0!k_1!\cdots k_{n-1}!)$.

Let $\mathcal{E}_n$ be the exchangeable measures in $\mathcal{P}_n$; then $\mathcal{E}_n$ is convex as a subset of $\mathcal{P}_n$.

Lemma 4.2.1. *$\mathcal{E}_n$ has $\binom{2n-1}{n}$ extreme points $\mathbf{h_0}, \mathbf{h_1}, \ldots, \mathbf{h}_{\binom{\mathbf{2n-1}}{\mathbf{n}}-\mathbf{1}}$, where $\mathbf{h_k}$ is the measure putting mass $k_0!k_1!\cdots k_{n-1}!/n!$ at each of the coordinates $i \in \Omega_n(\mathbf{k})$, and mass 0 elsewhere. Here, the index vector $\mathbf{k} = (k_0, k_1, \ldots, k_{n-1})$ runs over all possible distinct "urns", as described in the proof below.*

Each exchangeable probability $\mathbf{p} \in \mathcal{E}_n$ has a unique representation as a mixture of the $\binom{2n-1}{n}$ extreme points.

Proof. This is a direct generalization of Diaconis' [16] results to a higher dimensional setting. Intuitively, $\mathbf{h_k}$ is a column vector representing the measure associated with n drawings without replacement from the unique urn containing n balls, k_0 marked with s_0, k_1 marked with $s_1, \ldots,$ k_{n-1} marked with s_{n-1}; $\mathbf{h}$ stands for "hypergeometric". Each $\mathbf{h_k}$ is exchangeable, and finding the total number of distinct urns amounts to finding the number of ways to distribute n indistinguishable balls into n boxes marked $s_0, s_1, \ldots, s_{n-1}$ respectively. It is well known that the number of ways to distribute r balls into m (ordered) boxes is just $\binom{m+r-1}{r}$. In this case, $m = r = n$.

Note that when $S = \{s_0, s_1\} = \{0, 1\}$, this matches Diaconis' [16] result where $\mathcal{E}_n$ had $n+1$ extreme points, because in that case $m = 2$ and $r = n$, so that

$$\binom{2+n-1}{n} = \binom{n+1}{n} = n+1.$$

Now suppose that $\mathbf{h_k}$ is a mixture of two other exchangeable measures, that is $\mathbf{h_k} = p\mathbf{a} + (1-p)\mathbf{b}$, where $\mathbf{a}$ and $\mathbf{b}$ are exchangeable and $0 < p < 1$. Then $\mathbf{a}$ and $\mathbf{b}$ put mass zero at every point where $\mathbf{h_k}$ puts mass zero, namely, on the complement of $\Omega_n(\mathbf{k})$.

By exchangeability, outcomes with the same number of s_0's, s_1's, … s_{n-1}'s have the same probability. Therefore, the entries of both $\mathbf{a}$ and $\mathbf{b}$ must be equal at all coordinates $j \in \Omega_n(\mathbf{k})$. But the sum of the entries is 1 in each vector, so the mass at each $j \in \Omega_n(\mathbf{k})$ must be $k_0!k_1!\cdots k_{n-1}!/n!$. This implies $\mathbf{a} = \mathbf{b}$, and $\mathbf{h_k}$ is in fact an extreme point of $\mathcal{E}_n$.

It is well known that every point of a simplex has a unique representation as a mixture of extreme points. □

Now we see that $\mathcal{E}_n$ is a $\binom{2n-1}{n}$ sided polyhedron. The extreme points $\mathbf{h_0}, \mathbf{h_1}, \ldots, \mathbf{h}_{\binom{\mathbf{2n-1}}{\mathbf{n}}-\mathbf{1}}$ are linearly independent because they are supported on the disjoint sets $\Omega_n(\mathbf{j}), \mathbf{j} = \mathbf{0}, \ldots, \binom{\mathbf{2n-1}}{\mathbf{n}} - \mathbf{1}$. And, a probability $\mathbf{p}$ is in $\mathcal{E}_n$ if and only if it is constant on the sets $\Omega_n(\mathbf{k})$.

An interesting subclass is the class $\mathcal{M}_n \subset \mathcal{E}_n$ of i.i.d. probabilities on $\mathcal{P}_n$. $\mathcal{M}_n$ can be parameterized as an n^{th} degree polynomial of $n-1$ variables:

$$\mathbf{m} = (p_0^n, p_0^{n-1}p_1, \ldots, p_{n-2}p_{n-1}^{n-1}, p_{n-1}^n).$$

In general, for $j \in \Omega_n(\mathbf{k})$, we have $\mathbf{m}_j = \prod_{i=0}^{n-1} p_i^{k_i}$, where $p_i \geq 0,\ i = 0, \ldots, n-1$, and $p_0 + p_1 + \cdots + p_{n-1} = 1$. The set $\mathcal{M}_n$ is a smooth surface that twists through $\mathcal{E}_n$.

The class of mixtures of i.i.d. probabilities (with which de Finetti's Theorem is associated) is a convex set lying in $\mathcal{E}_n$. Measures in this set can be represented in uncountably many ways.

The above has similarly been a many-dimensional analogue of Diaconis [16], but we break new territory with the following

Proposition 4.2.1. *Each exchangeable probability* $\mathbf{p} \in \mathcal{E}_n$ *can be written as a (possibly) signed mixture of measures* $\{\mathbf{v}_i\} \subset \mathcal{M}_n$*.*

Proof. Consider the $\binom{2n-1}{n} - 1$ dimensional hyperplane $\mathcal{H}$ of Euclidean n^n space that is determined by the linearly independent hypergeometric vectors $\{\mathbf{h_k}\}$, and consider the surface $\mathcal{M}_n \subset \mathcal{H}$. We choose $\binom{2n-1}{n}$ linearly independent vectors $\{\mathbf{v}_i : i = 1, \ldots, \binom{2n-1}{n}\}$ in the uncountable collection $\mathcal{M}_n$ to construct a basis for $\mathcal{H}$. Then any $\mathbf{p} \in \mathcal{E}_n \subset \mathcal{H}$ can be written as a (possibly) signed linear combination $\mathbf{p} = \sum_i a_i \mathbf{v}_i$. The fact that $\sum_i a_i = 1$ follows from the observation that $\mathbf{v}_i(\prod_{j=1}^n S_j) = 1$ for all i, so that $\mathbf{p}(\prod_{j=1}^n S_j) = \sum_i a_i = 1$. $\square$

In symmetry to the above we may, as Diaconis [17] did, define the column vector $\mathbf{m_k}$ to represent the measure associated with n drawings with replacement from the urn containing k_0 balls marked with s_0, k_1 marked with $s_1, \ldots,$ k_{n-1} marked with s_{n-1}; $\mathbf{m}$ stands for "multinomial". From our Proposition 4.2.1 follows the

Corollary 4.2.1. *Each extreme point* $\mathbf{h_k}$ *can be written as a signed mixture of the multinomial measures* $\{\mathbf{m_j}\}$*.*

Proof. In essence, we have only chosen $\{\mathbf{m_j}\}$ as a particular basis for the hyperplane $\mathcal{H}$. For each $\mathbf{k}$, the measure $\mathbf{m_k}$ is exchangeable, so it is (from Lemma 4.2.1) a unique mixture of the measures $\{\mathbf{h_j}\}$. That is, there exist nonnegative weights $w_0, w_1, \ldots, w_{\binom{2n-1}{n}-1}$ that sum to one satisfying

$$\mathbf{m_k} = \sum_{j=0}^{\binom{2n-1}{n}-1} w_j \, \mathbf{h_j}.$$

A moment's reflection will convince us that the weights are *exactly* the ordinary multinomial distribution; in fact, we can display them explicitly in the form

$$\mathbf{m_k} = \sum_{\mathbf{j}} \binom{n}{j_0\, j_1 \cdots j_{n-1}} \left(\frac{k_0}{n}\right)^{j_0} \cdots \left(\frac{k_{n-1}}{n}\right)^{j_{n-1}} \cdot \mathbf{h_j},$$

where we interpret $0^0 = 1$ and the summation runs over the $\binom{2n-1}{n}$ indices $\mathbf{j}$.

Letting $\mathbf{w_k}$ be the column vector containing the weights associated with $\mathbf{m_k}$ we can write in matrix notation $\mathbf{M} = \mathbf{HW}$, where

$$\mathbf{M} = \left[\ \mathbf{m_0}\ \mathbf{m_1}\ \cdots\ \mathbf{m}_{\binom{\mathbf{2n-1}}{\mathbf{n}}-\mathbf{1}}\ \right],\ \mathbf{H} = \left[\ \mathbf{h_0}\ \mathbf{h_1}\ \cdots\ \mathbf{h}_{\binom{\mathbf{2n-1}}{\mathbf{n}}-\mathbf{1}}\ \right],\ \textit{and}$$

$$\mathbf{W} = \left[\ \mathbf{w_0}\ \mathbf{w_1}\ \cdots\ \mathbf{w}_{\binom{\mathbf{2n-1}}{\mathbf{n}}-\mathbf{1}}\ \right].$$

Now, the matrix $\mathbf{W}$ is invertible; indeed, $\mathbf{W}$ is merely the change of basis matrix from the basis $\{\mathbf{h_j}\}$ to the basis $\{\mathbf{m_j}\}$. Linear independence of the $\mathbf{w_j}$'s follows from the observation that any column vector $\boldsymbol{\beta}$ satisfying $\mathbf{W}\,\boldsymbol{\beta} = \mathbf{0}$ also satisfies $\mathbf{M}\,\boldsymbol{\beta} = \mathbf{H}\,(\mathbf{W}\,\boldsymbol{\beta}) = \mathbf{0}$. And the vectors $\mathbf{m_k}$, each corresponding to distinct urns $\mathbf{k}$, are easily verified to be linearly independent, hence $\boldsymbol{\beta}$ must be $\mathbf{0}$ and invertibility of $\mathbf{W}$ follows.

The required representation is obtained by inverting $\mathbf{W}$ to yield $\mathbf{H} = \mathbf{M}\,\mathbf{W}^{-1}$, which gives each extreme point $\mathbf{h_k}$ to be the mixture of the vectors $\{\mathbf{m_j}\}$, with the weights being the corresponding $\mathbf{k}$-th column of $\mathbf{W}^{-1}$. □

Note that in the above proof no claim was made either way regarding the sign of the weights of $\mathbf{W}^{-1}$. The fact of the matter is that except in the degenerate cases an extreme point $\mathbf{h_k}$ will be a signed mixture of the $\mathbf{m_j}$'s. For an arbitrary exchangeable measure nothing can in general be said. The deciding factor is the vector's location in $\mathcal{E}_n$; if it happens to fall in the convex set of mixtures of i.i.d. vectors in $\mathcal{M}_n$, then its representation will be classical. If it falls too close to the extreme points $\mathbf{h_k}$, then its existing mixture representation is necessarily signed.

Now we go to make precise that which has been mentioned above, together with direct examples of the geometric reasoning used. An illuminating example occurs already in case $n = 2$. Consider sampling without replacement from an urn containing two balls, one marked 0 and the other marked 1. Here the exchangeable random variables $X_1,\ X_2$ satisfy

$$\mathbb{P}(X_1 = 1, X_2 = 0) = \mathbb{P}(X_1 = 0, X_2 = 1) = \frac{1}{2},$$

$$\mathbb{P}(X_1 = 0, X_2 = 0) = \mathbb{P}(X_1 = 1, X_2 = 1) = 0.$$

Suppose for the moment that there existed a nonnegative mixing measure μ for this case. Then one would have

$$0 = \mathbb{P}(X_1 = 1, X_2 = 1) = \int_0^1 p^2 \mathrm{d}\mu(p),$$

implying that μ puts mass 1 at the point $p = 0$, but on the other hand,

$$0 = \mathbb{P}(X_1 = 0, X_2 = 0) = \int_0^1 (1-p)^2 \mathrm{d}\mu(p),$$

which implies that μ puts mass 1 at the point $p = 1$, which is impossible. This particular example is the one Diaconis [16], [17] used to display the inadequacy of de Finetti's theorem when applied to certain finite sequences of exchangeable random variables.

