Probabilistic Aspects of Moment Sequences: The Method of Moments

Bachelorprojekt i matematik
Institut for Matematiske Fag
Københavns Universitet

Bachelor Thesis in Mathematics
Department of Mathematical Sciences
University of Copenhagen

Frederik Riis Mikkelsen
Supervisor: Magdalena Musat

8th of June 2012

Resumé

I dette bachelorprojekt behandles det klassiske momentproblem og dets forbindelse til momentmetoden i sandsynlighedsregning. Forskellige resultater bevises undervejs med henblik på at løse det klassiske momentproblem og at etablere resultater indenfor momentmetoden. Nogle af disse resultater (herunder Kontinuitets sætningen) har yderligere anvendelser, som også vil blive behandlet i dette projekt. Disse anvendelser inkluderer blandt andet den centrale grænseværdis sætning i sandsynlighedsregning og Bochners sætning i analyse.

Abstract

This bachelor thesis in mathematics deals with the classical moment problem and its connection to the Method of moments in probability theory. Several important results are proven along the way, some of which are needed to establish the concept of the Method of moments, or to solve the classical moment problem. Furthermore, a number of applications of the results established along the way (of which, most notably, the Continuity Theorem) are given, in order to illustrate how fruitfull the subject is. Such applications include the Central Limit Theorem in probability and Bochner's theorem in analysis.

Introduction

This thesis deals with the classical moment problem and its relevance and applications in probability theory. The classical moment problem is as follows: given a sequence of real numbers, does there exist a measure μ on $\mathbb{R}$ having these numbers as its moments? Furthermore, if this is the case, is μ the only measure having this property?

The thesis is organized as follows: In the first chapter we introduce the basic concepts regarding moments and characteristic functions, along with relevant and basic results concerning them. The chapter concludes with an important connection between moments of a probability measure and properties of its characteristic function. The basic inequalities involving moments follow the approach from [3] and the proofs of the results regarding characteristic functions mainly follow [6]. Also a great thanks to Ernst Hansen for providing us with some very useful notes on the subject of characteristic functions.

In Chapter 2, we first consider the uniqueness part of the classical moment problem, i.e., we give sufficient conditions for a probability measure that has all moments, to be the only probability measure with these moments. We apply these conditions to adress the question whether a number of concrete probability distributions (which appear frequently in applications, cf. Table 1.1) are completely determined by their moments, or not. While the answer is yes in most cases considered, the Weibull distribution turned out to be more difficult to handle. We would like to thank Thomas Mikosch for pointing out the reference [8] to us, which proved instrumental in establishing the behaviour of this distribution. Furthermore, we also give examples of families of probability measures, that share the same moments. In the end of the second chapter we consider the existence part of the classical moment problem. To adress this, we need to employ important tools from functional analysis (such as the Hahn-Banach extension theorem and the Riesz representation theorem, discussed therein).

The way the classical moment problem is closely linked to probability theory is discussed in Chapter 3. We consider several different types of convergence of real-valued random variables, including distributional convergence (or weak convergence in measure theory). As it turns out, one can determine distributional convergence from convergence of the corresponding moments. This, however, requires the limit random variable to be uniquely determined by its moments. This is how the previous chapter is linked to this chapter. This approach to determine distributional convergence is called the Method of moments. In the end of the chapter, the applicability of the Method of moments is illustrated with a proof of the Central Limit theorem. The main tool used to prove the Method of moments is the Continuity Theorem, which basically links weak convergence of probability measures and pointwise convergence of the corresponding characteristic functions. The Continuity theorem has a broad range of applications. For instance, we use it to prove Stirling's formula and another version of the Central Limit theorem. And we also prove a little bonus result, giving necessary and sufficient conditions for a function to be the characteristic function of some probability measure on $\mathbb{R}$. (This is Bochner's theorem.)

The thesis ends with two appendices. The first contains several approximation results (of continuous functions by real polynomials, trigonometric polynomials and simple functions). These are needed in order to prove some of the results regarding characteristic functions and they are also used to show that probability measures concentrated on bounded intervals are uniquely determined by their moments. The second appendix contains the Carathéodory theorem, whose proof mainly follows [5]. This is used to prove that every right-continuous and non-decreasing function F on $\mathbb{R}$ corresponds to a measure μ on $(\mathbb{R}, \mathbb{B})$, related by $F(x) = \mu((-\infty, x])$, $x \in \mathbb{R}$. With this latter result available, we can prove some useful convergence results in Chapter 3.

Contents

Chapter 1

Moments and Characteristic Functions

1.1 Moments and Basic Inequalities

Let μ be a probability measure on the measure space $(\mathbb{R}, \mathbb{B})$, where $\mathbb{B}$ is the Borel σ-algebra on $\mathbb{R}$. The definition of the moments of μ naturally occurs when considering integrals of polynomials with respect to μ. More precisely,

Definition 1.1. *Let μ be a probability measure on $(\mathbb{R}, \mathbb{B})$. For $k \in \mathbb{N}$, the measure μ has k^{th} moment if the map $x \mapsto x^k$ is μ-integrable. In this case, the k^{th} moment of μ is defined as*

$$\int x^k \, d\mu(x),$$

while the k^{th} absolute moment of μ is

$$\int |x|^k \, d\mu(x).$$

A sequence of real numbers $(\alpha_n)_{n\in\mathbb{N}}$ is called a ***moment sequence****, if there exists a probability measure μ on $(\mathbb{R}, \mathbb{B})$, such that for each $n \in \mathbb{N}$, α_n is the n^{th} moment of μ.*

Let X be a real-valued random variable defined on a probability space $(\Omega, \mathbb{F}, P)$. Then the distribution of X, denoted by $X(P)$ is a probability measure on $(\mathbb{R}, \mathbb{B})$, so if $X(P)$ has k^{th} moment, we say that X has k^{th} moment. In this case we define the k^{th} moment of X as the k^{th} moment of $X(P)$. This is denoted by $E(X^k)$, while the absolute moment is denoted by $E\,|X|^k$. By the Change of variable theorem (cf. Theorem 16.13 in [2]),

$$\int x^k dX(P) = \int X^k dP,$$

which justifies the above notation. In Table 1.1, we list the moments of a random variable X which follows different distributions. All these moments can be derived from the definition.

Distribution	Support	Density function	Moments
Normal distribution $(0, \sigma^2)$, $\sigma > 0$	$\mathbb{R}$	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}}$	$E(X^k) = \begin{cases} \sigma^2 \frac{2^{k/2}}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right) & \text{if } k \text{ even} \\ 0 & \text{if } k \text{ odd} \end{cases}, \quad k \in \mathbb{N}$
Gamma distribution (λ, β), $\lambda, \beta > 0$	$(0, \infty)$	$f(x) = \frac{1}{\beta^\lambda \Gamma(\lambda)} x^{\lambda-1} e^{\frac{-x}{\beta}}$	$E(X^k) = \beta^k \frac{\Gamma(\lambda+k)}{\Gamma(\lambda)}, \quad k \in \mathbb{N}$
B-distribution (λ_1, λ_2), $\lambda_1, \lambda_2 > 0$	$(0, 1)$	$f(x) = \frac{x^{\lambda_1-1}(1-x)^{\lambda_2-1}}{B(\lambda_1,\lambda_2)}$	$E(X^k) = \frac{B(\lambda_1+k,\lambda_2)}{B(\lambda_1,\lambda_2)}, \quad k \in \mathbb{N}$
F-distribution (λ_1, λ_2), $\lambda_1, \lambda_2 > 0$	$(0, \infty)$	$f(x) = \frac{\lambda_1^{\lambda_1} \lambda_2^{\lambda_2}}{B(\lambda_1,\lambda_2)} \frac{x^{\lambda_1-1}}{(\lambda_1 x+\lambda_2)^{\lambda_1+\lambda_2}}$	$E(X^k) = \left(\frac{\lambda_2}{\lambda_1}\right)^k \frac{B(\lambda_1+k,\lambda_2-k)}{B(\lambda_1,\lambda_2)}, \quad k < \lambda_2$
t-distribution (λ), $\lambda > 0$	$\mathbb{R}$	$f(x) = \frac{\left(1+\frac{x^2}{\lambda}\right)^{-\frac{\lambda+1}{2}}}{\sqrt{\lambda} B(\lambda/2,1/2)}$	$E(X^k) = \begin{cases} \frac{\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{\lambda-k}{2}\right)\lambda^{k/2}}{\sqrt{\pi}\Gamma\left(\frac{\lambda}{2}\right)} & \text{if } k \text{ even} \\ 0 & \text{if } k \text{ odd} \end{cases}, \quad k < \lambda$
Log-normal distribution $(0, \sigma^2)$, $\sigma > 0$	$(0, \infty)$	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} e^{\frac{-\log(x)^2}{2\sigma^2}}$	$E(X^k) = e^{\sigma^2 \frac{k^2}{2}}, \quad k \in \mathbb{N}$
Uniform distribution (a, b), $a < b$	(a, b)	$f(x) = \frac{1}{b-a}$	$E(X^k) = \frac{1}{k+1}(b^{k+1} - a^{k+1}), \quad k \in \mathbb{N}$
Weibull distribution (c, β), $c, \beta > 0$	$(0, \infty)$	$f(x) = \frac{c}{\beta^c} x^{c-1} e^{\frac{-x^c}{\beta^c}}$	$E(X^k) = \beta^k \Gamma\left(1 + \frac{k}{c}\right), \quad k \in \mathbb{N}$
Fréchet distribution (c, β), $c, \beta > 0$	$(0, \infty)$	$f(x) = c \frac{\beta^c}{x^{c+1}} e^{\frac{-\beta^c}{x^c}}$	$E(X^k) = \beta^k \Gamma\left(1 - \frac{k}{c}\right), \quad k < c$
Laplace distribution $(0, \beta)$, $\beta > 0$	$\mathbb{R}$	$f(x) = \frac{1}{2\beta} e^{\frac{-\lvert x\rvert}{\beta}}$	$E(X^k) = \begin{cases} \beta^k k! & \text{if } k \text{ even} \\ 0 & \text{if } k \text{ odd} \end{cases}, \quad k \in \mathbb{N}$
Pareto distribution (α, β), $\alpha, \beta > 0$	(β, ∞)	$f(x) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}}$	$E(X^k) = \frac{\alpha\beta^k}{\alpha-k}, \quad k < \alpha$

Table 1.1: Examples of measures and their moments. Here $\Gamma(\cdot)$ and $B(\cdot)$ denote the gamma-function and beta-function, respectively.

The next lemma is an immediate result

Lemma 1.2. *Let μ be a probability measure on $(\mathbb{R}, \mathbb{B})$ and $m \in \mathbb{N}$. If μ has k^{th} moment and $m \leq k$, then μ has m^{th} moment, as well.*

Proof. Indeed, for $x \in \mathbb{R}$

$$|x|^m \leq \begin{cases} 1 & \text{if } |x| < 1 \\ |x|^k & \text{else.} \end{cases}$$

Hence $|x|^m \leq 1 + |x|^k$. Integration reveals

$$\int |x|^m \ d\mu(x) \leq \int 1 \ d\mu(x) + \int |x|^k \ d\mu(x) < \infty,$$

which shows that the m^{th} moment exists, as claimed. □

The following version of an important inequality will come in handy, later on.