Now consider the space $\mathcal{P}_2$ of all possible assignments of probabilities to the four events $\{X_1 = x_1, X_2 = x_2\}$, $x_i = 0, 1$. That is, all probabilities on $\{0,1\}^2$. $\mathcal{P}_2$ is a 3 dimensional tetrahedron, which is embedded in $\mathbb{R}^4$.

These probabilities are represented as the set of all points $\mathbf{p} = (p_0, p_1, p_2, p_3)$, where $p_i \geq 0$ and $p_0 + p_1 + p_2 + p_3 = 1$. Set

$$p_0 = \mathbb{P}(X_1 = 0, X_2 = 0),$$

$$p_1 = \mathbb{P}(X_1 = 0, X_2 = 1),$$

$$p_2 = \mathbb{P}(X_1 = 1, X_2 = 0),$$

$$p_3 = \mathbb{P}(X_1 = 1, X_2 = 1).$$

(Notice that p_j represents the probability of the outcome j, $j = 0, 1, 2, 3$, written in binary notation.) See Figure 1, page 82, as borrowed from Diaconis [16]. Here we have the sets

$$\Omega_2(2,0) = \{0\}, \quad \Omega_2(1,1) = \{1,2\},$$

$$\Omega_2(0,2) = \{3\}.$$

The subclass $\mathcal{E}_2$ of exchangeable probabilities is the set of all $\mathbf{p}$ where $p_1 = p_2$, that is, the set of all $\mathbf{p}$ which are constant on the sets $\Omega_2(\mathbf{k})$, $\mathbf{k}$ =(2,0), $(1,1)$, $(0,2)$. $\mathcal{E}$ is convex as a subset of $\mathcal{P}_2$; in fact, $\mathcal{E}_2$ is the triangle with vertices $(1,0,0,0)$, $(0,1/2,1/2,0)$, and $(0,0,0,1)$. The troublesome exchangeable point $(0,1/2,1/2,0)$ is indicated on the figure as well.

The vertices of the triangle are the hypergeometric vectors (extreme points) $\mathbf{h_k}$. In particular,

$$\mathbf{h}_{(2,0)} = (1,0,0,0), \quad \mathbf{h}_{(1,1)} = (0,1/2,1/2,0), \text{ and}$$

$$\mathbf{h}_{(0,2)} = (0,0,0,1).$$

It is clear that the $\mathbf{h_k}$ are linearly independent, since they are supported on the disjoint sets $\Omega_2(\mathbf{k})$.

Moving to the class $\mathcal{M}_2$, we recognize it as the parabola $((1-p)^2, (1-p)p, p(1-p), p^2)$ parametrized in p for the values $0 \le p \le 1$, where of course p would represent $\mathbb{P}(X_1 = 1)$. This situation is shown in Figure 2, page 82, also from Diaconis [16]. As special cases we identify the multinomial vectors

$$\mathbf{m}_{(2,0)} = (1,0,0,0), \quad \mathbf{m}_{(1,1)} = (1/4,1/4,1/4,1/4), \text{ and}$$

$$\mathbf{m}_{(0,2)} = (0, 0, 0, 1).$$

Notice that $\mathbf{m}_{(2,0)} = \mathbf{h}_{(2,0)}$ and $\mathbf{m}_{(0,2)} = \mathbf{h}_{(0,2)}$; in the degenerate cases it does not matter whether one samples with or without replacement.

Now consider the 2-dimensional plane $\mathcal{H}$ of Euclidean 4-space that is spanned by the vectors $\mathbf{h_k}$. Since $\mathcal{E}_2 \subset \mathcal{H}$, then any exchangeable $\mathbf{p}$ may be written as a (possibly signed) linear combination of the $\mathbf{h_k}$'s. For the exchangeable vectors $\mathbf{m_k}$ this mixture takes the form

$$\begin{pmatrix} \mathbf{m_{(2,0)}} & \mathbf{m_{(1,1)}} & \mathbf{m_{(0,2)}} \end{pmatrix} = \begin{pmatrix} \mathbf{h_{(2,0)}} & \mathbf{h_{(1,1)}} & \mathbf{h_{(0,2)}} \end{pmatrix} \cdot \mathbf{W}$$

where

$$\mathbf{W} = \begin{pmatrix} 1 & 0 & 1/4 \\ 0 & 1 & 1/4 \\ 0 & 0 & 1/2 \end{pmatrix}$$

Of course, the multinomial $\mathbf{m_k}$'s lie also in $\mathcal{H}$, and further span the subspace. By a change of basis the $\mathbf{h_k}$'s are signed mixtures of the $\mathbf{m_k}$'s. In other words $\mathbf{H} = \mathbf{M}\mathbf{W}^{-1}$, where

$$\mathbf{W}^{-1} = \begin{pmatrix} 1 & 0 & -1/2 \\ 0 & 1 & -1/2 \\ 0 & 0 & 2 \end{pmatrix}.$$

The two negative entries are evidence that when random variables are negatively correlated we must resort to extended notions of mixtures to retain de Finetti's convenient representation.

4.3 Proof of the Theorem

Proof. For $\boldsymbol{s} = (s_0, s_1, \ldots, s_{n-1}) \in S^n$, let $U(\boldsymbol{s})$ be the urn consisting of n balls, marked $s_0, s_1, \ldots, s_{n-1}$ respectively. Let $H_{U(\boldsymbol{s})}$ be the distribution of n draws made at random without replacement from $U(\boldsymbol{s})$. Thus, $H_{U(\boldsymbol{s})}$ is a probability on $(S^n, \mathcal{B}^n)$. The map $\boldsymbol{s} \mapsto H_{U(\boldsymbol{s})}(A)$ is measurable on $(S^n, \mathcal{B}^n)$ for each $A \in \mathcal{B}^n$.

Exchangeability of $\mathbb{P}$ entails that

$$\mathbb{P}(A) = \int_{S^n} H_{U(\boldsymbol{s})}(A)\, \mathbb{P}(\mathrm{d}\boldsymbol{s}).$$

This is true because $H_{U(\boldsymbol{s})}$ is the measure placing mass $1/n!$ at the $n!$ points which are permutations of $\boldsymbol{s} = (s_0, s_1, \ldots, s_{n-1})$, in which case $H_{U(\boldsymbol{s})}(A) = 1/n! \sum_\pi 1_A(\pi\boldsymbol{s})$, where π is a permutation of the first n positive integers and the summation extends over the $n!$ such permutations π. Next we calculate

$$\begin{aligned}
\int_{S^n} H_{U(\boldsymbol{s})}(A)\, \mathbb{P}(\mathrm{d}\boldsymbol{s}) &= \int_{S^n} \frac{1}{n!} \sum_\pi 1_A(\pi\boldsymbol{s})\, \mathbb{P}(\mathrm{d}\boldsymbol{s}) \\
&= \frac{1}{n!} \sum_\pi \int_{S^n} 1_A(\pi\boldsymbol{s})\, \mathbb{P}(\mathrm{d}\boldsymbol{s}) \\
&= \frac{1}{n!} \sum_\pi \int_{S^n} 1_A(\boldsymbol{s})\, \mathbb{P}(\mathrm{d}\boldsymbol{s}) \quad \text{(by exchangeability)} \\
&= \frac{1}{n!} \cdot n!\, \mathbb{P}(A) \\
&= \mathbb{P}(A).
\end{aligned}$$

Furthermore, from our Corollary 4.2.1 we may write

$$H_{U(\boldsymbol{s})}(A) = \int_{S^*} p^n(A)\, \mu(\boldsymbol{s}, \mathrm{d}p),$$

where $\mu(\boldsymbol{s}, \cdot)$ is a signed measure on S^*, the space of probability measures on $(S, \mathcal{B})$, and $p^n(A)$ is the probability of getting an outcome in A when doing an n-length i.i.d. experiment based on probability measure $p \in S^*$. We may think of μ as a signed transition kernel.

Thus, we may write

$$\begin{aligned}
\mathbb{P}(A) &= \int_{S^n} H_{U(\boldsymbol{s})}(A)\, \mathbb{P}(\mathrm{d}\boldsymbol{s}) \\
&= \int_{S^n} \left(\int_{S^*} p^n(A)\, \mu(\boldsymbol{s}, \mathrm{d}p) \right) \mathbb{P}(\mathrm{d}\boldsymbol{s}) \\
&= \int_{S^*} p^n(A)\, \nu(\mathrm{d}p),
\end{aligned}$$

where ν is the signed measure on S^* defined by

$$\nu(B) = \int_{S^n} \mu(\boldsymbol{s}, B)\, \mathbb{P}(\mathrm{d}\boldsymbol{s}), \quad B \in \mathcal{B}^*.$$

Of course, $\nu(S^*) = \int_{S^n} \mu(\boldsymbol{s}, S^*)\, \mathbb{P}(\mathrm{d}\boldsymbol{s}) = \int_{S^n} \mathbb{P}(\mathrm{d}\boldsymbol{s}) = 1$.

There are two remaining details to check. First, μ should have the right measurability properties, and second, ν should be of bounded variation.

It is clear on the one hand that for each $\boldsymbol{s}$, the function $\mu(\boldsymbol{s}, \cdot)$ is a measure on S^*. On the other hand, the claim that for each $B \in \mathcal{B}^*$, the function $\mu(\cdot, B)$ is a measurable function of $\boldsymbol{s}$ is not quite as clear. To see why this is the case, it is useful to examine the explicit form of the signed measure μ guaranteed by Corollary 4.2.1. It is defined by the formula

$$\mu(\boldsymbol{s}, B) = \sum_{k=1}^{\binom{2n-1}{n}} w_k\, 1_B(p^{(k)}),$$

where the (possibly negative) weights w_k are from the matrix $\mathbf{W}^{-1}$, and the measures $p^{(k)}$ are elements of S^* defined by

$$p^{(k)}(A) = \sum_{i=0}^{n-1} p_i^{(k)}\, 1_A(s_i), \quad \text{for } A \in \mathcal{B}.$$

The numbers $p_i^{(k)}$ are nonnegative and sum to one.

Since $\mu(\cdot, B)$ is a linear combination of indicator functions, it suffices to show that for fixed k the function $1_B(p^{(k)})$ is a measurable function of $\boldsymbol{s}$, or in other words, we just need to verify that the set $\{\boldsymbol{s} : p^{(k)} \in B\}$ is $\mathcal{B}^n$ measurable. Further, we remember that we endowed S^* with the weak* σ-algebra $\mathcal{B}^*$, generated by the class of sets $\{p : p(A) < t\}$, as A ranges over $\mathcal{B}$ and t over $[0,1]$. Thus, it is only necessary to confirm the measurability for such "nice" sets $B \in \mathcal{B}^*$. After these simplifications we find

$$\{\boldsymbol{s} : p^{(k)} \in B\} = \{\boldsymbol{s} : \sum p_i^{(k)} 1_A(s_i) < t\},$$

and this set is $\mathcal{B}^n$ measurable because the function $g(\boldsymbol{s}) = \sum p_i 1_A(s_i)$ is a measurable function of $\boldsymbol{s}$ when $A \in \mathcal{B}$.

To finish the proof we establish that ν is of bounded variation. First fix $\boldsymbol{s}$, let $\mathbf{W}^{-1} = (w_{ij})$ be as in Corollary 4.2.1 and notice

$$\|\mu(\boldsymbol{s}, \cdot)\|_{var} = \|\sum w_k\, \delta_{p^{(k)}}\|_{var} \leq \sum_{i,j} |w_{ij}| := M_n,$$

where M_n is a finite constant depending only on n. At last,

$$\|\nu\|_{var} \leq \int \|\mu(\boldsymbol{s}, \cdot)\|_{var}\, \mathbb{P}(\mathrm{d}\boldsymbol{s}) \leq \int M_n \mathrm{d}\mathbb{P} = M_n < \infty,$$

as required. $\square$

4.4 An Application to Bayesian Consistency

We turn to Bayesian Theory, where de Finetti's Theorem is commonly applied, for example in Predictive Inference. The standard setup involves some unknown (random) parameter θ, conditional on which a sequence of random variables $X_1, X_2, \ldots$ is distributed according to some family $\{f(\cdot|\theta) : \theta \in \Theta\}$ of p.d.f.'s indexed by θ, called the Likelihood family. In some situations θ may be interpreted as a strong law limit of a sequence of observations. The Bayesian has subjective beliefs about θ, represented by a Prior probability distribution $\pi(\theta)$ on Θ, the parameter space. The goal is to use the information contained in observations

$X_1, X_2, \ldots X_n$ to sequentially update the Prior distribution $\pi(\theta)$ to a Posterior distribution $\pi(\theta|\mathbf{x})$ via Bayes' Rule $\pi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\,\pi(\theta)$.

One hopes that with more and more information, one would become more confident about the location of θ, which would be reflected in the Posterior $\pi(\theta|\mathbf{x})$ by a concentration as $n \to \infty$ to a degenerate distribution centered at the true value of θ. It is a celebrated fact that under some regularity conditions such convergence does take place, see Berk [7], Le Cam [35], and Wald [47]. The usual method of proof is to suppose that the Likelihood may be written as a product of identical factors: $f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$. Then (depending on the particular proof) one takes logarithms, normalizes accordingly, and appeals to the Strong Law of Large Numbers to show that the Posterior does indeed concentrate as $n \to \infty$ to the true value of θ.

In this context the problem arises when one goes to use such a statement in practice; one is always confined by Nature to finite samples. As we have seen in the Introduction, under finite exchangeability the Likelihood certainly may *not* be written as a product of identical factors without first supposing that the sequence could be imbedded in an infinite exchangeable sequence and then applying de Finetti's theorem, that is, unless one asserts that the sequence could in principle continue indefinitely. In particular, this immediately restricts the Bayesian to observations which are nonnegatively correlated, and even then it is not guaranteed – in theory or practice – that a given sampling process could continue without end.

The theorem presented in this chapter shows that, with the introduction of an intermediate mixing generalized random variable β one may consider the sequence conditionally i.i.d. given β and θ without concern for the correlation or the conceptual difficulties associated with sampling from a "potentially infinite" exchangeable sequence.

A pointed criticism of this method would be that the resulting Likelihood $f(x|\beta,\theta)$ (possibly) takes negative values, which might seem unpleasant or unnatural. In fact, it seems as though we have traded one conceptual difficulty for another! However, it should be realized that the method is merely a mathematical technique meant to justify and explain what is strongly desired and commonly observed in practice: namely, the convergence of the sequence of Posteriors to a distribution degenerate at the true value of the parameter.

We go to demonstrate this fact, and for simplicity we prove it for the case when β and θ take at most a countable number of values by modifying an argument in Bernardo and Smith [8], a more general setting being much more technically cumbersome; see Berk [7]. Here we suppose that the mixing generalized random variable β takes countably many values $\beta_k,\ k = 1, 2, \ldots$, and the true parameter θ_{i^*} is distinguishable from the other values in the sense that the logarithmic divergences $\int |f(x|\beta_k, \theta_{i^*})| \log |f(x|\beta_k, \theta_{i^*})/f(x|\beta_l, \theta_i)| \,\mathrm{d}x$ are strictly positive for any $k \neq l$ or $i \neq i^*$.

Proposition 4.4.1. *Let $\mathbf{x} = (x_1, x_2, \ldots x_n)$ be exchangeable observations from a Likelihood family $\{f(\cdot|\theta) : \theta \in \Theta\}$, where $\Theta = \{\theta_1, \theta_2, \ldots\}$ is countable. Suppose that θ_{i^*} is the true value of θ and the following regularity conditions hold:*

1. *The Likelihood $f(\cdot|\beta_k, \theta_i) > 0$ w.p.1, for all $k, i \geq 1$.*

2. *The Prior satisfies $\pi(\theta_{i^*}) > 0$ and there exists k^* such that $g(\beta_{k^*}|\theta_{i^*}) \neq 0$, where g denotes the mass function of β.*

3. *The joint mass function $g(\beta, \theta)$ is of bounded variation: $\sum_{k,i} |g(\beta_k, \theta_i)| < \infty$.*

4. *For any $k \neq l$ or $i \neq i^*$,*

$$\int f(x|\beta_k, \theta_{i^*}) \log \left[\frac{f(x|\beta_k, \theta_{i^*})}{f(x|\beta_l, \theta_i)} \right] \mathrm{d}x > 0.$$

Then

$$\lim_{n\to\infty} \pi(\theta_{i^*}|\mathbf{x}) = 1, \quad \lim_{n\to\infty} \pi(\theta_i|\mathbf{x}) = 0, \; i \neq i^*.$$

Proof. By Bayes' Rule,

$$\pi(\theta_i|\mathbf{x}) = \frac{f(\mathbf{x}|\theta_i)\pi(\theta_i)}{\sum_m f(\mathbf{x}|\theta_m)\pi(\theta_m)}$$

But by Theorem 4.1.1 we may write

$$f(\mathbf{x}|\theta) = \sum_k \prod_{j=1}^n f(x_j|\beta_k, \theta) \cdot g(\beta_k|\theta),$$

where $f(\cdot|\beta,\theta) \geq 0$ and $g(\beta|\theta)$ is (possibly) signed. The above expression then becomes

$$\begin{aligned} \pi(\theta_i|\mathbf{x}) &= \frac{\sum_k \prod_{j=1}^n f(x_j|\beta_k, \theta_i)\, g(\beta_k|\theta_i)\pi(\theta_i)}{\sum_m \sum_l \prod_{j=1}^n f(x_j|\beta_l, \theta_m)\, g(\beta_l|\theta_m)\pi(\theta_m)} \\ &= \sum_k \frac{\prod_{j=1}^n f(x_j|\beta_k, \theta_i)\, g(\beta_k, \theta_i)}{\sum_{m,l} \prod_{j=1}^n f(x_j|\beta_l, \theta_m)\, g(\beta_l, \theta_m)}. \end{aligned}$$

By condition 2, there exists k^* such that $g(\beta_{k^*}, \theta_{i^*}) \neq 0$; consequently after using condition 1 and dividing numerator and denominator by $\prod_{j=1}^n f(x_j|\beta_{k^*}, \theta_{i^*})$ there is obtained

$$= \sum_k \frac{\exp\{S_{k,i}\} g(\beta_k, \theta_i)}{\sum_{m,l} \exp\{S_{l,m}\} g(\beta_l, \theta_m)}$$

where

$$S_{k,i} = \log\left(\prod_{j=1}^n \frac{f(x_j|\beta_k, \theta_i)}{f(x_j|\beta k^*, \theta_{i^*})}\right) = \sum_{j=1}^n \log \frac{f(x_j|\beta_k, \theta_i)}{f(x_j|\beta_{k^*}, \theta_{i^*})}.$$

Now, conditional on $(\beta_{k^*}, \theta_{i^*})$, $S_{k,i}$ is the sum of n i.i.d. random variables and therefore

$$\lim_{n\to\infty} \frac{1}{n} S_{k,i} = \int f(x|\beta_{k^*}, \theta_{i^*}) \log\left[\frac{f(x|\beta_k, \theta_i)}{f(x|\beta_{k^*}, \theta_{i^*})}\right] \mathrm{d}x,$$

by the Strong Law of Large Numbers. Condition 4 implies that the right hand side is negative for $k \neq k^*$ or $i \neq i^*$, and it of course equals zero for $k = k^*$, $i = i^*$; therefore as $n \to \infty$, $S_{k,i} \to -\infty$ for $k \neq k^*$ or $i \neq i^*$ and $S_{k^*,i^*} \to 0$. Proposition 4.4.1 now follows from Condition 3 and the Dominated Convergence Theorem. $\square$

4.5 An Application to Statistical Physics

Following Johnson, Kotz, Kemp [32], Bach [3], Constantini and Garibaldi [10], consider a physical system comprising a number n of particles of some kind, for example electrons, protons, or photons. Suppose that there are d states (energy levels) in which each particle can be. If X_i represents the state of particle i, $i = 1, \ldots, n$, then the overall state of the system is $(X_1, \ldots, X_n)$, and equilibrium is defined as the overall state with the highest probability of occurrence.

If all d^n arrangements are equally likely the system is said to behave according to **Maxwell-Boltzmann statistics** (MB), where "statistics" is used here in a sense meaningful to physicists. Assumptions on the system that lead to such behavior are:

1. The particles are identical in terms of physical properties but distinguishable in terms of position. This is equivalent to the statement that the particle size is small compared with the average distance between particles.

2. There is no theoretical limit on the fraction of the total number of particles in a given energy state, but the density of particles is sufficiently low and the temperature sufficiently high that no more than one particle is likely to be in a given state at the same time.

However, modern experimentation (particularly at low temperatures) has yielded two more plausible sets of hypotheses concerning physical systems; these result in **Bose-Einstein statistics** (BE) and **Fermi-Dirac statistics** (FE).

For both, one supposes that the particles are indistinguishable (thereby granting exchangeability). The BE statistics are obtained by retaining the second assumption that there is no limit on the number of particles that may occupy a particular energy state. Particles that are observed to obey BE statistics are called bosons and include photons, alpha particles, and deuterons.

For FD statistics one stipulates instead that only one particle can occupy a particular state at a given time (a condition well known as the Pauli exclusion principle). Particles that obey the principle are called fermions and include protons, neutrons, and electrons. All known elementary particles fall into one of the two above categories.

Much work has been done studying these different models and their consequences for the interpretation of the physical concepts. Constantini, Galavotti, and Rosa [11] proposed a new set of ground hypotheses for deriving the three models. Bach, Blank, and Francke [4], using the argument that exchangeable random variables are appropriate for describing indistinguishable particles, used a multivariate de Finetti's theorem to derive BE statistics. Bach [2] explored the quantum properties of indistinguishable particles, while Bach [3], Constantini and Garibaldi [10] attempted to base the derivation of the different statistics on the correlation structure and an introduced relevance quotient.

This section, in the spirit of Bach, Blank, and Francke [4], shows that the Fermi-Dirac statistics of indistinguishable (exchangeable) particles can be derived from the statistics of classical (*i.e.* independent) particles by means of Theorem 4.1.1.

More precisely, we are concerned with the statistical problem of distributing n particles into d cells. We introduce a probability space $(\Omega, \mathcal{B}, \mu)$ and random variables $X_i : \Omega \to \{1, \ldots, d\}$, where the event $\{X_i = j\}$ represents the outcome that particle i is in cell j. As remarked above, there are d^n different configurations, which are characterized by the events $\{\mathbf{X} = \mathbf{j}\}$, where $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{j} \in \{1, \ldots, d\}^n$.