Lemma 1.3 (Chebyshev's inequality). *Let X be a real-valued random variable defined on $(\Omega, \mathbb{F}, P)$. Let ϕ be an even function that is strictly positive and increasing on $(0, \infty)$. Assume that $\phi(X)$ has first moment. Then, for every $r > 0$,*

$$P(|X| \geq r) \leq \frac{E\left(\phi(X)\right)}{\phi(r)}.$$

Proof. The function ϕ is non-negative and has minimum $\phi(r)$ on the set $\{|X| \geq r\}$, so

$$E(\phi(X)) = \int \phi(X) \ dP \geq \int_{\{|X| \geq r\}} \phi(X) \ dP \geq \phi(r) P(|X| \geq r).$$

Rearranging this gives us the result. □

One of the many applications of Chebyshev's inequality is in the proof of the Weierstrass approximation theorem with $\phi(t) = t^2$ (see Appendix A). It will also be used in the case $\phi(t) = |t|^r$ for a positive integer r in the proof of a later theorem (cf. Theorem 3.35).

Lemma 1.4 (Lyapounov's inequality). *Let X be a real-valued random variable defined on $(\Omega, \mathbb{F}, P)$. If X has k^{th} moment, for some $k \geq 1$, then for every $m = 1, 2, ..., k-1$, we have that*

$$(E(|X|^m))^{1/m} \leq (E(|X|^k))^{1/k}.$$

Proof. Fix $m = 1, 2, ..., k-1$. By Lemma 1.2 we know that X has m^{th} moment. We observe that the map $x \mapsto x^{k/m}$ is convex on $[0, \infty)$ (since $k/m > 1$). Jensen's inequality implies that

$$(E(|X|^m))^{k/m} \leq E((|X|^m)^{k/m}) = (E(|X|^k)).$$

Rearranging this gives us the result. □

Remark 1.5. With this lemma, we now see that a **necessary** condition for a sequence of real numbers $(\alpha_n)_{n\in\mathbb{N}}$ to be a moment sequence, is that $((\alpha_n)^{1/n})_{n\in\mathbb{N}}$ is non-decreasing. However, it is **not** a **sufficient** condition, as we will see next. Consider the following sequence

$$a_1 = 1, \quad a_2 = 1, \quad a_3 = 2, \quad a_4 = 3, \quad \text{and} \quad a_m = m^m, \quad \text{for } m \geq 5.$$

Then $((\alpha_n)^{1/n})_{n\in\mathbb{N}}$ is non-decreasing. Now assume that it is a moment sequence, i.e., that there exists a probability measure μ, such that $\int x^n \, d\mu(x) = \alpha_n$ for all $n \in \mathbb{N}$. Then the non-negative function $x \mapsto (3x - 2x^2)^2$ is integrable and must have non-negative integral. But we see that

$$0 \leq \int (3x - 2x^2)^2 \, d\mu(x) = \int 9x^2 + 4x^4 - 12x^3 \, d\mu(x) = 9\alpha_2 + 4\alpha_4 - 12\alpha_3 = -3,$$

which is a contradiction. So we conclude that no such measure μ exists, hence the condition cannot be sufficient. For a necessary and sufficient condition, see chapter 2.

Let $C_b(\mathbb{R})$ be the class of all real-valued, continuous and bounded functions on $\mathbb{R}$. The next result will provide a criteria for showing that two probability measures are equal.

Theorem 1.6. *Let μ and ν be probability measures on $(\mathbb{R}, \mathbb{B})$. If*

$$\int f \, d\mu = \int f \, d\nu, \qquad \text{for all } f \in C_b(\mathbb{R}),$$

then $\mu = \nu$.

Proof. Let (a, b) denote a bounded interval in $\mathbb{R}$. There exists $N \in \mathbb{N}$ such that $n > 2/(b - a)$ for all $n > N$. For each such positive integer n, define the following function

$$f_n(x) = \begin{cases} 1 & \text{if } x \in \left[a + \frac{1}{n}, b - \frac{1}{n}\right] \\ n(x - a) & \text{if } x \in \left(a, a + \frac{1}{n}\right) \\ n(b - x) & \text{if } x \in \left(b - \frac{1}{n}, b\right) \\ 0 & \text{else.} \end{cases}$$

Then $f_n \to 1_{(a,b)}$, as $n \to \infty$, pointwise and dominated by the constant function 1. By the Lebesgue dominated convergence theorem,

$$\int f_n \, d\mu \to \int 1_{(a,b)} \, d\mu = \mu(a, b), \qquad \text{as } n \to \infty.$$

Similarly for ν, and since $f_n \in C_b(\mathbb{R})$, μ and ν agree (a, b). Hence they also agree on all open bounded intervals. These subsets form an intersection-stable generator for the Borel σ-algebra $\mathbb{B}$. So by the uniqueness theorem for probability measures (cf. Theorem 3.3 in [2]), $\mu = \nu$. □

1.2 Characteristic Functions

We will need another concept in order to obtain a new approach to moments of probability measures.

Definition 1.7. *Let μ be a probability measure on $(\mathbb{R}, \mathbb{B})$. The characteristic function of μ is the function $\phi : \mathbb{R} \to \mathbb{C}$ defined by*

$$\phi(\theta) = \int e^{i\theta x} \, d\mu(x), \quad \theta \in \mathbb{R}.$$

Note that ϕ is well-defined, since the map $x \mapsto e^{i\theta x}$ is measurable for all $\theta \in \mathbb{R}$ and $\left|e^{i\theta x}\right| = 1$, so this map is also μ-integrable. The above formula involves a complex-valued integral, so we rewrite it as

$$\phi(\theta) = \int \cos(\theta x) \, d\mu(x) + i \int \sin(\theta x) \, d\mu(x), \quad \theta \in \mathbb{R}$$

From this we see that ϕ is conjugate symmetric (i.e., $\phi(-\theta) = \overline{\phi(\theta)}$ for all $\theta \in \mathbb{R}$). Since sin is an odd function and cos is even, we get for all $\theta \in \mathbb{R}$,

$$\begin{aligned} \phi(-\theta) &= \int \cos(-\theta x) \, d\mu(x) + i \int \sin(-\theta x) \, d\mu(x) \\ &= \int \cos(\theta x) \, d\mu(x) - i \int \sin(\theta x) \, d\mu(x) = \overline{\phi(\theta)}. \end{aligned}$$

If X is a real-valued random variable, then we can define the characteristic function of X. We simply define it as the characteristic function of the distribution of X. This approach is justified by the Change of variable theorem:

$$\int e^{i\theta x} \, dX(P) = \int e^{i\theta X} \, dP = E(e^{i\theta X}).$$

Lemma 1.8. *Let X be a real-valued random variable with characteristic function ϕ. Then $aX + b$ has characteristic function ψ, given by*

$$\psi(\theta) = e^{i\theta b} \phi(a\theta), \quad \theta \in \mathbb{R}.$$

Proof. The proof is a simple calculation:

$$\psi(\theta) = \int e^{i\theta(aX+b)} \, dP = e^{i\theta b} \int e^{ia\theta X} \, dP = e^{i\theta b} \phi(a\theta),$$

as wanted. □

Lemma 1.9. *Let $X_1, X_2, ..., X_n$ be real-valued independent random variables with characteristic function ϕ_i, respectively. Then $S_n = \sum_{i=1}^{n} X_i$ has characteristic function ψ, given by*

$$\psi(\theta) = \prod_{i=1}^{n} \phi_i(\theta), \quad \theta \in \mathbb{R}.$$

Proof. Since $X_1, X_2, ..., X_n$ are independent, then so are $e^{i\theta X_1}, e^{i\theta X_2}, ..., e^{i\theta X_n}$, for each $\theta \in \mathbb{R}$. So

$$\psi(\theta) = E(e^{i\theta S_n}) = E\left(\prod_{i=1}^{n} e^{i\theta X_i}\right) = \prod_{i=1}^{n} E(e^{i\theta X}) = \prod_{i=1}^{n} \phi_i(\theta),$$

as claimed. □

By using some of the approximation results from Appendix A, we shall see that probability measures are uniquely determined by their characteristic functions.

Theorem 1.10. *Let μ and ν be two probability measures on $(\mathbb{R}, \mathbb{B})$, and let ϕ_μ and ϕ_ν be their respective characteristic functions. If $\phi_\mu = \phi_\nu$, then $\mu = \nu$.*

Proof. Let $q : \mathbb{R} \to \mathbb{C}$ be a generalized trigonometric polynomial, i.e., of the form (see Appendix A)

$$q(\theta) = \sum_{j=1}^{N} c_j e^{i\alpha_j \theta}, \quad \theta \in \mathbb{R}.$$

Note that the integral of q with respect to μ is uniquely determined by the characteristic function ϕ_μ. More precisely,

$$\int q(\theta)\ d\mu(\theta) = \sum_{j=1}^{N} c_j \int e^{i\alpha_j\theta}\ d\mu(\theta) = \sum_{j=1}^{N} c_j \phi_\mu(\alpha_j).$$

A similar result holds, of course, for ν. So μ and ν agree on the generalized trigonometric polynomials, when it comes to integration. Given $f \in C_b(\mathbb{R})$, by Corollary A.6 there exists a sequence of real generalized trigonometric polynomials $(q_n)_{n\in\mathbb{N}}$ such that for every $n \in \mathbb{N}$,

$$\sup_{\theta\in[-n,n]} |f(\theta) - q_n(\theta)| < \frac{1}{n}.$$

Furthermore, the fact that $\|q_n\| \leq \|f\| + 1$, for all n, ensures that the Lebesgue dominated convergence theorem applies to the sequence $(q_n)_{n\in\mathbb{N}}$ converging pointwise to f, as $n \to \infty$. Hence

$$\int q_n\ d\mu \to \int f\ d\mu, \quad \text{as } n \to \infty.$$

A similar argument applies to ν. So the integrals of any $f \in C_b(\mathbb{R})$ are the same for both measures. By Theorem 1.6, we conclude that $\mu = \nu$. □

There are further important and useful results regarding characteristic functions. First of all, we will establish a connection between existence of moments of a measure and smoothness properties of its characteristic function. For this we need the following lemma:

Lemma 1.11. *The following holds*

$$\int_0^y \frac{\sin(t)}{t}\ dt \to \frac{\pi}{2}, \qquad \textit{as}\ \ y \to \infty.$$

Proof. Note that the map $t \mapsto \sin(t)/t$ is not integrable on $(0,\infty)$. For instance, observe that

$$\sum_{j=1}^{\infty} \frac{1}{2t} 1_{(\frac{\pi}{3}+\pi j, \frac{2\pi}{3}+\pi j)} \leq \left| \frac{\sin(t)}{t} \right|,$$

where it is easily seen that the left hand side has infinite integral, so we have to be careful.

We consider the function $f : (0,\infty)^2 \to \mathbb{R}$ given by $f(t,x) = \sin(t)e^{-tx}$. For all $y > 0$ we see that f is m_2-integrable over $(0,y) \times (0,\infty)$ by use of Tonelli's theorem,

$$\int_{(0,y)\times(0,\infty)} |f| \; dm_2 = \int_0^y \int_0^\infty |\sin(t)| \, e^{-tx} \; dxdt \leq \int_0^y \int_0^\infty te^{-tx} \; dxdt = \int_0^y 1 \; dt = y < \infty.$$

Fubini's theorem then applies and both orders of integration represent the integral of f on $(0,y) \times (0,\infty)$ with respect to m_2. The first order of integration reveals

$$\int_{(0,y)\times(0,\infty)} f \; dm_2 = \int_0^y \int_0^\infty \sin(t)e^{-tx} \; dxdt = \int_0^y \frac{\sin(t)}{t} dt.$$

In the other order we first consider the inner integral. Since the integration domain is a bounded interval and the integrand is continuous on the whole real line, the integral exists both as a Riemann integral and a Lebesgue integral. Therefore the integral can be calculated as a Riemann integral (see section 17 in [2]). Using the substitution $z = tx$ on the inner integral gives us

$$\begin{aligned}
\int_0^y \sin(t)e^{-tx} \; dt &= \int_0^{yx} \sin\left(\frac{z}{x}\right) e^{-z} \frac{1}{x} \; dz = \frac{1}{x} \left[\frac{-x}{1+x^2} e^{-z} \left(\cos\left(\frac{z}{x}\right) + x \sin\left(\frac{z}{x}\right) \right) \right]_{z=0}^{z=yx} \\
&= \frac{1}{x} \left(\frac{-x}{1+x^2} e^{-yx} \left(\cos(y) + x\sin(y)\right) - \frac{-x}{1+x^2} \right) \\
&= \left(\frac{1}{1+x^2} - \frac{1}{1+x^2} e^{-yx} \cos(y) - \frac{x}{1+x^2} e^{-yx} \sin(y) \right),
\end{aligned}$$

where we used the antiderivative $F(z) = \frac{-x}{1+x^2} e^{-z} \left(\cos\left(\frac{z}{x}\right) + x\sin\left(\frac{z}{x}\right)\right)$. Now we get that

$$\begin{aligned}
\int_{(0,y)\times(0,\infty)} f \; dm_2 &= \int_0^\infty \int_0^y \sin(t)e^{-tx} \; dtdx \\
&= \int_0^\infty \left(\frac{1}{1+x^2} - \frac{1}{1+x^2} e^{-yx} \cos(y) - \frac{x}{1+x^2} e^{-yx} \sin(y) \right) \; dx \\
&= \frac{\pi}{2} - \int_0^\infty \left(\frac{1}{1+x^2} e^{-yx} \cos(y) + \frac{x}{1+x^2} e^{-yx} \sin(y) \right) \; dx.
\end{aligned}$$

The last equality follows since the integrand is integrable and $\int_0^\infty \frac{1}{1+x^2} \; dx = \pi/2$. Combining the two representations of the integral reveals

$$\left| \int_0^y \frac{\sin(t)}{t} \; dt - \frac{\pi}{2} \right| = \left| \int_0^\infty \frac{1}{1+x^2} e^{-yx} \cos(y) + \frac{x}{1+x^2} e^{-yx} \sin(y) dx \right| \leq \int_0^\infty 2e^{-yx} \; dx \leq \frac{2}{y}.$$

Letting $y \to \infty$ yields the result. $\square$

As a consequence of the above theorem, we see that the map $y \mapsto \int_0^y \frac{\sin(t)}{t}\, dt$ is bounded on $(0, \infty)$. It will also be useful to observe that

$$\int_{-y}^{y} \frac{\sin(\alpha t)}{t}\, dt = 2\int_0^y \frac{\sin(\alpha t)}{t}\, dt = 2\int_0^{\alpha y} \frac{\sin(z)}{z}\, dz \to \begin{cases} \pi & \text{if } \alpha > 0 \\ -\pi & \text{if } \alpha < 0 \\ 0 & \text{if } \alpha = 0 \end{cases}, \quad \text{as } y \to \infty \tag{1.1}$$

by using that $t \mapsto \sin(\alpha t)/t$ is an even function and by the substitution $z = \alpha t$.

As we will see next, integrating the characteristic functions reveals some very interesting results. First we need to make sure that the characteristic function of a measure can be integrated, in fact, as it turns out, it is uniformly continuous:

Lemma 1.12. *Let μ be a probability measure on $(\mathbb{R}, \mathbb{B})$. Let $g : \mathbb{R} \to \mathbb{C}$ be a μ-integrable function and let $\lambda \in \mathbb{R}$. The complex-valued function*

$$f(x) = \int e^{i\lambda x t} g(t)\, d\mu(t), \qquad x \in \mathbb{R},$$

is uniformly continuous.

Proof. Integrability of g ensures that f is well-defined. Let $x \in \mathbb{R}$ be arbitrary. We want to prove that $|f(x + \Delta x) - f(x)| \to 0$, as $\Delta x \to 0$. We see that

$$\begin{aligned} |f(x+\Delta x) - f(x)| &= \left| \int e^{i\lambda(x+\Delta x)t} g(t) - e^{i\lambda x t} g(t)\, d\mu(t) \right| \\ &\le \int |g(t)| \left| \left(e^{i\lambda \Delta x t} - 1\right) e^{i\lambda x t} \right| d\mu(t) = \int |g(t)| \left| e^{i\lambda \Delta x t} - 1 \right| d\mu(t). \end{aligned}$$

Note that the integrand converges pointwise to 0, as $\Delta x \to 0$, dominated by the μ-integrable function $t \mapsto 2\,|g(t)|$. Therefore, an application of the Lebesgue dominated convergence theorem shows that $|f(x + \Delta x) - f(x)| \to 0$, as $\Delta x \to 0$. ☐

Setting $\lambda = 1$ and $g(t) \equiv 1$ we see that the characteristic function of μ is uniformly continuous. It is now possible to integrate the characteristic function with respect to the Lebesgue measure over a bounded interval.

Theorem 1.13 (The inversion formula). *Let ϕ be the characteristic function of a probability measure μ on $(\mathbb{R}, \mathbb{B})$. For $a < b$,*

$$\frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-iat} - e^{-ibt}}{it} \phi(t)\, dt \to \mu((a, b)) + \frac{\mu(\{a\}) + \mu(\{b\})}{2}, \qquad \text{as } T \to \infty. \tag{1.2}$$

Proof. For fixed a, b we introduce the function $f(t) = (e^{-iat} - e^{-ibt})/(it)$, for $t \in \mathbb{R}$. Observe that f is bounded, since for $t \in \mathbb{R}$

$$|f(t)| = \left| \int_a^b e^{-itz}\, dz \right| \le \int_a^b \left| e^{-itz} \right| dz = b - a,$$

so the integral in (1.2) is well-defined. Also f is conjugate symmetric, hence $f\phi$ is also conjugate symmetric. This implies that the imaginary part of the function is odd, so $\int_{-T}^{T} f(t)\phi(t)dt$ is real. We write out the characteristic function in (1.2),

$$\int_{-T}^{T} f(t)\phi(t)\ dt = \int_{-T}^{T}\int f(t)e^{itx}\ d\mu(x)dt.$$

We want to interchange the order of integration, by use of Fubini's theorem. It applies since we have the sufficient integrability conditions

$$\int_{-T}^{T}\int \left|f(t)e^{itx}\right|\ d\mu(x)dt \leq \int_{-T}^{T}\int b-a\ d\mu(x)dt = 2T(b-a) < \infty.$$

We write out the integrand

$$\begin{aligned} f(t)e^{itx} &= \frac{e^{-i(a-x)t} - e^{-i(b-x)t}}{it} \\ &= \frac{\cos((a-x)t) - i\sin((a-x)t) - \cos((b-x)t) + i\sin((b-x)t)}{it} \\ &= \frac{\sin((b-x)t) - \sin((a-x)t)}{t} - i\frac{\cos((a-x)t) - \cos((b-x)t)}{t}. \end{aligned}$$

The new inner integral is therefore (only the real part is relevant because of conjugate symmetry)

$$\int_{-T}^{T}\frac{\sin((b-x)t) - \sin((a-x)t)}{t}\ dt = \int_{-T}^{T}\frac{\sin((b-x)t)}{t}\ dt - \int_{-T}^{T}\frac{\sin((a-x)t)}{t}\ dt.$$

We consider the limit $T \to \infty$. Depending on the sign of $a-x$ and $b-x$ the limits will be either π, 0 or $-\pi$, as we saw in (1.1). This then gives us

$$\int_{-T}^{T} f(t)e^{itx}\ dt \to \begin{cases} 2\pi & x \in (a,b) \\ \pi & x \in \{a,b\} \\ 0 & \text{else} \end{cases}, \qquad \text{as } T \to \infty.$$

This convergence is dominated (since the function $y \mapsto \int_0^y \frac{\sin(t)}{t}\ dt$ is bounded). Applying the Lebesgue dominated convergence theorem we obtain

$$\int\int_{-T}^{T} f(t)e^{itx}\ dtd\mu(x) \to \int 2\pi 1_{(a,b)} + \pi 1_{\{a,b\}} d\mu, \qquad \text{as } T \to \infty.$$

Thus proving the claim. □

As an interesting application, we obtain the following:

Corollary 1.14. *Let ϕ be the characteristic function of a probability measure μ on $(\mathbb{R}, \mathbb{B})$. If ϕ is integrable with respect to the Lebesgue measure m, then μ has density with respect to m. The density is*

$$g(x) = \frac{1}{2\pi}\int e^{-ixt}\phi(t)\ dm(t), \qquad x \in \mathbb{R}.$$

Proof. One might think that the function g from above is complex-valued. However we quickly realize that the integrand is conjugate symmetric, hence g is a real-valued function. There can be at most countably many points of discontinuity for μ. So let $a < b$ be two points of continuity. Let f denote the function from the previous proof. For every $T > 0$ the map $t \mapsto 1_{(-T,T)} f\phi$ is bounded by an m-integrable function

$$\left|1_{(-T,T)}(t)f(t)\phi(t)\right| \leq \left|\frac{e^{-iat} - e^{-ibt}}{it}\phi(t)\right| \leq (b-a)\left|\phi(t)\right|,$$

since we assume that ϕ is integrable. Now letting $T \to \infty$ we get that $1_{(-T,T)} f\phi \to f\phi$. So by the Lebesgue dominated convergence theorem combined with the inversion formula we get (using that a, b are points of continuity, so that the point masses are zero)

$$\mu((a,b)) = \frac{1}{2\pi}\int f\phi \; dm = \int \int_a^b \frac{1}{2\pi} e^{-itx}\phi(t) \; dxdm(t),$$

Here we have used the integral representation of f that we saw in the previous proof. We want to apply Fubini's theorem, so we check for integrability

$$\int \int_a^b \left|\frac{1}{2\pi} e^{-itx}\phi(t)\right| \; dxdm(t) = \int \int_a^b \frac{1}{2\pi}\left|\phi(t)\right| \; dxdm(t) = \frac{b-a}{2\pi}\int \left|\phi(t)\right| \; dm(t) < \infty,$$

since ϕ is integrable by assumption. Applying Fubini's theorem reveals

$$\mu((a,b)) = \int_a^b \int \frac{1}{2\pi} e^{-itx}\phi(t) \; dm(t)dx = \int_a^b g(x)dx. \tag{1.3}$$

So $\mu = g \cdot m$ on all Borel sets of the form (a, b), where a, b belongs to a dense subset in $\mathbb{R}$. These sets form an intersection-stable generator for the Borel σ-algebra on $\mathbb{R}$. By the uniqueness theorem for probability measures, we conclude that $\mu = g \cdot m$.