For a given configuration $\{\mathbf{X} = \mathbf{j}\}$, we define the occupation numbers $n_k,\ k = 1, \ldots, d,$

by

$$n_k(\mathbf{j}) = \sum_{i=1}^{n} \delta_{k,j_i}.$$

For the particles of MB statistics we obtain

$$\mathbb{P}_{\mathrm{MB}}(\mathbf{X} = \mathbf{j}) = d^{-n} \qquad \mathbf{j} \in \{0, \ldots, n\}^d.$$

However, under BE statistics we have

$$\mathbb{P}_{\mathrm{BE}}(\mathbf{X} = \mathbf{j}) = \binom{n}{n_1(\mathbf{j}) \cdots n_d(\mathbf{j})}^{-1} \binom{n+d-1}{n}^{-1} \qquad \mathbf{j} \in \{0, \ldots, n\}^d.$$

For FD statistics, according to the Pauli exclusion principle we must assume $n \leq d$. Also, whenever there exists an occupation number n_k larger than 1 we must set the corresponding probability zero, with the remaining configurations being equally likely:

$$\mathbb{P}_{\mathrm{FD}}(\mathbf{X} = \mathbf{j}) = \begin{cases} 0 & \text{if } \mathbf{n}(\mathbf{j}) \notin \{0,1\}^d, \\ \binom{d}{n}^{-1} (n!)^{-1} & \text{if } \mathbf{n}(\mathbf{j}) \in \{0,1\}^d. \end{cases}$$

It is immediately clear that $\mathbb{P}_{\mathrm{MB}}$, $\mathbb{P}_{\mathrm{BE}}$, and $\mathbb{P}_{\mathrm{FD}}$ are exchangeable. It was shown in Bach, Blank, and Francke [4] that $\mathbb{P}_{\mathrm{BE}}$ is a nonnegative Dirichlet mixture of multinomial distributions, *i.e.* statistics corresponding to classical, independent particles. And in Bach [3] the correlation structures for the models were found to be

$$\mathrm{Corr}_{\mathrm{BE/FD}}(X_i, X_j) = \pm(d \pm 1)^{-1}$$

where of course $\mathrm{Corr}_{\mathrm{MB}}(X_i, X_j) = 0$. The negative correlation in FD statistics (due to the Pauli exclusion principle) shows that a classical mixture representation will not hold, and also suggests why no such derivation of them using classical particles has been done until now. However, by exchangeability we may immediately apply Theorem 4.1.1 to see that $\mathbb{P}_{\mathrm{FD}}$ is indeed a (necessarily signed) mixture, that is linear combination, of multinomial distributions.

Two remarks are in order. First, it is unnecessary to repeat all of the hard work present in the above papers to assert the mixture representation; it is a natural corollary of the study of this chapter. Secondly, an obvious (and perhaps tempting) question would concern the physical *interpretation* of the mixing generalized random variable and its signed distribution. Let it be clear that no such explanation is made or implied here. In all of the above mentioned papers there was no attempt to understand the physical significance of the mixture; throughout the goal was to find some way to mathematically base the modern, quantum particles on the more familiar classical particles of Maxwell-Boltzmann.

It seems that in the 1980's an asymmetry was created in the literature by Bach and his colleagues since such a basis was possible only for Bose-Einstein statistics, while the status of Fermi-Dirac statistics remained unresolved. It is hoped that the theorem of this chapter and the representation of this section may help to restore the symmetry to this long standing problem.

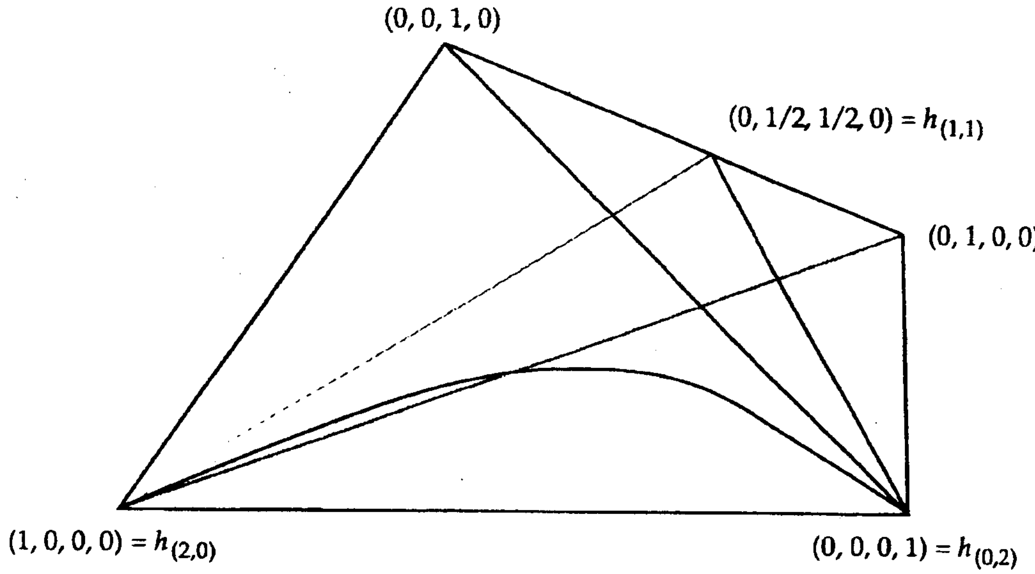

Figure 1: The set $\mathcal{E}_1$

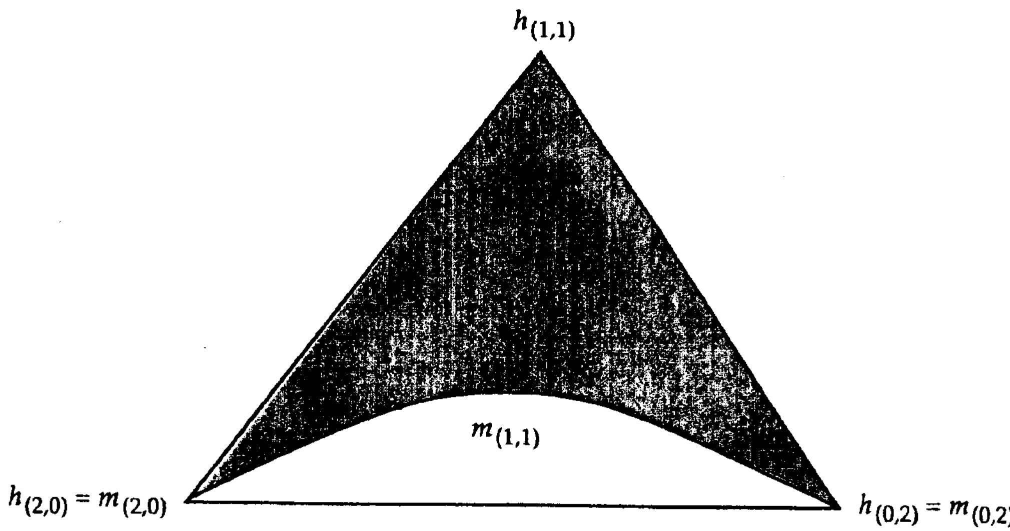

Figure 2: The set $\mathcal{M}_2$

REFERENCES

[1] Aldous, D.J. (1985). *Exchangeability and Related Topics.* Lecture Notes in Mathematics 117 Springer, New York.

[2] Bach, A. (1985). On the Quantum Properties of Indistinguishable Classical Particles. *Lettere al Nuovo Cimento* **43** 383-387.

[3] Bach, A. (1988). The Concept of Indistinguishable Particles in Classical and Quantum Physics. *Foundations of Physics* **18** 639-649.

[4] Bach, A., H. Blank, and H. Francke (1985). Bose-Einstein Statistics Derived from the Statistics of Classical Particles. *Lettere al Nuovo Cimento* **43** 195-198.

[5] Báez-Duarte, L. (1993). Central Limit Theorem for Complex Measures. *J. Theor. Prob.* **6** 33-56.

[6] Bartlett, M.S. (1945) Negative Probability. *Proc. Cambridge Phil. Soc.* **41** 71-73.

[7] Berk, R. (1970). Consistency a Posteriori. *Ann. Math. Stat.* **41** 894-906.

[8] Bernardo, J.M. and Smith, A.F.M. (1993). *Bayesian Theory.* Wiley, New York.

[9] Chung, K.L. (2001). *A Course in Probability Theory.* Academic Press, San Diego.

[10] Constantini, D. and U. Garibaldi (1989). Classical and Quantum Statistics as Finite Random Processes. *Foundations of Physics* **19** 743-754.

[11] Constantini, D., M. Galavotti, and R. Rosa (1983). A Set of Ground Hypotheses for Elementary-Particle Statistics. *Il Nuovo Cimento* **74** 151-158.

[12] De Moivre, A. (1718). *The Doctrine of Chances.* London.

[13] de Finetti, B. (1931). Funzione caratteristica di un fenomeno allatorio. *Atti della R. Accademia Nazionale dei Lincii Ser. 6, Memorie, Classe di Scienze, Fisiche, Matematiche e Naturali.* **4** 251-299.

[14] de Finetti, B. (1937). La prévision: ses logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* **7** 1-68.

[15] de Finetti, B. (1975). *Theory of Probability, Vol. 2.* Wiley, New York.

[16] Diaconis, P. (1977). Finite Forms of de Finetti's Theorem on Exchangeability. *Synthese* **36** 271-281.

[17] Diaconis, P and D. Freedman (1980). Finite Exchangeable Sequences. *Ann. Prob.* **8** 745-764.

[18] Diaconis, P and D. Freedman (1987). A Dozen de Finitti Style Results in Search of a Theory. *Annales de l'Institut Henri Poincaré* **23** 397-423.

[19] Diaconis, P.W., M.L. Eaton and S.L. Lauritzen (1992). Finite de Finetti Theorems in Linear Models and Multivariate Analysis. *Scand. J. Statist.* **19** 289-315.

[20] Dubins, L.E. and D. Freedman (1979). Exchangeable Processes Need Not Be Distributed Mixtures of Independent, Identically Distributed Random Variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **48** 115-132.

[21] Dynkin, E.B. (1953). Klassy ekvivalentnyh Slučaĭnyh veličin. *Uspehi Matematičeskih Nauk 8* **54** 125-134.

[22] Feller, W. (1967). A Direct Proof of Stirling's Formula. *Amer. Math. Monthly* **74** 1223-1225.

[23] Feynman, R.P. (1987). Negative Probability. *Quantum implications*, Routledge and Kegan Paul, London. 235-248.

[24] Folland, G.B. (1999). *Real Analysis: modern techniques and their applications.* Wiley, New York.

[25] Freedman, D. (1977). A Remark on the Difference Between Sampling With and Without Replacement. *J. Amer. Statist. Assoc.* **72** 681.

[26] Gnedenko, B.V. and A.N. Kolmogorov (1968). *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Massachusetts.

[27] Gnedin, A.V. (1996). On a Class of Exchangeable Sequences. *Statist. Probab. Lett.* **28** 159-164.

[28] González, M. (1992) *Classical Complex Analysis.* Marcel Dekker, New York.

[29] Haag, J. (1928). Sur un problème général de probabilitès et ses diverses applications. *Proceedings of the International Congress of Mathematicians, Toronto, 1924*, 659-674.

[30] Hewitt, E. and L.J. Savage (1955). Symmetric Measures on Cartesian Products. *Trans. Amer. Math. Soc.* **80** 470-501.

[31] Jaynes, E. (1986). Some Applications and Extensions of the De Finetti Representation Theorem. *Bayesian Inference and Decision Techniques*. Elsevier. 31-42.

[32] Johnson, K., S. Kotz, and A. Kemp (1993). *Univariate Discrete Distributions 2nd ed.* Wiley, New York.

[33] Kendall, D.K. (1967). On Finite and Infinite Sequences of Exchangeable Events. *Studia Sci. Math. Hung.* **2** 319-327.