Strictly speaking, we have not yet shown that $g \cdot m$ is a probability measure. But we can pick a sequence of endpoints a_n, b_n consisting of continuity points of μ, such that

$$(a_n, b_n) \to (-\infty, \infty), \qquad \text{as } n \to \infty.$$

So $\int_{-\infty}^{\infty} g(x)dx = \mu((-\infty,\infty)) = 1$. The density g is also continuous (actually uniformly continuous) due to Lemma 1.12. This then implies that g cannot be negative. If g were negative in a point x_0 then it would also be negative in an interval (a, b) around x_0. But then (1.3) would give $\mu((a, b))$ a negative value, which is impossible. □

Now we can establish the connection between the characteristic function of a probability measure μ and its moments (if they exists).

Theorem 1.15. *Let ϕ be the characteristic function of a probability measure μ on $(\mathbb{R}, \mathbb{B})$. If μ has k^{th} moment, then ϕ is a C^k-function. Furthermore, its derivatives are given by*

$$\phi^{(k)}(\theta) = i^k \int x^k e^{i\theta x} \; d\mu(x),$$

for all $\theta \in \mathbb{R}$.

Proof. Suppose that μ has n^{th} moment. We prove the assertion by simple induction, letting $k \leq n$ denote the order of differentiability. Let $k = 1$, we observe that $\left|\frac{d}{d\theta}\cos(\theta x)\right| = |-x\sin(\theta x)| \leq |x|$. By the assumption that μ has 1^{st} moment, the last function is integrable. Hence the function $\theta \mapsto \int \cos(\theta x)\ d\mu(x)$ is differentiable and (see Appendix A.9 in [4])

$$\frac{d}{d\theta}\int \cos(\theta x)\ d\mu(x) = -\int x\sin(\theta x)\ d\mu(x).$$

The same argument also gives that

$$\frac{d}{d\theta}\int \sin(\theta x)\ d\mu(x) = \int x\cos(\theta x)\ d\mu(x).$$

Hence $\phi(\theta) = \int \cos(\theta x)\ d\mu(x) + i\int \sin(\theta x)\ d\mu(x)$ is also differentiable and

$$\phi'(\theta) = -\int x\sin(\theta x)\ d\mu(x) + i\int x\cos(\theta x)\ d\mu(x) = i\int xe^{i\theta x}\ d\mu(x).$$

Continuity of ϕ' follows from Lemma 1.12. This proves the induction start. Now let ϕ be C^{k-1}. We want to show that $\phi \in C^k$. We have, respectively, that

$$\begin{aligned}
\left|\frac{d}{d\theta}x^{k-1}\cos(\theta x)\right| &= \left|-x^k\sin(\theta x)\right| \leq \left|x^k\right| \\
\left|\frac{d}{d\theta}x^{k-1}\sin(\theta x)\right| &= \left|x^k\cos(\theta x)\right| \leq \left|x^k\right|,
\end{aligned}$$

so the left hand side is integrable (μ has n^{th} moment, hence also k^{th} moment). It follows that the functions $\theta \mapsto \int x^{k-1}\cos(\theta x)\ d\mu(x)$ and $\theta \mapsto \int x^{k-1}\sin(\theta x)\ d\mu(x)$ are differentiable, hence $\phi^{(k-1)}$ is also differentiable. The induction assumption gives

$$\begin{aligned}
\phi^{(k)}(\theta) &= (\phi^{(k-1)})'(\theta) = i^{(k-1)}\left(-\int x^k\sin(\theta x)\ d\mu(x) + i\int x^k\cos(\theta x)\ d\mu(x)\right) \\
&= i^k\int x^k e^{i\theta x}\ d\mu(x).
\end{aligned}$$

Continuity follows from Lemma 1.12. □

We quickly see that this result gives a new approach to obtain the moments of a measure. If we know that the measure has k^{th} moment, and we know its characteristic function ϕ, then the k^{th} moment of μ is equal to $-i^k\phi^{(k)}(0)$. As a further consequence of the above result, it is possible, under additional assumptions, to approximate the characteristic function of a measure with its moments. In order to obtain this result, we will need the following lemma:

Lemma 1.16. *For any $n \geq 0$, we have that*

$$\left|e^{iy} - \sum_{k=0}^{n}\frac{(iy)^k}{k!}\right| \leq \min\left(2\frac{|y|^n}{n!}, \frac{|y|^{n+1}}{(n+1)!}\right), \quad y \in \mathbb{R}. \tag{1.4}$$

Proof. We prove (1.4) for all $y > 0$, the proof for $y \leq 0$ is completely analogous. Let $y > 0$, we will prove the following version of Taylor's formula with remainder

$$e^{iy} = \sum_{k=0}^{n} \frac{(iy)^k}{k!} + \frac{i^{n+1}}{n!} \int_0^y (y-x)^n e^{ix} \, dx, \tag{1.5}$$

which will be proved by induction on the non-negative integer n. The induction start is clear. For the induction step fix $n \in \mathbb{N}$ and observe that integration by parts gives us

$$\begin{aligned} \int_0^y (y-x)^n e^{ix} \, dx &= \left[-\frac{(y-x)^{n+1}}{n+1} e^{ix} \right]_{x=0}^{x=y} - \int_0^y -\frac{(y-x)^{n+1}}{n+1} i e^{ix} \, dx \\ &= \frac{y^{n+1}}{n+1} + \frac{i}{n+1} \int_0^y (y-x)^{n+1} e^{ix} \, dx. \end{aligned}$$

Hence

$$\begin{aligned} e^{iy} &= \sum_{k=0}^{n} \frac{(iy)^k}{k!} + \frac{i^{n+1}}{n!} \int_0^y (y-x)^n e^{ix} \, dx \\ &= \sum_{k=0}^{n} \frac{(iy)^k}{k!} + \frac{i^{n+1}}{n!} \left(\frac{y^{n+1}}{n+1} + \frac{i}{n+1} \int_0^y (y-x)^{n+1} e^{ix} \, dx \right) \\ &= \sum_{k=0}^{n+1} \frac{(iy)^k}{k!} + \frac{i^{n+2}}{(n+1)!} \int_0^y (y-x)^{n+1} e^{ix} \, dx, \end{aligned}$$

thus proving (1.5). With this available we observe that, for $n \geq 0$

$$\left| e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!} \right| \leq \frac{1}{n!} \int_0^y (y-x)^n \, dx = \frac{y^{n+1}}{(n+1)!}.$$

So we have proved the first half of (1.4). For the other half, consider (1.5) for $n-1$, (where $n \geq 1$), then

$$\begin{aligned} e^{iy} &= \sum_{k=0}^{n-1} \frac{(iy)^k}{k!} + \frac{i^n}{(n-1)!} \int_0^y (y-x)^{n-1} e^{ix} \, dx \\ &= \sum_{k=0}^{n} \frac{(iy)^k}{k!} + \frac{i^n}{(n-1)!} \int_0^y (y-x)^{n-1} (e^{ix} - 1) \, dx. \end{aligned}$$

Hence

$$\left| e^{iy} - \sum_{k=0}^{n} \frac{(iy)^k}{k!} \right| \leq \frac{1}{(n-1)!} \left| \int_0^y (y-x)^{n-1} (e^{ix} - 1) \, dx \right| \leq \frac{1}{(n-1)!} \frac{2y^n}{n}.$$

The conclusion also holds for $n = 0$, thus completing the proof. $\square$

Theorem 1.17. *Let X be a real-valued random variable defined on a probability space $(\Omega, \mathbb{F}, P)$ and let ϕ be the characteristic function of the probability measure. If X has n^{th} moment, for some $n \in \mathbb{N}$, then*

$$\phi(\theta) = 1 + \sum_{k=1}^{n} \frac{(i\theta)^k}{k!} E(X^k) + o(|\theta|^n), \qquad \textit{as } \theta \to 0.$$

Proof. What we need to prove, is that

$$\frac{\left|\phi(\theta) - 1 - \sum_{k=1}^{n} \frac{(i\theta)^k}{k!} E(X^k)\right|}{|\theta|^n} \to 0, \quad \text{as } \theta \to 0.$$

We replace y by θX in (1.4) and obtain

$$\left| e^{i\theta X} - \sum_{k=0}^{n} \frac{(i\theta X)^k}{k!} \right| \leq |\theta|^n \min\left(2\frac{|X|^n}{n!}, \frac{|\theta|\,|X|^{n+1}}{(n+1)!} \right).$$

Integrating the above with respect to P and using Theorem 1.15 reveals

$$\begin{aligned} \left|\phi(\theta) - 1 - \sum_{k=1}^{n} \frac{(i\theta)^k}{k!} E(X^k)\right| &\leq \int \left| e^{i\theta X} - \sum_{k=0}^{n} \frac{(i\theta X)^k}{k!} \right| \, dP \\ &\leq |\theta|^n \int \min\left(2\frac{|X|^n}{n!}, \frac{|\theta|\,|X|^{n+1}}{(n+1)!} \right) \, dP. \end{aligned}$$

The last integrand converges to 0, as $\theta \to 0$, bounded by the function $\theta \mapsto 2\,|X|^n / n!$, which is integrable by assumption. By the Lebesgue dominated convergence theorem, we have proved the claim. $\square$

Chapter 2

The Classical Moment Problem

Let $(\alpha_n)_{n\in\mathbb{N}_0}$ be a sequence of real numbers, where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. The classical moment problem focuses on two questions: Under what conditions is $(\alpha_n)_{n\in\mathbb{N}_0}$ the moment sequence of some measure μ on $\mathbb{R}$? This we call the existence part. Secondly if $(\alpha_n)_{n\in\mathbb{N}_0}$ is a moment sequence of some measure μ, under what conditions is μ the only measure having this property? This we call the uniqueness part. In this chapter we consider these two questions.

2.1 Uniqueness

If a distribution has all moments and it is the only distribution with these moments, then we say that the distribution is *uniquely determined by its moments* (or *M-determinate*); if not we say it is *M-indeterminate*. The focus of the first section in this chapter is to establish sufficient conditions for a given sequence of real numbers to be the moments of **at most one** measure. One of these sufficient conditions, which we discuss next, involves the support of the measure.

2.1.1 Bounded Support

As a consequence of Theorem 1.6 and the Weierstrass approximation theorem (see Appendix A), we obtain the following result:

Theorem 2.1. *Let μ and ν be probability measures on $(\mathbb{R}, \mathbb{B})$ concentrated on a bounded interval I. If μ and ν have the same moments, then $\mu = \nu$.*

Proof. The maps $x \mapsto x^k$ are bounded on I, so the moments always exist for measures concentrated on bounded intervals. By hypothesis, the measures μ and ν have the same moments, hence polynomials of the form $p(x) = \sum_{i=1}^{n} a_i x^i$ are integrable and

$$\int p(x)\ d\mu(x) = \sum_{i=1}^{n} a_i \int x^i\ d\mu(x) = \sum_{i=1}^{n} a_i \int x^i\ d\nu(x) = \int p(x)\ d\nu(x).$$

Let $f \in C_b(\mathbb{R})$. It follows that f is continuous on $[a, b]$ for some $a, b \in \mathbb{R}$ such that $I \subseteq [a, b]$. For any $\epsilon > 0$, the Weierstrass approximation theorem ensures the existence of a polynomial p such that $|f(x) - p(x)| < \epsilon$, for all $x \in I$. Since μ and ν are concentrated on I, we deduce that

$$\left|\int f\ d\mu - \int p\ d\mu\right| \leq \int_I |f - p|\, d\mu \leq \epsilon.$$

Similarly for ν. It follows that

$$\left|\int f\ d\mu - \int f\ d\nu\right| \quad \leq \quad \left|\int f\ d\mu - \int p\ d\mu\right| + \left|\int f\ d\mu - \int p\ d\nu\right| \leq 2\epsilon.$$

An application of Theorem 1.6 shows that $\mu = \nu$. □

So if a probability measure μ is concentrated on a bounded interval, it is uniquely determined by its moments. We therefore have the following corollary:

Corollary 2.2. *The B-distribution and the uniform distribution are uniquely determined by their moments.*

2.1.2 Further Sufficient Conditions

Lemma 2.3. *Let X be a real-valued random variable on a probability space, which has all moments, and let $a, b \in \mathbb{R}$. If X is uniquely determined by its moments, then so is $aX + b$.*

Proof. Since X has all moments, so does $aX + b$. Let Y be another real-valued random variable also having all moments. Assume that $E((aX+b)^k) = E(Y^k)$ holds for all $k \in \mathbb{N}$, we want to show that $aX + b = Y$. Set $Z = (Y - b)/a$, then we have $E((aX + b)^k) = E((aZ + b)^k)$ for all $k \in \mathbb{N}$. We prove, by complete induction on k, that $E(X^k) = E(Z^k)$ holds for all $k \in \mathbb{N}$. The induction start clearly holds, since

$$aE(X) + b = E(aX + b) = E(aZ + b) = aE(Z) + b.$$

Now suppose that $E(X^r) = E(Z^r)$ holds for all positive integers $r < k$. Then

$$\sum_{i=0}^{k} \binom{k}{i} b^{r-i} a^i E(X^i) = E((aX + b)^k) = E((aZ + b)^k) = \sum_{i=0}^{k} \binom{k}{i} b^{r-i} a^i E(Z^i).$$

We see that all the terms, except the last, cancel. Hence $E(X^k) = E(Z^k)$, by the assertion that X is uniquely determined by its moments, we conclude that $X = Z$. Therefore $aX + b = aZ + b = Y$, so the proof is complete. □

Theorem 2.4. *Let μ be a probability measure on $(\mathbb{R}, \mathbb{B})$ having all moments $(\alpha_k)_{k\in\mathbb{N}}$. If the power series*

$$\sum_{k=1}^{\infty} \alpha_k \frac{r^k}{k!},$$

has a positive radius of convergence r, then μ is the only measure having $(\alpha_k)_{k\in\mathbb{N}}$ as moments.

Proof. Let $(\beta_k)_{k\in\mathbb{N}}$ denote the sequence of absolute moments of μ. First we want to show that

$$\frac{\beta_k r^k}{k!} \to 0, \qquad \text{as } k \to \infty, \tag{2.1}$$

for some positive r. By hypothesis, there exists $s \in (0,1)$ such that $\lim_{k\to\infty}(\alpha_k s^k)/k! = 0$, since the series converges. Now choose $0 < r < s$, then (2.1) holds for the sequence of even k's, since in this case $\alpha_k = \beta_k$. Furthermore $2kr^{2k-1} < s^{2k}$ holds for k large enough, so

$$\frac{|x|^{2k-1}\, r^{2k-1}}{(2k-1)!} \leq \frac{r^{2k-1}}{(2k-1)!} + \frac{|x|^{2k}\, r^{2k-1}}{(2k-1)!} \leq \frac{r^{2k-1}}{(2k-1)!} + \frac{|x|^{2k}\, s^{2k}}{(2k)!},$$

where we used that $|x|^{2k-1} \leq 1 + |x|^{2k}$. Integrating with respect to μ reveals

$$\frac{\beta_{2k-1} r^{2k-1}}{(2k-1)!} \leq \frac{r^{2k-1}}{(2k-1)!} + \frac{\beta_{2k} s^{2k}}{(2k)!} = \frac{r^{2k-1}}{(2k-1)!} + \frac{\alpha_{2k} s^{2k}}{(2k)!} \to 0, \qquad \text{as } k \to \infty.$$

Hence (2.1) also holds for the sequence of odd k's. Hence (2.1) must hold. By Lemma 1.16, with $y = hx$ we have

$$\left| e^{i(\theta+h)x} - \sum_{k=0}^{n} \frac{(ihx)^k}{k!} e^{i\theta x} \right| = \left| e^{i\theta x} \left(e^{ihx} - \sum_{k=0}^{n} \frac{(ihx)^k}{k!} \right) \right| = \left| e^{ihx} - \sum_{k=0}^{n} \frac{(ihx)^k}{k!} \right| \leq \frac{|hx|^{n+1}}{(n+1)!}$$

Hence by integrating with respect to μ and using Theorem 1.15 we get that

$$\begin{aligned} \left| \phi(\theta+h) - \sum_{k=0}^{n} \frac{h^k}{k!} \phi^{(k)}(\theta) \right| &\leq \int \left| e^{i(\theta+h)x} - \sum_{k=0}^{n} \frac{(ihx)^k}{k!} e^{i\theta x} \right| \, d\mu(x) \\ &\leq \int \frac{|hx|^{n+1}}{(n+1)!} \, d\mu(x) = \frac{|h|^{n+1} \beta_{n+1}}{(n+1)!}. \end{aligned}$$

Using (2.1) we conclude that

$$\phi(\theta+h) = \sum_{k=0}^{\infty} \frac{\phi^{(k)}(\theta)}{k!} h^k, \qquad \text{for } |h| \leq r.$$

Similarly, if ν were another probability measure on $(\mathbb{R}, \mathbb{B})$ with moments $(\alpha_k)_{k\in\mathbb{N}}$ and characteristic function ψ, then

$$\psi(\theta+h) = \sum_{k=0}^{\infty} \frac{\psi^{(k)}(\theta)}{k!} h^k, \qquad \text{for } |h| \leq r.$$

Let $t = 0$. Since $\phi^{(k)}(0) = i^k \alpha_k = \psi^{(k)}(0)$, it follows that ϕ and ψ agree on $(-r, r)$. Hence their derivatives will also agree on $(-r, r)$. Then letting $t = r - \epsilon$ and $t = -r + \epsilon$ reveals that ϕ and ψ agree on $(-2r + \epsilon, 2r - \epsilon)$. Therefore they agree on $(-2r, 2r)$. Repeating this argument gives us that ϕ and ψ will agree on $\mathbb{R}$. By Theorem 1.10, the proof is complete. $\square$

Now we can prove the following corollary:

Corollary 2.5. *The normal-, gamma- and Laplace-distribution are uniquely determined by their moments.*

Proof. We consider the power series in Theorem 2.4,

$$\sum_{k=1}^{\infty} \alpha_k \frac{r^k}{k!}.$$

Here α_k denotes the moment k^{th} moment, and they are listed in Table 1.1. We know from the root-criteria for power series, that the above has positive radius of convergence if

$$\limsup_{k\to\infty} \left|\frac{\alpha_k}{k!}\right|^{1/k} < \infty.$$

Due to Lemma 2.3, we only need to consider the standard normal distribution, the standard Laplace distribution and the gamma distribution with scale parameter $\beta = 1$.

For the standard normal distribution, we only need to consider the even moments. We use that the gamma-function is increasing, when the argument is greater than 2. So for $k > 2$ and even,

$$\left(\frac{\alpha_k}{k!}\right)^{1/k} = \left(2^{k/2}\frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi}k!}\right)^{1/k} = \sqrt{2}\left(\frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi}\Gamma(k+1)}\right)^{1/k} \leq \sqrt{2}.$$

So the sequence $((\alpha_k/k!)^{1/k})_{k\in\mathbb{N}}$ is bounded, by the root-criteria we conclude that the conditions in Theorem 2.4 are met.

For gamma distribution with scale parameter 1, we see that

$$\left(\frac{\alpha_k}{k!}\right)^{1/k} = \left(\frac{\frac{\Gamma(\lambda+k)}{\Gamma(\lambda)}}{k!}\right)^{1/k} = \left(\frac{\prod_{j=1}^{k}(\lambda+j-1)}{\prod_{j=1}^{k} j}\right)^{1/k} \leq \left(\prod_{j=1}^{k}(\lambda+1)\right)^{1/k} = \lambda+1.$$

We therefore see that the sequence $((\alpha_k/k!)^{1/k})_{k\in\mathbb{N}}$ is bounded, by the root-criteria we conclude that the conditions in Theorem 2.4 are met. Hence the gamma distribution is uniquely determined by its moments.

For the standard Laplace distribution, we have, for k even

$$\left(\frac{\alpha_k}{k!}\right)^{1/k} = \left(\frac{k!}{k!}\right)^{1/k} = 1.$$

Hence the Laplace distribution is uniquely determined by its moments. $\square$

2.1.3 Examples of M-indeterminate Distributions

Given a measure μ on $(\mathbb{R}, \mathbb{B})$ having all moments, in general it may not be the only measure with those moments. The following two examples, illustrate this (cf. [4], section 2.3).

Example 2.6 (Heyde (1963))**.** Consider the *lognormal distribution*, i.e., the measure on $(\mathbb{R}, \mathbb{B})$ given by the following density with respect to the Lebesgue measure:

$$f_0(x) = \begin{cases} (2\pi)^{-\frac{1}{2}} x^{-1} e^{-\log(x)^2/2}, & x > 0 \\ 0 & \text{else.} \end{cases}$$

It is well-known that $\mu_0 = f_0 \cdot m$ is a probability measure. For $\alpha \in [-1, 1]$ we will justify that $f_\alpha(x) = f_0(x)\,(1 + \alpha \sin(2\pi \log(x)))$ is a probability density with respect to the Lebesgue measure. First f_α is clearly measurable. We also have that $\alpha \sin(2\pi \log(x)) \in [-1, 1]$ for all x, so f_α is non-negative. For $k \in \mathbb{N}_0$ we have

$$\begin{aligned} \int x^k \; d\mu_\alpha(x) &= \int_0^\infty x^k f_0(x)\,(1 + \alpha \sin(2\pi \log(x))) \; dx \\ &= \int_0^\infty x^k f_0(x) \; dx + \alpha \int_0^\infty x^k f_0(x) \sin(2\pi \log(x)) \; dx. \end{aligned}$$

Since the above integrands are integrable (they are bounded by the m-integrable function $x \mapsto 2\,|x|^k\, f_0(x)$). So now we calculate the last integral. We use the substitution $y = \log(x)$

$$\int_0^\infty x^k f_0(x) \sin(2\pi \log(x)) \; dx = \int_{-\infty}^\infty (2\pi)^{-\frac{1}{2}} e^{ky - y^2/2} \sin(2\pi y) \; dy.$$

We rewrite the exponent

$$ky - \frac{y^2}{2} = -\frac{1}{2}(y^2 - 2ky) = -\frac{1}{2}((y-k)^2 - k^2) = \frac{k^2}{2} - \frac{(y-k)^2}{2},$$

so the integral is

$$\begin{aligned} \int_0^\infty x^k f_0(x) \sin(2\pi \log(x)) \; dx &= \int_{-\infty}^\infty (2\pi)^{-\frac{1}{2}} e^{k^2/2} e^{(y-k)^2/2} \sin(2\pi y) \; dy \\ &= \int_{-\infty}^\infty (2\pi)^{-\frac{1}{2}} e^{k^2/2} e^{t^2/2} \sin(2\pi(t+k)) \; dt \\ &= \int_{-\infty}^\infty (2\pi)^{-\frac{1}{2}} e^{k^2/2} e^{t^2/2} \sin(2\pi t) \; dt = 0. \end{aligned}$$

With use of the substition $t = y - k$, periodicity of sin and the last integrand being an odd function. So we have found uncountably many probability measures all having the same moments. By setting α to 0, we conclude that the standard log-normal distribution is not uniquely determined by its moments.