[34] Kingman, J. (1978). Uses of Exchangeability. *Ann. Prob.* **6** 183-197.

[35] Le Cam, L. (1953). On some Asymptotic Properties of Maximum Likelihood Estimates and Related Bayes Estimates. *Univ. Calif. Publ. Statist.* **1** 277-300.

[36] Lukacs, E. (1970). *Characteristic Functions*. Griffin, London.

[37] Mückenheim, W. (1986). A Review of Extended Probabilities. *Phys. Rep. 133* **6** 337-401.

[38] O'Brien, G. and F. Steutel (1981). Divisibility Properties of Lebesgue Measure. Proceedings of Koninklijke Nederlandse Akademie Van Wetenschappen Series A **84** 393-398.

[39] Ressel, P. (1985). de Finetti-type Theorems: an Analytical Approach. *Ann. Prob.* **13**, 898-922.

[40] Robbins, H. (1955). A Remark on Stirling's Formula. *Amer. Math Monthly* **62**, 26-28.

[41] Ruzsa, I. and G. Székely (1988) *Algebraic Probability Theory*. Wiley, New York.

[42] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions.* University Press, Cambridge.

[43] Spizzichino, F. (2001). *Subjective Probability Models for Lifetimes.* Monographs on Statistics and Applied Probability 91, Chapman and Hall, New York.

[44] Székely, G.J. (2003). Mystical Half Coins. *unpublished*, 6-7.

[45] Varadarajan, V.S. (1961). Measures on Topological Spaces. *Amer. Math. Soc. Translation Series 2* **48**, 161-228.

[46] von Plato, J. (1991). Finite Partial Exchangeability. *Statist. Probab. Lett.* **11** 99-102.

[47] Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.* **20** 595-601.

[48] Widder, D. (1946) *The Laplace Transform.* Princeton, Univ. Press.

Appendix A
COMPLEX MEASURE THEORY

A.1 Signed and Complex Measures

Definition A.1.1. *Let $(\Omega, \mathcal{A})$ be a measurable space. A* ***signed measure*** *on $(\Omega, \mathcal{A})$ is a function $\nu : \mathcal{A} \to [-\infty, \infty]$ such that*

1. *$\nu(\emptyset) = 0$,*

2. *ν assumes at most one of the values $\pm\infty$,*

3. *if $\{E_j\}$ is a sequence of disjoint sets in $\mathcal{A}$, then $\nu\left(\bigcup_1^\infty E_j\right) = \sum_1^\infty \nu(E_j)$, where the latter sum converges absolutely if $\nu\left(\bigcup_1^\infty E_j\right)$ is finite.*

Thus every measure is a signed measure; for emphasis we will sometimes refer to measures as **positive measures**. If ν is a signed measure on $(\Omega, \mathcal{A})$, a set $E \in \mathcal{A}$ is called **positive** (resp. **negative**, **null**) for ν if $\nu(F) \geq 0$ (resp. $\nu(F) \leq 0,\ \nu(F) = 0$) for all $F \in \mathcal{A}$ such that $F \subset E$.

Theorem A.1.1 (The Hahn Decomposition Theorem). *If ν is a signed measure on $(\Omega, \mathcal{A})$, there exist a positive set P and a negative set N for ν such that $P \cup N = \Omega$ and $P \cap N = \emptyset$. If P', N' is another such pair, then $P \triangle P'$ $(= N \triangle N')$ is null for ν.*

We say that two signed measures μ and ν on $(\Omega, \mathcal{A})$ are **mutually singular** if there exist $E, F \in A$ such that $E \cap F = \emptyset,\ E \cup F = \Omega,\ E$ is null for μ, and F is null for ν. Informally speaking, mutual singularity means that μ and ν “live on different sets”. We express this relationship symbolically by writing $\mu \perp \nu$.

Theorem A.1.2 (The Jordan Decomposition Theorem). *If ν is a signed measure, then there exist unique positive measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$.*

The measures ν^+ and ν^- are called the **positive** and **negative variations** of ν, and $\nu = \nu^+ - \nu^-$ is called the **Jordan decomposition** of ν. Furthermore, we define the **total variation** of ν to be the measure $|\nu|$ defined by

$$|\nu| = \nu^+ + \nu^-.$$

Integration with respect to a signed measure is defined in the obvious way. As usual, for a positive measure μ we denote

$$L^1(\mu) = \{f : \Omega \to \mathbb{C} \text{ such that } \int |f| \, \mathrm{d}\mu < \infty\},$$

we set $L^1(\nu) = L^1(\nu^+) \cap L^1(\nu^-)$, and for $f \in L^1(\nu)$ we define

$$\int f \, \mathrm{d}\nu = \int f \, \mathrm{d}\nu^+ - \int f \, \mathrm{d}\nu^-.$$

Proposition A.1.1. *Let ν, ν_1, and ν_2 be signed measures on $(\Omega, \mathcal{A})$.*

1. $L^1(\nu) = L^1(|\nu|)$.

2. *If* $f \in L^1(\nu)$, $\left|\int f \, \mathrm{d}\nu\right| \leq \int |f| \, \mathrm{d}|\nu|$.

3. *If* $E \in \mathcal{A}$, $|\nu|(E) = \sup\{|\int_E f \, \mathrm{d}\nu| : |f| \leq 1\}$.

4. *If* $E \in \mathcal{A}$, $|\nu|(E) = \sup\{\sum_1^n |\nu(E_j)| : n \in \mathbb{N}$, $E_1, \ldots, E_n$ *are disjoint, and* $\bigcup_1^n E_j = E\}$

5. *If* ν_1 *and* ν_2 *both omit the value* $\pm\infty$, *then* $|\nu_1 + \nu_2| \leq |\nu_1| + |\nu_2|$.

Definition A.1.2. *Let* $(\Omega, \mathcal{A})$ *be a measurable space. A* ***complex measure*** *on* $(\Omega, \mathcal{A})$ *is a map* $\nu : \mathcal{A} \to \mathbb{C}$ *such that*

1. $\nu(\emptyset) = 0$,

2. *if* $\{E_j\}$ *is a sequence of disjoint sets in* $\mathcal{A}$*, then* $\nu\left(\bigcup_1^\infty E_j\right) = \sum_1^\infty \nu(E_j)$*, where the series converges absolutely.*

In particular, infinite values are not allowed, so a positive measure is a complex measure if and only if it is finite. If ν is a complex measure, we write ν_r and ν_i for the real and imaginary parts of ν. Thus ν_r and ν_i are signed measures that do not assume the values $\pm\infty$; hence they are finite and so the range of ν is a bounded subset of $\mathbb{C}$.

The notions developed so far generalize easily to the complex case. For example, we define $L^1(\nu) = L^1(\nu_r) \cap L^1(\nu_i)$, and for $f \in L^1(\nu)$ we define $\int f \,\mathrm{d}\nu = \int f \,\mathrm{d}\nu_r + i \int f \,\mathrm{d}\nu_i$. If μ and ν are complex measures we say $\nu \perp \mu$ if $\nu_a \perp \mu_b$ for $a, b = r, i$, and if λ is a positive measure we say $\nu \ll \lambda$ if $\nu_r \ll \lambda$ and $\nu_i \ll \lambda$.

Theorem A.1.3 (The Lebesgue-Radon-Nikodym Theorem). *If* ν *is a complex measure and* μ *is a* σ*-finite positive measure on* $(\Omega, \mathcal{A})$*, there exist a complex measure* λ *and an* $f \in L^1(\mu)$ *such that* $\lambda \perp \mu$ *and* $\mathrm{d}\nu = \mathrm{d}\lambda + f \,\mathrm{d}\mu$*. If also* $\lambda' \perp \mu$ *and* $\mathrm{d}\nu = \mathrm{d}\lambda' + f' \,\mathrm{d}\mu$*, then* $\lambda' = \lambda$ *and* $f' = f$ $\mu - a.e.$

The **total variation** of a complex measure ν is the positive measure $|\nu|$ determined by the property that if $\mathrm{d}\nu = f \,\mathrm{d}\mu$ where μ is a positive measure, then $\mathrm{d}|\nu| = |f| \,\mathrm{d}\mu$.

Proposition A.1.2. *Let* ν*,* ν_1*, and* ν_2 *be complex measures on* $(\Omega, \mathcal{A})$*.*

1. $|\nu(E)| \leq |\nu|(E)$ *for all* $E \in \mathcal{A}$*.*

2. $\nu \ll |\nu|$, *and* $\mathrm{d}\nu/\mathrm{d}|\nu|$ *has modulus 1* $|\nu|$*-a.s.*

3. $L^1(\nu) = L^1(|\nu|)$, *and if* $f \in L^1(\nu)$, *then* $\left|\int f \,\mathrm{d}\nu\right| \leq \int |f| \,\mathrm{d}|\nu|$.

4. If $E \in \mathcal{A}$, $|\nu|(E) = \sup\{\left|\int_E f \,\mathrm{d}\nu\right| : |f| \leq 1\}$.

5. $|\nu_1 + \nu_2| \leq |\nu_1| + |\nu_2|$.

A.2 Topological and L^p Spaces, Inequalities, and Complex Radon Measures

Let Ω be a locally compact Hausdorff (LCH) space and let $\mathcal{B}_\Omega$ denote the Borel σ-algebra on Ω, that is, the σ-algebra generated by the open sets of Ω.

We introduce some terminology. If $\mathcal{X}$ is any set, we denote by $B(\mathcal{X}, \mathbb{R})$ (resp. $B(\mathcal{X}, \mathbb{C})$) the space of all bounded real- (resp. complex-) valued functions on $\mathcal{X}$. If Ω is any topological space we have also the spaces $C(\Omega, \mathbb{R})$ and $C(\Omega, \mathbb{C})$ of continuous functions on Ω and we define

$$BC(\Omega, F) = B(\Omega, F) \cap C(\Omega, F) \quad (F = \mathbb{R} \text{ or } \mathbb{C}).$$

In speaking of complex-valued functions we shall usually omit the $\mathbb{C}$ and simply write $B(\Omega)$, $C(\Omega)$, and $BC(\Omega)$.

If $f \in B(\Omega)$, we define the **uniform norm** of f to be

$$\|f\|_u = \sup\{|f(\omega)| : \omega \in \Omega\}.$$

It is not hard to show that if Ω is a topological space then $BC(\Omega)$ is a closed subspace of $B(\Omega)$ in the metric induced by the uniform norm; in particular, $BC(\Omega)$ is complete.

For $f \in C(\Omega)$, the **support** of f, denoted $\mathsf{supp}(f)$, is the smallest closed set outside of which f vanishes, that is, the closure of $\{\omega : f(\omega) \neq 0\}$. If $\mathsf{supp}(f)$ is compact, we say that f is **compactly supported**, and we define

$$C_c(\Omega) = \{f \in C(\Omega) : \mathsf{supp}(f) \text{ is compact}\}.$$

Moreover, if $f \in C(\Omega)$, we say that f **vanishes at infinity** if for every $\epsilon > 0$ the set $\{\omega : |f(\omega)| \geq \epsilon\}$ is compact, and we define

$$C_0(\Omega) = \{f \in C(\Omega) : f \text{ vanishes at infinity}\}.$$

Clearly $C_c(\Omega) \subset C_0(\Omega)$. In fact, $C_0(\Omega) \subset BC(\Omega)$, because for $f \in C_0(\Omega)$ the image of the set $\{\omega : |f(\omega)| \geq \epsilon\}$ is compact, and $|f| < \epsilon$ on its complement.