Example 2.7. Let $\lambda \in (0, 1)$ and $\alpha \in [-1, 1]$, then set

$$f_{\alpha,\lambda}(x) = c_\lambda e^{-|x|^\lambda} \left(1 + \alpha \sin(\beta\,|x|^\lambda \operatorname{sgn}(x))\right),$$

where $\beta = \tan(\lambda\pi/2)$ and $1/c_\lambda = \int e^{-|x|^\lambda} < \infty$. It is well-known that $c_\lambda e^{-|x|^\lambda}$ is a probability density with respect to m and it has all moments. The functions $f_{\alpha,\lambda}$ are measurable and non-negative since $\alpha \in [-1,1]$. The following integral exists (the integrand is bounded by $|x|^n e^{-|x|^\lambda}$)

$$\int x^n e^{-|x|^\lambda} \sin(\beta |x|^\lambda \operatorname{sgn}(x)) \, dx.$$

We claim it is 0 for all $n \in \mathbb{N}_0$, we would then get

$$\begin{aligned}\int x^n f_{\alpha,\lambda}(x) \, dm(x) &= \int x^n c_\lambda e^{-|x|^\lambda} \, dm(x) + \alpha c_\lambda \int x^n e^{-|x|^\lambda} \sin(\beta |x|^\lambda \operatorname{sgn}(x)) \, dm(x) \\ &= \int x^n c_\lambda e^{-|x|^\lambda} \, dm(x).\end{aligned}$$

Thus $(\mu_\alpha = f_{\alpha,\lambda} \cdot m)_{\alpha \in [-1,1]}$ is a family of probability measures having the same moments for each fixed λ. So now we prove the claim. We look at the integrand $x \mapsto x^n e^{-|x|^\lambda} \sin(\beta |x|^\lambda \operatorname{sgn}(x))$. If n is even then the integrand is odd and the claim holds. So now let n be odd, then the integrand is even and we only have to prove the claim with integration domain $[0,\infty)$, so we need to show

$$\int_0^\infty x^n e^{-x^\lambda} \sin(\beta x^\lambda) \, dx = 0,$$

for n odd. Using the identity with $q \in \mathbb{C}$ and $\mathrm{Re}(q) > 0$

$$\int_0^\infty t^{p-1} e^{-qt} \, dt = \frac{\Gamma(p)}{q^p},$$

with $p = (n+1)/\lambda$, $q = 1 + i\beta$ and the substitution $t = x^\lambda$ we get

$$\begin{aligned}\frac{\Gamma(\frac{n+1}{\lambda})}{(1+i\beta)^{\frac{n+1}{\lambda}}} &= \int_0^\infty x^{\lambda(\frac{n+1}{\lambda}-1)} e^{-(1+i\beta)x^\lambda} \lambda x^{\lambda-1} \, dx \\ &= \lambda \int_0^\infty x^n e^{-x^\lambda} \cos(\beta x^\lambda) \, dx - i\lambda \int_0^\infty x^n e^{-x^\lambda} \sin(\beta x^\lambda) \, dx. \qquad (2.2)\end{aligned}$$

We then focus on the denominator $(1+i\beta)^{\frac{n+1}{\lambda}}$. Since

$$1 + i\beta = 1 + i\tan(\lambda\pi/2) = 1 + \frac{i\sin(\lambda\pi/2)}{\cos(\lambda\pi/2)} = \frac{e^{i\lambda\pi/2}}{\cos(\lambda\pi/2)},$$

we get that $(1+i\beta)^{(n+1)/\lambda} = e^{(i\lambda\pi(n+1))/(2\lambda)} / \cos(\lambda\pi/2)^{(n+1)/\lambda}$. Since $n+1$ is even the right-hand side is real, thus (2.2) reveals the claim.

2.1.4 The Weibull Distribution and the Krein Condition

So far we have studied whether certain commonly used distributions are uniquely determined by their moments. The only distribution listed in Table 1.1 that has all moments, but we have not yet considered is the Weibull distribution. As it turns out, this is more difficult to handle, and surprisingly the question whether or not it is uniquely determined by its moments depends on the parameter value. In order to show this, we state the following theorem from [8] without proof:

Theorem 2.8. *Let f be a probability density with respect to the Lebesgue measure, that has all moments and is concentrated on $(0,\infty)$. If the following Krein condition holds:*

$$\int_0^\infty \frac{-\log(f(x^2))}{1+x^2}\,dx < \infty,$$

then the measure $f \cdot m$ is M-indeterminate. If, however,

$$\int_0^\infty \frac{-\log(f(x^2))}{1+x^2}\,dx = \infty,$$

and f is differentiable and there exists $x_0 > 0$, such that for $x \geq x_0$ we have that $(-xf'(x))/f(x) \nearrow \infty$, as $x \to \infty$, then $f \cdot m$ is M-determinate.

Corollary 2.9. *The Weibull distribution with parameter $c > 0$ (see Table 1.1) is uniquely determined by its moments if and only if $c \geq 1/2$.*

Proof. The density of the Weibull distribution is given by $f(x) = c\beta^{-c}x^{c-1}e^{-x^c/\beta^c}$, for $x > 0$. We want to apply Theorem 2.8. Observe that

$$\int_0^\infty \frac{-\log(f(x^2))}{1+x^2}\,dx = \int_0^\infty \frac{-\log\left(\frac{c}{\beta^c}\right)}{1+x^2} - \frac{2(c-1)\log(x)}{1+x^2} + \frac{1}{\beta^c}\frac{x^{2c}}{1+x^2}\,dx.$$

The two first terms in the integrand have finite integral, so the above if finite (respectively infinite) if and only if $\int_1^\infty x^{2c}/(1+x^2)\,dx$ is finite (respectively infinite).

If $c \geq 1/2$, then we have that

$$\int_1^\infty \frac{x^{2c}}{1+x^2}\,dx \geq \int_1^\infty \frac{x}{1+x^2}\,dx \geq \int_1^\infty \frac{x/2}{x^2}\,dx = \infty.$$

Furthermore we see that f is differentiable and

$$f'(x) = \frac{c}{\beta^c}\left((c-1)x^{c-2}e^{-\frac{x^c}{\beta^c}} - x^{c-1}e^{-\frac{x^c}{\beta^c}}\frac{cx^{c-1}}{\beta^c}\right) = \frac{c}{\beta^c}e^{-\frac{x^c}{\beta^c}}\left((c-1)x^{c-2} - \frac{c}{\beta^c}x^{2c-2}\right).$$

So we get that for all $c > 0$

$$\frac{-xf'(x)}{f(x)} = \frac{x\frac{c}{\beta^c}e^{-\frac{x^c}{\beta^c}}\left((c-1)x^{c-2} - \frac{c}{\beta^c}x^{2c-2}\right)}{\frac{c}{\beta^c}x^{c-1}e^{-\frac{x^c}{\beta^c}}} = (1-c) + \frac{c}{\beta^c}x^c \nearrow \infty, \quad \text{as } x \to \infty.$$

By Theorem 2.8 we conclude that if $c \geq 1/2$, then the Weibull distribution with parameter c is M-determinate.

Conversely, if $0 < c < 1/2$, then we see that

$$\int_1^\infty \frac{x^{2c}}{1+x^2}\, dx \leq \int_1^\infty x^{2c-2}\, dx < \infty,$$

since $2c - 2 < -1$. By Theorem 2.8 we conclude that the Weibull distribution is M-indeterminate for parameter values in $(0, 1/2)$. Hence the proof is complete. □

2.2 Existence

In this section, we will prove the existence part of the classical moment problem. We will need some definitions and lemmas. First we prove that every non-negative polynomial is a sum of squares, more precisely:

Lemma 2.10. *If $p : \mathbb{R} \to \mathbb{R}$ is a non-negative polynomial, then there exists polynomials $q_1, q_2, ..., q_k$ such that*

$$p(x) = q_1(x)^2 + q_2(x)^2 + ... + q_k(x)^2, \quad \textit{for all } x \in \mathbb{R}.$$

Proof. We prove it by complete induction on $n = \deg(p)$. For $n = 0$, p is a non-negative constant and the claim obviously holds. Let $n > 0$ and $\deg(p) = n$ and assume that the claim holds for all non-negative polynomials of order less than n. Set

$$k = \inf \{p(x) \mid x \in \mathbb{R}\} = \min \{p(x) \mid x \in \mathbb{R}\} \geq 0.$$

Set $g(x) = p(x) - k$, for all $x \in \mathbb{R}$. Then g is a non-negative polynomial and g has at least one real root c. The multiplicity of c is necessarily even, since the complex roots come in conjugate pairs. We factor the polynomial g, so

$$g(x) = (x - c)^2 h(x), \quad x \in \mathbb{R},$$

for some polynomial h, which has degree less than n. By the induction hypothesis h is a sum of squares, hence

$$f(x) = (x - c)^2 h(x) + (\sqrt{k})^2, \quad x \in \mathbb{R}$$

is a sum of squares. Thus completing the proof. □

Next we will need a very important result from functional analysis, the Hahn-Banach extension theorem. First we need to define the following

Definition 2.11. *Let X be vector space over $\mathbb{R}$. A* sublinear *functional on X is a map $p : X \to \mathbb{R}$, such that*

1. $p(x + y) \leq p(x) + p(y), \quad$ *for all $x, y \in X$*
2. $p(\lambda x) = \lambda p(x), \quad$ *for all $x \in X, \lambda \geq 0$*

A seminorm on X is an example of a sublinear functional on X. The proof of The Hahn-Banach extension Theorem follows [5] section 5.2.

Theorem 2.12 (The Hahn-Banach extension Theorem)**.** *Let X be a vector space over $\mathbb{R}$ and let $p : X \to \mathbb{R}$ be sublinear functional on X. If M is subspace of X and $f : M \to \mathbb{R}$ is a linear functional on M, such that $f(x) \leq p(x)$ for all $x \in M$, then there exists a linear functional $F : X \to \mathbb{R}$ such that $F|_M = f$ and $F(x) \leq p(x)$ for all $x \in X$.*

Proof. First we show that if $x \in X \setminus M$, then there exists a linear functional $g : M + \mathbb{R}x \to \mathbb{R}$, such that $g|_M = f$ and $g(y) \leq p(y)$, for all $y \in M + \mathbb{R}x$. Observe that for $y_1, y_2 \in M$ we have that

$$f(y_1) + f(y_2) = f(y_1 + y_2) \leq p(y_1 + y_2) = p(y_1 - x + x + y_2) \leq p(y_1 - x) + p(x + y_2).$$

Hence we deduce that $\sup\{f(y) - p(y - x) \mid y \in M\} \leq \inf\{p(x + y) - f(y) \mid y \in M\}$. Let α be a number satisfying

$$\sup\{f(y) - p(y - x) \mid y \in M\} \leq \alpha \leq \inf\{p(x + y) - f(y) \mid y \in M\},$$

and define $g : M + \mathbb{R}x \to \mathbb{R}$ by $g(y + \lambda x) = f(y) + \lambda\alpha$, for $y \in M$. Clearly g is a linear functional on $M + \mathbb{R}x$ and $g|_M = f$. So we have that $g(y) \leq p(y)$, for $y \in M$. We have for $\lambda > 0$

$$g(y + \lambda x) = \lambda\left(f\left(\frac{y}{\lambda}\right) + \alpha\right) \leq \lambda\left(f\left(\frac{y}{\lambda}\right) + p\left(x + \frac{y}{\lambda}\right) - f\left(\frac{y}{\lambda}\right)\right) = p(y + \lambda x).$$

For $\lambda < 0$, set $\mu = -\lambda > 0$ and we get that

$$g(y + \lambda x) = \mu\left(f\left(\frac{y}{\mu}\right) - \alpha\right) \leq \mu\left(f\left(\frac{y}{\mu}\right) - f\left(\frac{y}{\mu}\right) + p\left(\frac{y}{\mu} - x\right)\right) = p(y + \lambda x).$$

Hence $g(y) \leq p(y)$, for all $y \in M + \mathbb{R}x$.

Now we consider the family of maps

$$\mathcal{F} = \{F : Y \to \mathbb{R} \mid M \subseteq Y \text{ subspace of } X,\ F \text{ linear},\ F|_M = f,\ F(y) \leq p(y), \text{ for all } y \in Y\}$$

We define a relation $\preceq$ on $\mathcal{F}$ as follows: If $F_1, F_2 \in \mathcal{F}$ with domains Y_1, Y_2, respectively, we say that $F_1 \preceq F_2$ if $Y_1 \subseteq Y_2$ and $F_2|_{Y_1} = F_1$. It is easily verified that $\preceq$ is a partial ordering on $\mathcal{F}$. We claim that any family of maps $(F_i)_{i\in A}$ in $\mathcal{F}$, that is totally ordered with respect to $\preceq$, has an upper bound. For every $i \in A$, let Y_i be the domain of F_i. Set $Y = \bigcup_{i\in A} Y_i$. Then Y is a subspace of X, containing M. Indeed if $y_1, y_2 \in Y$, then $y_1 \in Y_{i_1}$ and $y_2 \in Y_{i_2}$ for some $i_1, i_2 \in A$. By the total ordering of $(F_i)_{i\in A}$, we conclude that $y_1, y_2 \in Y_{i_1}$ or $y_1, y_2 \in Y_{i_2}$, either way $\lambda_1 y_1 + \lambda_2 y_2 \in Y$ for all $\lambda_1, \lambda_2 \in \mathbb{R}$, since Y_{i_1} and Y_{i_2} are subspaces of X.

Then define $\tilde{F} : Y \to \mathbb{R}$ as follows: for $y \in Y$, then $y \in Y_i$ for a $i \in A$, set

$$\tilde{F}(y) = F_i(y).$$

First, we justify that $\tilde{F}$ is well-defined. Suppose $y \in Y_i$ and $y \in Y_{i'}$. By the total ordering of $(F_i)_{i\in A}$, we can assume that $F_i \preceq F_{i'}$. Then $y \in Y_i \subseteq Y_{i'}$ and $F_{i'}|_{Y_i} = F_i$, so

$F_{i'}(y) = F_i(y)$. Hence $\tilde{F}$ is well-defined. It is easily verified that $\tilde{F} \in \mathcal{F}$, simply by using the properties of F_i. Also clearly $F_i \preceq \tilde{F}$, for all $i \in A$, hence $\mathcal{F}$ has an upper bound.

By Zorn's Lemma there exists a maximal element $F \in \mathcal{F}$ for the partial ordered set $(\mathcal{F}, \preceq)$. We now verify that F is the extension of f that the theorem claims exist. By construction it is indeed a extension of f, however we need to show that F is defined on the whole space X. Now suppose that the domain X' of F is not the whole space X. Let $x \in X \setminus X'$. By the argument in the beginning of the proof, there exists a linear functional $g : X' + \mathbb{R}x \to \mathbb{R}$, such that $g|_{X'} = F$ and $g(y) \leq p(y)$, for all $y \in X' + \mathbb{R}x$. It is easily seen that $g \in \mathcal{F}$ and that $F \preceq g$ and $F \neq g$, contradicting the fact that F was maximal in $(\mathcal{F}, \preceq)$. We conclude that $X' = X$ and thus proving the theorem. □

The following Corollary, due to Choquet (1962), follows [1].

Corollary 2.13. *Let X be a vector space over $\mathbb{R}$ and let M be a subspace of X. Let E be a non-empty, convex subset of X, that is closed under multiplication with non-negative constants and with the property that $x \in E$ and $-x \in E$ implies $x = 0$. Assume furthermore that $M + E = X$. Then every linear functional f on M, which is non-negative on $M \cap E$, can be extended to a linear functional F on X, which is non-negative on E.*

Proof. Define a relation $\preceq$ on X as follows: $x \preceq y$ if $y - x \in E$. We justify that it is a partial order relation. Since $0 \in E$ we have that $x \preceq x$, hence $\preceq$ is reflexive. If $x \preceq y$ and $y \preceq x$, then set $z = y - x$. We then have that $z, -z \in E$, hence $y - x = z = 0$, so indeed $x = y$. Therefore $\preceq$ is anti-symmetric. Finally assume that $x \preceq y$ and $y \preceq z$, then $y - x, z - y \in E$ so by convexity $(z - x)/2 \in E$, hence $z - x \in E$, thus proving transitivity.

Observe that for every $x \in X$, there exist $y_1, y_2 \in M$ such that $y_1 \preceq x \preceq y_2$. It follows since $x \in X = M + E$, so $x = y_1 + y'$ for some $y_1 \in M$ and $y' \in E$. Then $x - y_1 = y' \in E$, so $y_1 \preceq x$. A similar argument gives the existence of y_2. Also observe that if $y_1, y_2 \in M$ and $y_1 \preceq y_2$, then $y_2 - y_1 \in M \cap E$, so since f is linear and non-negative on $M \cap E$, we get that $f(y_1) \leq f(y_2)$.

For $x \in X$, consider the set $\{f(y) \mid y \in M, x \preceq y\}$. The first observation above gives us that the set is non-empty. Furthermore it gives us that there exists $y_1 \in M$ such that $y_1 \preceq x$, so for all $y \in M$ where $x \preceq y$, we get that $f(y) \geq f(y_1)$. Hence the set is bounded from below. So

$$p(x) = \inf \{f(y) \mid y \in M, x \preceq y\}, \quad x \in X,$$

is a well-defined real-valued function on X. We also see by the second observation above, that $f(y) = p(y)$ for all $y \in M$. We easily see that for $\lambda \geq 0$ we have that $y \preceq x$ if and only if $\lambda y \preceq \lambda x$, so $p(\lambda x) = \lambda p(x)$. We also see that for $x, x' \in X$,

$$\begin{aligned} p(x) + p(x') &= \inf \{f(y) \mid y \in M, x \preceq y\} + \inf \{f(y) \mid y \in M, x' \preceq y\} \\ &= \inf \{f(y) + f(y') \mid y, y' \in M, x \preceq y, x' \preceq y'\} \\ &\geq \inf \{f(y + y') \mid y + y' \in M, (x + x')/2 \preceq (y + y')/2\} = p(x + x') \end{aligned}$$

so p is sublinear. So f is dominated by the sublinear functional p on M. By The Hahn-Banach extension Theorem, there exists a linear functional F on X, so $F|_M = f$ and $F \leq p$ on X. We only need to verify that F is non-negative on E. Let $x \in E$. We have that $x = 0 - (-x) \in E$, hence $-x \preceq 0$, so $F(-x) \leq p(-x) \leq f(0) = 0$, hence F is non-negative on E. $\square$

Please observe that in this section we let moment-sequences be label with $\mathbb{N}_0$, in contrary to $\mathbb{N}$ in the definition. The 0^{th} moment is then just understood as the measure of the whole space.

Theorem 2.14. *Let $(\alpha_n)_{n\in\mathbb{N}_0}$ be a sequence of real-numbers. There exists a Radon measure μ on $\mathbb{R}$ such that*

$$\alpha_n = \int x^n \, d\mu(x), \quad \text{for all } n \in \mathbb{N}_0,$$

if and only if for all $n \geq 0$ and all $(n+1)$-tuples $(c_0, c_1, ..., c_n)$ of real numbers we have

$$\sum_{i,j=0}^{n} c_i c_j \alpha_{i+j} \geq 0. \tag{2.3}$$

Note that the measure μ above is a probability measure if and only if $\alpha_0 = 1$.

Proof. Assume that there exists a measure μ on $\mathbb{R}$ with $(\alpha_n)_{n\in\mathbb{N}_0}$ as moments. Let $n \geq 0$ and let $(c_0, c_1, ..., c_n)$ be any $(n+1)$-tuple of real numbers. Then set the following polynomial

$$p(x) = \left(\sum_{i=0}^{n} c_i x^i\right)^2 = \sum_{i,j=0}^{n} c_i c_j x^{i+j}.$$

Then clearly $p(x) \geq 0$ for all $x \in \mathbb{R}$. So the integral with respect to μ is well-defined and non-negative:

$$0 \leq \int p \, d\mu = \int \sum_{i,j=0}^{n} c_i c_j x^{i+j} \, d\mu(x) = \sum_{i,j=0}^{n} c_i c_j \alpha_{i+j},$$

hence (2.3) holds.

Conversely, assume that (2.3) holds for all $n \geq 0$ and all $(n+1)$-tuples of real numbers. Consider the following vector spaces of functions

$$C(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous}\}, \quad C_c(\mathbb{R}) = \{f \in C(\mathbb{R}) \mid \text{supp}(f) \text{ is compact}\},$$

$$\text{Pol} = \{f \in C(\mathbb{R}) \mid f \text{ is a polynomial}\}, \ H = \{f \in C(\mathbb{R}) \mid \text{exists } p \in \text{Pol such that } |f| \leq p\}$$

Here $\text{supp}(f) = \overline{\{x \in \mathbb{R} \mid f(x) \neq 0\}}$ is the support of f. It is easily verified that these are all vector spaces. Let $H^+ = H \cap C(\mathbb{R})^+$ and $\text{Pol}^+ = \text{Pol} \cap C(\mathbb{R})^+$, where $C(\mathbb{R})^+$ is the space on non-negative continuous functions on $\mathbb{R}$.