With respect to L^p spaces, if f is a measurable function on the measure space $(\Omega, \mathcal{B}, \mu)$ and $0 < p < \infty$, we define

$$\|f\|_p = (\mathbb{E}|f|^p)^{1/p}\,,$$

and we define

$$L^p(\Omega, \mathcal{B}, \mu) = \{f : \Omega \to \mathbb{C} \text{ such that } \|f\|_p < \infty\}.$$

Theorem A.2.1 (Hölder's Inequality). *Suppose $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$. Then for the complex measure ν on $\mathbb{R}^2$ and any two random variables X and Y,*

$$\mathbb{E}_{|\nu|}|XY| \leq \left(\mathbb{E}_{|\nu|}|X|^p\right)^{1/p} \left(\mathbb{E}_{|\nu|}|Y|^q\right)^{1/q}.$$

In particular, if $X \in L^p$ and $Y \in L^q$ then $XY \in L^1$, and equality holds iff

$\alpha|X|^p = \beta|Y|^q$ a.s., for some constants α, β with $\alpha\beta \neq 0$.

Remark. *We will have occasion to use Hölder's Inequality written in a slightly different form: for $X \sim \nu$ and measurable functions f and g,*

$$|\mathbb{E}||f(X)g(X)| \leq (|\mathbb{E}||f(X)|^p)^{1/p}\,(|\mathbb{E}||g(X)|^q)^{1/q}.$$

Theorem A.2.2 (Minkowski's Inequality). *If $1 \le p < \infty$ and $X, Y \in L^p(|\nu|)$ then*

$$\left(\mathbb{E}_{|\nu|}|X+Y|^p\right)^{1/p} \le \left(\mathbb{E}_{|\nu|}|X|^p\right)^{1/p} + \left(\mathbb{E}_{|\nu|}|Y|^p\right)^{1/p}.$$

Proposition A.2.1 (Folland). *If the complex measure ν satisfies $|\nu|(\mathbb{R}) < \infty$, then for $0 < p < q \le \infty$ we have $|\mathbb{E}||X|^p \le (|\mathbb{E}||X|^q)^{p/q} \cdot |\nu|(\mathbb{R})^{1-p/q}$.*

Proof. If $q = \infty$, this is obvious:

$$|\mathbb{E}||X|^p \le \|X\|_\infty^p \cdot \int \mathrm{d}|\nu| = \|X\|_\infty^p \cdot |\nu|(\mathbb{R}).$$

If $q < \infty$ we use the second Hölder's Inequality A.2.1 above, with conjugate exponents q/p and $q/(q-p)$:

$$|\mathbb{E}||X|^p = \int |x|^p \cdot 1 \,\mathrm{d}|\nu|(x) \le \||X|^p\|_{q/p}\|1\|_{q/(q-p)} = (|\mathbb{E}||X|^q)^{p/q} \cdot |\nu|(\mathbb{R})^{(q-p)/q}.$$

□

Theorem A.2.3 (Chebychev's Inequality). *If $X \sim \nu$ satisfies $|\mathbb{E}||X|^p < \infty$, $0 < p < \infty$, then for any $\epsilon > 0$,*

$$|\mathbb{P}|\left(|X| > \epsilon\right) \le \frac{|\mathbb{E}||X|^p}{\epsilon^p}.$$

Let μ be a Borel measure on Ω and E a Borel subset of Ω. The measure μ is called **outer regular** on E if

$$\mu(E) = \inf\{\, \mu(U) : U \supset E,\ U \text{ open} \,\}$$

and **inner regular** on E if

$$\mu(E) = \sup\{\, \mu(K) : K \subset E,\ K \text{ compact} \,\}.$$

A Borel measure that is outer and inner regular on all Borel sets is called **regular**.

Definition A.2.1. *A **Radon measure** on Ω is a Borel measure that is finite on all compact sets, outer regular on all Borel sets, and inner regular on all open sets.*

We next list some properties of Radon measures.

Proposition A.2.2 (Folland). *Suppose that μ is a finite Radon measure on $(\Omega, \mathcal{B})$ and $E \in \mathcal{B}$ is a Borel set.*

1. *μ is regular.*

2. *For every $\epsilon > 0$ there exist an open U and a closed F with $F \subset E \subset U$ and $\mu(U \backslash F) < \epsilon$;*

3. *There exist an F_σ set A and a G_δ set B such that $A \subset E \subset B$ and $\mu(B \backslash A) = 0$.*

4. *C_c is dense in $L^p(\mu)$ for $1 \leq p < \infty$.*

Theorem A.2.4 (Lusin's Theorem). *Suppose that μ is a Radon measure on Ω and $f : \Omega \to \mathbb{C}$ is a measurable function that vanishes outside a set of finite measure. Then for any $\epsilon > 0$ there exists $\phi \in C_c(\Omega)$ such that $\phi = f$ except on a set of measure $< \epsilon$. If f is bounded, ϕ can be taken to satisfy $\|\phi\|_u \leq \|f\|_u$.*

Definition A.2.2. *A **signed Radon measure** is a signed Borel measure whose positive and negative variations are Radon, and a **complex Radon measure** is a complex Borel measure whose real and imaginary parts are Radon. We denote the space of complex Radon measures on Ω by $M(\Omega)$, and for $\nu \in M(\Omega)$ we define*

$$\|\nu\| = |\nu|(\Omega),$$

where $|\nu|$ is the total variation of ν.

Remark. *On a second countable LCH space every complex Borel measure is Radon. This follows since complex measures are bounded.*

Proposition A.2.3 (Folland). *If ν is a complex Borel measure, then ν is Radon iff $|\nu|$ is Radon. Moreover, $M(\Omega)$ is a vector space and $\nu \to \|\nu\|$ is a norm on it.*

Notation. *By M_+ and $M_\pm$ we mean the subspaces of positive measures and signed measures on Ω, respectively, with the inherited total variation norm.*

Lemma A.2.1 (Scheffé's Lemma). *For complex probability distributions μ and ν, with respective densities $f, g \in L^1$, we have $\|f - g\|_1 = 2\|\mu - \nu\|$.*

Proof. Given $A \in \mathcal{B}(\mathbb{R})$, $\int(f - g) = 1 - 1 = 0$ implies

$$0 = \int_A (f - g) + \int_{A^c} (f - g), \quad \text{which implies}$$

$$\left|\int_A (f - g)\right| = \left|\int_{A^c} (f - g)\right|.$$

Then

$$\begin{aligned} 2|\mu(A) - \nu(A)| &= 2\left|\int_A (f - g)\right| = \left|\int_A (f - g)\right| + \left|\int_{A^c} (f - g)\right| \\ &\leq \int_A |f - g| + \int_{A^c} |f - g| = \|f - g\|_1, \end{aligned}$$

in which case

$$2\|\mu - \nu\| = \sup_{A \in \mathcal{B}(R)} |\mu(A) - \nu(A)| \leq \|f - g\|_1.$$

But there exists a set for which equality holds: let $A = \{f \geq g\}$. Then

$$\begin{aligned} 2|\mu(A) - \nu(A)| &= \left|\int_A (f - g)\right| + \left|\int_{A^c} (f - g)\right| \\ &= \int_A |f - g| + \int_{A^c} |g - f| = \|f - g\|_1, \end{aligned}$$

and we are finished. $\square$

A.3 Properties of FST's and L^1 Convolution

It is sometimes possible to go in reverse and invert the Fourier transform. If $f \in L^1(m)$, we define

$$f^{\vee}(x) = \widehat{f}(-x) = \int e^{2\pi itx} f(t)\, \mathrm{d}m(t).$$

Theorem A.3.1 (The Fourier Inversion Theorem). *If $f \in L^1$ and $\widehat{f} \in L^1$ then f agrees almost everywhere with a continuous function f_0 and $(\widehat{f})^{\vee} = \widehat{(f^{\vee})} = f_0$.*

Notation. *We denote by C^{∞} the space of all functions on $\mathbb{R}$ whose derivatives of order k exist and are continuous for any k and by $\mathcal{S}$ the **Schwartz space** consisting of those C^{∞} functions which, together with all their derivatives, vanish at infinity faster than any power of $|x|$. More precisely, for any $k, N \in \mathbb{N}$ we define*

$$\|f\|_{(N,k)} = \sup_{x \in \mathbb{R}} (1 + |x|)^N |f^{(k)}(x)|;$$

then

$$\mathcal{S} = \{\, f \in C^{\infty} : \|f\|_{(N,k)} < \infty \text{ for all } k, N \,\}.$$

Proposition A.3.1 (Folland). *$\mathcal{F}$ is an isomorphism of $\mathcal{S}$ onto itself.*

Theorem A.3.2 (Parseval's Identity). *If $f,\ g \in L^1(m)$ then*

$$\int f(x)\overline{g(x)}\, \mathrm{d}m(x) = \int \widehat{f}(t)\overline{\widehat{g}(t)}\, \mathrm{d}m(t).$$

Proposition A.3.2 (Folland). *The subspace C_c^{∞} (and hence also $\mathcal{S}$) is dense in L^p $(1 \le p < \infty)$ and in C_0.*

Definition A.3.1. *Let f and g be measurable functions on $\mathbb{R}$. The **convolution** of f and g is the function $f * g$ defined by*

$$f * g(x) = \int f(x - y) g(y)\, \mathrm{d}y,$$

for all x such that the integral exists.

Various conditions may be imposed on f and g to guarantee that $f * g$ is defined at least almost everywhere, for example if f is bounded and compactly supported then g may be any locally integrable function.

In order to generalize the concept of convolution it is necessary to use products of complex measures on product spaces, which is facilitated by Radon-Nikodym derivatives. Namely, if $\mu,\ \nu \in M(\mathbb{R})$, we define $\mu \times \nu \in M(\mathbb{R} \times \mathbb{R})$ by

$$\mathrm{d}(\mu \times \nu)(x, y) = \frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}(x)\frac{\mathrm{d}\nu}{\mathrm{d}|\nu|}(y)\, \mathrm{d}(|\mu| \times |\nu|)(x, y).$$

The last theorem is a recollection of the not so well known Leibnitz's Rule, an application of the Fundamental Theorem of Calculus and the Chain Rule.

Theorem A.3.3 (Leibnitz's Rule). *If $f(x, \theta)$, $a(\theta)$, and $b(\theta)$ are differentiable with respect to θ, then*

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\int_{a(\theta)}^{b(\theta)} f(x, \theta)\mathrm{d}x = f(b(\theta), \theta)\frac{\mathrm{d}}{\mathrm{d}\theta}b(\theta) - f(a(\theta), \theta)\frac{\mathrm{d}}{\mathrm{d}\theta}a(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta}f(x, \theta)\mathrm{d}x$$

Appendix B
MODES OF CONVERGENCE

In what follows we will be very much concerned with the asymptotic behavior of certain sequences of random variables; at the same time we will also be mindful of particular classes of random variables and we would wish that these classes would be nice, in some sense. There are many ways in which signed measures may converge, and the question quickly narrows to the problem of finding a suitable convergence type that induces a topology with favorable properties.

In order to find the appropriate sense of convergence we are led to consider the following well known result:

Theorem B.1 (The Riesz Representation Theorem). *Let Ω be an LCH space, and for $\nu \in M(\Omega)$ and $f \in C_0(\Omega)$ let $I_\nu(f) = \int f \mathrm{d}\nu$. Then the map $\nu \to I_\nu$ is an isometric isomorphism from $M(\Omega)$ to $C_0(\Omega)^*$.*

The weak* topology on $M(\Omega) = C_0(\Omega)$, in which $\nu_\alpha \to \nu$ if and only if $\int f \mathrm{d}\nu_\alpha \to \int f \mathrm{d}\nu$ for all $f \in C_0(\Omega)$, is of considerable importance in what follows; it is called the **vague topology**, and we denote it by $\nu_\alpha \xrightarrow{v} \nu$.

We continue by giving a useful criterion for vague convergence.

Proposition B.1 (Folland). *Suppose $\nu_1, \nu_2, \ldots, \nu \in M(\mathbb{R})$, and let $F_n(x) = \nu_n(-\infty, x]$ and $F(x) = \nu(-\infty, x]$.*

1. *If $\sup_n \|\nu_n\| < \infty$ and $F_n(x) \to F(x)$ for every x at which F is continuous, then $\nu_n \xrightarrow{v} \nu$.*

2. *If* $\nu_n \xrightarrow{v} \nu$, *then* $\sup_n \|\nu_n\| < \infty$. *If, in addition, the* ν_n*'s are positive, then* $F_n(x) \to F(x)$ *at every* x *at which* F *is continuous.*

Proof. (a) Since F is continuous except at countably many points, $F_n \to F$ a.e. with respect to Lebesgue measure. Also, $\|F_n\|_u \leq \|\nu_n\| \leq C$, so the F_n's are uniformly bounded. If f is continuously differentiable and has compact support then integration by parts and the Dominated Convergence Theorem yield

$$\int f \,\mathrm{d}\nu_n = \int f'(x)F_n(x)\,\mathrm{d}x \to \int f'(x)F(x)\,\mathrm{d}x = \int f\,\mathrm{d}\nu.$$

But the set of all such f's is dense in $C_0(\mathbb{R})$ (Proposition A.3.2), so $\int f\,\mathrm{d}\nu_n \to \int f\,\mathrm{d}\nu$ for all $f \in C_0(\mathbb{R})$ by Lemma B.1. That is, $\nu_n \xrightarrow{v} \nu$.

(b) If $\nu_n \xrightarrow{v} \nu$, then $\sup_n \|\nu_n\| < \infty$ by the Principle of Uniform Boundedness. Suppose that $\nu_n \geq 0$, and hence $\nu \geq 0$, and that F is continuous at $x = a$. If $f \in C_c(\mathbb{R})$ is the function that is 1 on $[-N, a]$, is 0 on $(-\infty, -N-\epsilon]$ and $[a+\epsilon, \infty)$, and linear in between, we have

$$F_n(a) - F_n(-N) = \nu_n(-N, a] \leq \int f\,\mathrm{d}\nu_n \to \int f\,\mathrm{d}\nu \leq F(a+\epsilon) - F(-N-\epsilon).$$

But as $N \to \infty$, $F_n(-N)$ and $F(-N-\epsilon)$ tend to zero, so

$$\limsup_{n\to\infty} F_n(a) \leq F(a+\epsilon).$$

Similarly, by considering the function that is 1 on $[-N+\epsilon, a-\epsilon]$, 0 on $(-\infty, N]$ and $[a, \infty)$, and linear in between, we see that

$$\liminf_{n\to\infty} F_n(a) \geq F(a-\epsilon).$$

Since ϵ is arbitrary and F is continuous at a, we have $F_n(a) \to F(a)$, as desired. $\square$

Proposition B.2. *If* $\nu_n \xrightarrow{v} \nu$ *in* $M(\Omega, \mathcal{B})$, *then* $|\nu|(O) \leq \liminf_n |\nu_n|(O)$ *for any open set* $O \in \mathcal{B}$.

Proof. Let O be an open set and $\epsilon > 0$. Then since for any $E \in \mathcal{B}$,

$$|\nu|(E) = \sup\{\left|\int_E f \, \mathrm{d}\nu\right| : |f| \leq 1\},$$

there exists f such that $\left|\int_E f \, \mathrm{d}\nu\right| \geq |\nu|(O) - \epsilon/2$, and Lusin's Theorem A.2.4 guarantees the existence of $\phi \in C_c$ with support contained in O and satisfying $|\phi| \leq |f| \leq 1$ such that $\phi = f$ except possibly on a set of variation measure $< \epsilon/2$. Then

$$\begin{aligned} \left|\int \phi \, \mathrm{d}\nu\right| &= \left|\int f \, \mathrm{d}\nu - \int (f - \phi) \, \mathrm{d}\nu\right| \\ &\geq \left|\int f \, \mathrm{d}\nu\right| - \left|\int (f - \phi) \, \mathrm{d}\nu\right| \\ &\geq |\nu|(O) - \epsilon/2 - \int |(f - \phi)| \, \mathrm{d}\nu \\ &> |\nu|(O) - \epsilon. \end{aligned}$$

And vague convergence allows us to write

$$\left|\int \phi \, \mathrm{d}\nu\right| = \lim_n \left|\int \phi \, \mathrm{d}\nu_n\right| \leq \liminf_n |\nu_n|(O).$$

Since ϵ can be made arbitrarily small we are finished. ☐

Proposition B.3. *If* $\{\nu_n\} \subset M(\Omega)$, $\nu_n \longrightarrow \nu$ *vaguely, and* $\|\nu_n\| \to \|\nu\|$, *then* $\int f \mathrm{d}\nu_n \to \int f \mathrm{d}\nu$ *for every* $f \in BC(\Omega)$. *Moreover, the hypothesis* $\|\nu_n\| \to \|\nu\|$ *cannot be omitted.*

Proposition B.4 (Folland). *Suppose that* $\nu, \nu_1, \nu_2, \ldots \in M(\mathbb{R})$. *If* $\|\nu_k\| \leq C < \infty$ *for all* k *and* $\widehat{\nu_n} \to \widehat{\nu}$ *pointwise, then* $\nu_n \xrightarrow{v} \nu$. *Conversely, if* $\nu_n \xrightarrow{v} \nu$ *and* $\|\nu_n\| \to \|\nu\|$, *then* $\widehat{\nu_n} \to \widehat{\nu}$ *pointwise.*

Proof. If $f \in \mathcal{S}$, then choose $g \in \mathcal{S}$ such that $\widehat{g} = f$ (which is possible by Proposition A.3.1), and then by the Fourier Inversion Theorem A.3.1,

$$\int f \, \mathrm{d}\nu_k = \int \int g(y) e^{-2\pi i y x} \, \mathrm{d}y \, \mathrm{d}\nu_k(x) = \int g(y) \widehat{\nu_k}(y) \, \mathrm{d}y.$$

Since $g \in L^1$ and $\|\widehat{\nu_k}\|_u \leq C$, the Dominated Convergence Theorem implies that $\int f \, \mathrm{d}\nu_k \to \int f \, \mathrm{d}\nu$. But $\mathcal{S}$ is dense in $C_0(\mathbb{R})$ (Proposition A.3.2), so Lemma B.1 shows that $\int f \, \mathrm{d}\nu_k \to \int f \, \mathrm{d}\nu$ for all $f \in C_0(\mathbb{R})$, that is, $\nu_n \xrightarrow{v} \nu$.

The converse follows immediately from Proposition B.3. □

It turns out that vague convergence is not strong enough for our purposes; however, the type of convergence discussed in Proposition B.3, in which $\nu_\alpha \to \nu$ if and only if $\int f \mathrm{d}\nu_\alpha \to \int f \mathrm{d}\nu$ for all $f \in BC(\Omega)$, is of primary importance in what follows and is termed **weak convergence**. We denote it by $\nu_\alpha \Rightarrow \nu$. The weak topology is stronger than the vague topology and turns out to have more useful properties that we are going to explore.

In order to discuss weak convergence fruitfully we need the following definition.

Definition B.1. *A family $\mathcal{A}$ of complex Borel measures on a topological space is said to be* ***tight*** *if for all $\epsilon > 0$ there exists a compact set K such that $|\nu|(K^c) < \epsilon$ for all $\nu \in \mathcal{A}$.*

Proposition B.5. *If there exists $\delta > 0$ such that $\sup_n |\mathbb{E}| \, |X_n|^\delta = C < \infty$, then $\{\nu_n\}$ is tight.*

Proof. This is an easy consequence of Chebychev's Inequality A.2.3: Let

$$p_{n,M} = |\mathbb{P}| \left(|X_n| > M \right).$$

Then

$$p_{n,M} \leq \frac{|\mathbb{E}||X_n|^\delta}{M^\delta} \leq \frac{C}{M^\delta} \longrightarrow 0, \text{ as } M \to \infty.$$

For $\epsilon > 0$, choose $M_\epsilon > (C/\epsilon)^{1/\delta}$. Then $\sup_n p_{n,M_\epsilon} < \epsilon$, and the compact set $[-M_\epsilon, M_\epsilon]$ suffices in the definition. □

Proposition B.6. *Let $\{X_n\}$ and $\{Y_n\}$ be independent. If $\{X_n\}$ and $\{Y_n\}$ are bounded and tight, then so is $\{X_n + Y_n\}$.*

Proof. Suppose $X_n \sim \mu_n$ and $Y_n \sim \nu_n$ with $\|\mu_n\|,\ \|\nu_n\| \le C < \infty$. Let $\epsilon > 0$. We notice that for any $M > 0$ the event

$$\{|X_n + Y_n| > M\} \quad \text{implies} \quad \{|X_n| > M/2 \text{ or } |Y_n| > M/2\}.$$

That is, by letting $f = 1\{|X_n + Y_n| > M\}$, $h_1 = 1\{|X_n| > M/2\}$, and $h_2 = 1\{|Y_n| > M/2\}$, we may say $f \le h_1 + h_2$. Now let M be so large that $|\mu_n|([-M/2, M/2]^c) < \epsilon/2C$ and $|\nu_n|([-M/2, M/2]^c) < \epsilon/2C$ for all n. Denote $g = \mathrm{d}|\mu_n * \nu_n|/\mathrm{d}(\mu_n * \nu_n)$, and remember that Proposition A.1.2 implies $|g| = 1$ a.e.. We are ready to compute

$$\begin{aligned}
|\mu_n * \nu_n|([-M, M]^c) &= |\mathbb{P}|(|X_n + Y_n| > M) \\
&= \int f \,\mathrm{d}|\mu_n * \nu_n| = \int f\, g\, \mathrm{d}(\mu_n * \nu_n) \\
&\le \int |f|\,|g|\,\mathrm{d}(|\mu_n| \times |\nu_n|) = \int f\,\mathrm{d}(|\mu_n| \times |\nu_n|) \\
&\le \int (h_1 + h_2)\,\mathrm{d}(|\mu_n| \times |\nu_n|) \\
&= |\mathbb{P}|(|X_n| > M/2) \cdot \|\nu_n\| + \|\mu_n\| \cdot |\mathbb{P}|(|Y_n| > M/2) \\
&< \epsilon/2 + \epsilon/2 = \epsilon.
\end{aligned}$$

Therefore, $\{X_n + Y_n\}$ is tight and boundedness follows directly from Proposition 1.2.1. □

The usefulness of the concept of tightness is well displayed by the following theorem, as generalized by Varadarajan [45].

Theorem B.2 (Prohorov's Theorem). *A family $\mathcal{A}$ of Borel probability measures on a separable complete metric space is relatively compact in the weak topology if and only if $\mathcal{A}$ is tight.*

When discussing weak convergence in classical Probability theory one quickly encounters Lévy's *Continuity Theorem* for characteristic functions which states that weak convergence of a sequence of probability distributions is equivalent to the pointwise convergence of the corresponding sequence of characteristic functions, provided the limit is continuous at zero.

In the complex case such a convenient statement is simply not true; the extra ingredients needed are supplied by Prohorov's Theorem and the result is the **Continuity Theorem for Complex Measures**, as developed by Báez-Duarte [5]:

Theorem B.3 (Báez-Duarte). *$\nu_n \Rightarrow \nu$ if and only if $\widehat{\nu_n} \to \widehat{\nu}$ pointwise and $\{\nu_n\}$ is bounded and tight.*

Theorem B.4 (Báez-Duarte). *If $\widehat{\nu_n} \to g$ pointwise and $\{\nu_n\}$ is bounded and tight, then there is a complex measure $\nu \in M$ such that $\widehat{\nu} = g$ and $\nu_n \Rightarrow \nu$.*

Proof of Theorem B.3. If $\nu_n \Rightarrow \nu$ then clearly $\widehat{\nu_n} \to \widehat{\nu}$. The Principle of Uniform Boundedness guarantees that $\{\nu_n\}$ is bounded, and tightness follows from Prohorov's Theorem. Thus the only if part is done, and the if part of the theorem is a particular case of Theorem B.4. $\square$

Proof of Theorem B.4. Suppose that $\widehat{\nu_n} \to g$ pointwise and $\{\nu_n\}$ is bounded and tight. For any subsequence $(n_{(1)})$, Prohorov's Theorem gives a sub-subsequence $(n_{(2)})$ such that $\nu_{n_{(2)}} \Rightarrow \nu$ for some $\nu \in M$. Then $\widehat{\nu_{n_{(2)}}} \to \widehat{\nu}$, so that $g = \widehat{\nu}$. The uniqueness of the Fourier-Stieltjes transform implies that the limiting measure ν obtained in the previous fashion does not depend on the subsequence $(n_{(1)})$. From this we conclude that $\nu_n \Rightarrow \nu$. $\square$

The two preceding theorems give criteria which are, in practice, somewhat inconvenient to check. Next we give a theorem which has stronger hypotheses and yet will prove to be more useful in the context of our later discussions.

Theorem B.5 (Báez-Duarte). *If* $\widehat{\nu_n} \to \widehat{\nu}$ *pointwise and* $\limsup \|\nu_n\| \leq \|\nu\|$, *then* $\nu_n \Rightarrow \nu$. *Furthermore, in this case also* $|\nu_n| \Rightarrow |\nu|$, *and, in particular,* $\|\nu_n\| \to \|\nu\|$.

Remark. $\nu_n^{\pm} \Rightarrow \nu^{\pm}$, *respectively, also follow.*

Proof. We begin by establishing that $\nu_n \xrightarrow{v} \nu$. To prove it for $\phi \in \mathcal{S}$ take $\psi \in \mathcal{S}$ such that $\widehat{\psi} = \phi$ (use Proposition A.3.1). Since $|\psi \widehat{\nu_n}| \leq |\psi| \sup_n \|\nu_n\|$, Parseval's Identity A.3.2 and Dominated Convergence imply

$$\int \phi \, \mathrm{d}\nu_n = \int \psi \widehat{\nu_n} \, \mathrm{d}x \to \int \psi \widehat{\nu} \, \mathrm{d}x = \int \phi \, \mathrm{d}\nu.$$

Since $\mathcal{S}$ is uniformly dense in C_c, Lemma B.1 gives the vague convergence.

The convergence of the total variations follows immediately by taking $O = \Omega$ in Proposition B.2. Now, for $\epsilon > 0$ we find an open set O with compact closure such that $|\nu|(O^c) < \epsilon$, which, on account of Proposition B.2, implies

$$\begin{aligned} \|\nu\| - \epsilon < |\nu|(O) \leq \liminf_n |\nu_n|(O) &= \liminf_n (\|\nu_n\| - |\nu_n|(O^c)) \\ &= \|\nu\| - \limsup_n |\nu_n|(O^c), \end{aligned}$$

that is, $\limsup_n |\nu_n|(O^c) < \epsilon$. Tightness of $\{\nu_n\}$ follows, and Theorem B.3 then yields $\nu_n \Rightarrow \nu$.

It remains to show that $|\nu_n| \Rightarrow |\nu|$. First discard the case $\nu = 0$, for then

$$\left| \int \phi \, \mathrm{d}|\nu_n| \right| \leq \|\phi\|_\infty \|\nu_n\| \to 0,$$

by convergence of the total variations.

Next we show that

$$|\nu_n|(O) \to |\nu|(O), \quad (O \text{ open}, \ |\nu|(\partial O) = 0). \tag{B.0.1}$$

To accomplish this write the disjoint union $\Omega = O \cup \partial O \cup U$. The assumption $|\nu|(\partial O) = 0$ and Proposition B.2 imply

$$\begin{aligned}
\|\nu\| = |\nu|(O) + |\nu|(U) &\le \liminf_n |\nu_n|(O) + \liminf_n |\nu_n|(U) \\
&\le \liminf_n |\nu_n|(O \cup U) \\
&\le \|\nu\|.
\end{aligned}$$

Therefore

$$\|\nu\| = \liminf_n |\nu_n|(O) + \liminf_n |\nu_n|(U) \tag{B.0.2}$$

so that

$$\begin{aligned}
\|\nu\| - \liminf_n |\nu_n|(O) &= \liminf_n |\nu_n|(U) \\
&= \liminf_n (\|\nu_n\| - |\nu_n|(O \cup \partial O)) \\
&\le \|\nu\| - \limsup_n |\nu_n|(O).
\end{aligned}$$

Thus $\lim |\nu_n|(O)$ exists for every open set with $|\nu|(\partial O) = 0$, and $\lim |\nu_n|(O) \ge |\nu|(O)$. As this applies to both O and U in equation (B.0.2), we see that it is impossible to have $\lim |\nu_n|(O) > |\nu|(O)$. Therefore equation (B.0.1) has been established.

To conclude weak convergence of the v-measures we employ a standard argument included here for the sake of completeness. Choose $f \in BC(\Omega)$ and $\epsilon > 0$. Let $C = \sup_n \|\nu_n\|$. Since the possible atoms of $|\nu| \circ f^{-1}$ form a countable set it is possible to subdivide the bounded range of f by means of points $a_1 < a_2 < \cdots < a_m$ such that $\max(a_{k+1} - a_k) < \epsilon/(2C)$ and the open sets $A_k = \{f < a_k\}$ have $|\nu|(\partial A_k) = 0$. Then for the simple function $s = \sum a_k 1_{A_k}$ one has $\|f - s\|_\infty < \epsilon/(2C)$. It is clear from equation (B.0.1) that

$$\omega_n = \left| \int s \,\mathrm{d}|\nu_n| - \int s \,\mathrm{d}|\nu| \right| \to 0,$$

so that

$$\begin{aligned} \left| \int f \,\mathrm{d}|\nu_n| - \int f \,\mathrm{d}|\nu| \right| &\leq \left| \int (f - s) \,\mathrm{d}|\nu_n| \right| + \omega_n + \left| \int (s - f) \,\mathrm{d}|\nu| \right| \\ &\leq 2\|f - s\|_\infty C + \omega_n \\ &\leq \epsilon + \omega_n. \end{aligned}$$

We have thus arrived at $|\nu_n| \Rightarrow |\nu|$. □

It is very useful to have weak convergence of the v-measures, as the following theorem shows.

Theorem B.6. *If $|\nu_n| \Rightarrow |\nu|$ and for some $\delta > 0$, $\sup_n |\mathbb{E}|\,|X_n|^\delta = M < \infty$, then*

$$|\mathbb{E}|\,|X|^\delta \leq M, \quad \textit{and for each } p < \delta, \quad \lim_{n\to\infty} |\mathbb{E}|\,|X_n|^p = |\mathbb{E}|\,|X|^p < \infty.$$

Remark. *If p above is a positive integer then we may replace $|X_n|^p$ and $|X|^p$ with X_n^p and X^p, respectively.*

Proof. We prove the remark since the assertion of the theorem is similar. For $A > 0$ define f_A on $\mathbb{R}$ as follows:

$$f_A(x) = \begin{cases} x^p, & \text{if } |x| \leq A, \\ A^p, & \text{if } x > A, \\ (-A)^p & \text{if } x < -A. \end{cases}$$

Then $f_A \in BC$ and by weak convergence

$$\int f_A \,\mathrm{d}|\nu_n| \to \int f_A \,\mathrm{d}|\nu|.$$

Next we have

$$\begin{aligned}
\int |f_A(x) - x^p| \, \mathrm{d}|\nu_n|(x) &\leq \int_{|x|>A} |x|^p \, \mathrm{d}|\nu_n|(x) \\
&\leq A^p \, |\mathrm{P}|(|X_n| > A) \\
&\leq A^p \left(\frac{|\mathrm{E}||X_n|^\delta}{A^\delta} \right) \\
&\leq \frac{M}{A^{\delta - p}}.
\end{aligned}$$

The last term does not depend on n, and converges to zero as $A \to \infty$. It follows that as $A \to \infty$, $\int f_A \, \mathrm{d}|\nu_n|$ converges uniformly in n to $\int x^p \, \mathrm{d}|\nu|$. Hence by a standard theorem on the inversion of repeated limits we have

$$\begin{aligned}
\int x^p \, \mathrm{d}|\nu| &= \lim_{A\to\infty} \int f_A \, \mathrm{d}|\nu| = \lim_{A\to\infty} \lim_{n\to\infty} \int f_A \, \mathrm{d}|\nu_n| \\
&= \lim_{n\to\infty} \lim_{A\to\infty} \int f_A \, \mathrm{d}|\nu_n| = \lim_{n\to\infty} \int x^p \, \mathrm{d}|\nu_n|.
\end{aligned}$$

It remains to show the first assertion of the theorem. Similar to the above for $A > 0$ we define g_A on R as follows:

$$g_A(x) = \begin{cases} |x|^\delta, & \text{if } |x| \leq A, \\ A^\delta, & \text{if } |x| > A. \end{cases}$$

Then $g_A \in BC$ and increases monotonically to $|x|^\delta$ as $A \to \infty$. But then

$$\begin{aligned}
\int |x|^\delta \, \mathrm{d}|\nu| &= \lim_{A\to\infty} \int g_A \, \mathrm{d}|\nu|, \quad \text{by the Monotone Convergence Theorem,} \\
&= \lim_{A\to\infty} \lim_{n\to\infty} \int g_A \, \mathrm{d}|\nu_n|, \quad \text{by weak convergence,} \\
&\leq \lim_{A\to\infty} M, \quad \text{since } \textstyle\int g_A \, \mathrm{d}|\nu_n| \leq M \text{ for each } n, \\
&= M.
\end{aligned}$$

□

Lemma B.1. *Let $\mathcal{F} \subset \mathcal{G}$ be normed spaces of measurable functions, and let the complex measures $\nu, \{\nu_n\} \subset M(\Omega)$ be bounded. Suppose $\int f \, \mathrm{d}\nu_n \to \int f \, \mathrm{d}\nu$ for all $f \in \mathcal{F}$ and further suppose that $\mathcal{F}$ is dense in $\mathcal{G}$. Then $\int g \, \mathrm{d}\nu_n \to \int g \, \mathrm{d}\nu$ for all $g \in \mathcal{G}$.*

Proof. Let $C = \sup_n \|\nu_n\|$. Given $g \in \mathcal{G}$ and $\epsilon > 0$, choose $f \in \mathcal{F}$ such that $\|f - g\| < \epsilon/(3C)$. If n is large enough so that $\left|\int f \, \mathrm{d}\nu_n - \int f \, \mathrm{d}\nu\right| < \epsilon/3$, we have

$$\begin{aligned}
\left|\int g \, \mathrm{d}\nu_n - \int g \, \mathrm{d}\nu\right| &\leq \left|\int g \, \mathrm{d}\nu_n - \int f \, \mathrm{d}\nu_n\right| + \left|\int f \, \mathrm{d}\nu_n - \int f \, \mathrm{d}\nu\right| \\
&+ \left|\int f \, \mathrm{d}\nu - \int g \, \mathrm{d}\nu\right| \\
&\leq 2C\|f - g\| + \epsilon/3 < \epsilon,
\end{aligned}$$

so that $\int g \, \mathrm{d}\nu_n \to \int g \, \mathrm{d}\nu$. $\square$