Now consider again the sequence $(\alpha_n)_{n\in\mathbb{N}_0}$. Define the linear functional $F : \text{Pol} \to \mathbb{R}$ by

$$F(c_0 + c_1 x + c_2 x^2 + ... + c_n x^n) = \sum_{i=0}^{n} c_i \alpha_i.$$

It is easily verified that F is linear. We claim that F is a non-negative functional, i.e., $F(p) \geq 0$ whenever p is a non-negative polynomial. By Lemma 2.10 and linearity of F it suffices to prove that $F(q^2) \geq 0$ for every polynomial $q(x) = c_0 + c_1 x + c_2 x^2 + ... + c_n x^x$. We quickly realize that it must hold, since (using that (2.3) is assumed to hold)

$$F(q(x)^2) = F\left(\sum_{i,j=0}^{n} c_i c_j x^{i+j}\right) = \sum_{i,j=0}^{n} c_i c_j \alpha_{i+j} \geq 0.$$

So now we have constructed a linear functional on Pol, which is non-negative on $\mathrm{Pol}^+ = \mathrm{Pol} \cap H^+$. Observe that H^+ is convex and that if both the function g and $-g$ lie in H^+, then g must be the zero-function. We also see that $H = \mathrm{Pol} + H^+$, since $\mathrm{Pol} + H^+ \subseteq H$ obviously holds and given a function f bounded by a polynomial p, then $f = -p + (f + p) \in \mathrm{Pol} + H^+$. Now Corollary 2.13 (with $X = H$, $M = \mathrm{Pol}$, and $E = H^+$) gives us that we can extend F to a linear functional $\tilde{F}$ on H, which is non-negative on H^+. Further, since $C_c(\mathbb{R}) \subset H$, we can restrict $\tilde{F}$ to a linear functional F_0 on $C_c(\mathbb{R})$. By the Riesz representation theorem (cf. Theorem 7.2 in [5]), there exists a unique non-negative Radon measure μ on $\mathbb{R}$, such that $F_0(f) = \int f \, d\mu$ for all $f \in C_c(\mathbb{R})$.

We show that polynomials are μ-integrable. For each $n \in \mathbb{N}$, construct the function

$$\tilde{1}_{(-n,n)}(x) = \begin{cases} 1 & \text{for } x \in (-n, n) \\ x + (n+1) & \text{for } x \in [-n-1, -n] \\ -x + n + 1 & \text{for } x \in [n, n+1] \\ 0 & \text{else} \end{cases}$$

Then $\tilde{1}_{(-n,n)} \in C_c(\mathbb{R})$, for all $n \in \mathbb{N}$. Let p be a polynomial, then $|p| \in H^+$, and let q denote a polynomial that bounds $|p|$. Then $\tilde{1}_{(-n,n)} \, |p| \in C_c(\mathbb{R})$, so we get that

$$\begin{aligned} \int \tilde{1}_{(-n,n)} \, |p| \; d\mu &= F_0(\tilde{1}_{(-n,n)} \, |p|) = \tilde{F}(\tilde{1}_{(-n,n)} \, |p|) \leq \tilde{F}(\tilde{1}_{(-n,n)} \, |p|) + \tilde{F}((1 - \tilde{1}_{(-n,n)}) \, |p|) \\ &= \tilde{F}(|p|) = \tilde{F}(|p|) + \tilde{F}(q - |p|) = \tilde{F}(q) = F(q) < \infty, \end{aligned}$$

where we used that $\tilde{F}((1 - \tilde{1}_{(-n,n)}) \, |p|) \geq 0$ and $\tilde{F}(q - |p|) \geq 0$. The bound is independent of n, so by Fatou's Lemma $\int |p| \; d\mu \leq F(q)$, since $\tilde{1}_{(-n,n)} \, |p|$ converges pointwise to $|p|$, as $n \to \infty$. Hence p is μ-integrable.

Let $p \in \mathrm{Pol}^+$ be a non-negative polynomial. We want to show that $\int p \, d\mu = F(p)$. Because p is μ-integrable, we conclude by the Lebesque dominated convergence theorem, that

$$\int (1 - \tilde{1}_{(-n,n)}) p \; d\mu \to 0, \quad \text{as } n \to \infty.$$

So we get that

$$\begin{aligned} \int p \; d\mu &= \int \tilde{1}_{(-n,n)} p \; d\mu + \int (1 - \tilde{1}_{(-n,n)}) p \; d\mu \leq \tilde{F}(\tilde{1}_{(-n,n)} p) + \int (1 - \tilde{1}_{(-n,n)}) p \; d\mu \\ &\leq \tilde{F}(p) + \int (1 - \tilde{1}_{(-n,n)}) p \; d\mu = F(p) + \int (1 - \tilde{1}_{(-n,n)}) p \; d\mu, \end{aligned}$$

holds for all $n \in \mathbb{N}$. Letting $n \to \infty$ yields $\int p \, d\mu \leq F(p)$. For the other inequality observe that we can find another positive polynomial q such that $p(x)/q(x) \to 0$, as $n \to \infty$. Hence there exists $x_0 \in \mathbb{R}$, such that $x \geq x_0$ implies that $p(x) < \epsilon q(x)$. Let n be a positive integer, such that $n \geq x_0$. Then $1_{[-x_0,x_0]} \leq \tilde{1}_{(-n,n)} \leq 1$, where $\tilde{1}_{(-n,n)} \in C_c(\mathbb{R})$. Then we get that

$$0 \leq p(1 - \tilde{1}_{(-n,n)}) \leq p(1 - 1_{[-x_0,x_0]}) \leq \epsilon q, \quad \text{for all } x \in \mathbb{R}.$$

Since $\tilde{F}$ is non-negative on H^+, we get that

$$0 \leq \tilde{F}(p) - \tilde{F}(p\tilde{1}_{(-n,n)}) \leq \epsilon \tilde{F}(q).$$

Hence

$$F(p) = \tilde{F}(p) \leq \tilde{F}(p\tilde{1}_{(-n,n)}) + \epsilon \tilde{F}(q) = \int p\tilde{1}_{(-n,n)} \, d\mu + \epsilon \tilde{F}(q) \leq \int p \, d\mu + \epsilon \tilde{F}(q).$$

Since ϵ was arbitrary and independent of q, we get that $F(p) \leq \int p \, d\mu$. Hence $F(p) = \int p \, d\mu$ for all $p \in \mathrm{Pol}^+$.

Let $n \in \mathbb{N}_0$. If n is even, then x^n is a non-negative polynomial, so in particular it is the difference of two non-negative polynomials. If n is odd, then we have that $x^n = (1+x^{n+1}+x^n)-(1+x^{n+1})$, so x^n is still a difference of two non-negative polynomials. For $n \in \mathbb{N}$, let $p_{1,2} \in \mathrm{Pol}^+$ such that $x^n = p_1 - p_2$, then

$$\int x^n \, d\mu(x) = \int p_1 \, d\mu - \int p_2 \, d\mu = F(p_1) - F(p_2) = F(x^n) = \alpha_n.$$

Thus $(\alpha_n)_{n \in \mathbb{N}_0}$ is the moment-sequence of the measure μ. $\square$

Chapter 3

Method of Moments

The goal of this chapter will be to establish an important connection between two kinds of convergence: weak convergence of measures and convergence of the corresponding moments.

3.1 Weak Convergence of Probability Measures

The next definition opens up for a new concept regarding convergence of a sequence of measures. It is of great importance in asymptotic theory of mathematical statistics.

Definition 3.1. *Let $\mu, (\mu_n)_{n\in\mathbb{N}}$ be probability measures on $(\mathbb{R}, \mathbb{B})$. Let C denote the set of points of continuity of μ (i.e., the points $a \in \mathbb{R}$ where $\mu(\{a\}) = 0$). The sequence $(\mu_n)_{n\in\mathbb{N}}$ is said to converge weakly to μ if for all $a \in C$*

$$\mu_n(-\infty, a] \to \mu(-\infty, a], \quad \text{as } n \to \infty. \tag{3.1}$$

We write $\mu_n \xrightarrow{w} \mu$, as $n \to \infty$.

Note that the limit is unique. If ν were another probability measure on $(\mathbb{R}, \mathbb{B})$ satisfying (3.1), then μ and ν would agree on the collection $\{(-\infty, a] \mid a \in C\}$, which is an intersection-stable generator of $\mathbb{B}$. By the uniqueness theorem for probability measures, we get that $\mu = \nu$.

Weak convergence of a sequence of probability measures can also be expressed in terms of pointwise convergence of the associated distribution functions. Let us recall the definition.

Definition 3.2. *Let μ be a probability measure on $(\mathbb{R}, \mathbb{B})$.* The distribution function of μ *is the function $F : \mathbb{R} \to [0, 1]$ given by*

$$F(x) = \mu(-\infty, x], \quad x \in \mathbb{R}.$$

It is easily verified that the distribution function F of a probability measure on $(\mathbb{R}, \mathbb{B})$ has the following properties: (1) F is non-decreasing, (2) F is right-continuous, (3) $\lim_{x\to-\infty} F(x) = 0$ and (4) $\lim_{x\to\infty} F(x) = 1$. It is a deep result that if a function F

satisfies the above conditions (1)-(4), then there exists a unique probability measure μ on $(\mathbb{R}, \mathbb{B})$ such that F is the distribution function of μ. Details concerning the construction can be found in Appendix B. We are now ready to spell out the above mentioned connection to weak convergence of a sequence of probability measures. If F and F_n denote the distribution functions of μ and μ_n, respectively, then the condition that $\mu_n \xrightarrow{w} \mu$, as $n \to \infty$, is equivalent to

$$F_n(a) \to F(a), \quad \text{as } n \to \infty$$

for every continuity point a of F. In establishing this equivalence we have also used the fact that left-continuity of F at a point x is equivalent with $\mu(\{x\}) = 0$. So F is continuous at x if and only if $\mu(\{x\}) = 0$.

We can express weak convergence in terms of integrals of certain functions.

Theorem 3.3. *Let $\mu, (\mu_n)_{n \in \mathbb{N}}$ be probability measures on $(\mathbb{R}, \mathbb{B})$. Then $\mu_n \xrightarrow{w} \mu$, as $n \to \infty$, if and only if*

$$\int f \; d\mu_n \to \int f \; d\mu, \quad \textit{as } n \to \infty, \quad \textit{for all } f \in C_b(\mathbb{R}). \tag{3.2}$$

Proof. Suppose (3.2) holds and let a be a continuity point of μ. Define for each $k \in \mathbb{N}$, the functions g_k and f_k by

$$g_k(x) = \begin{cases} 1 & \text{if } x \in \left(-\infty, a - \frac{1}{k}\right) \\ k(x-a) & \text{if } x \in \left[a - \frac{1}{k}, a\right] \\ 0 & \text{else} \end{cases}, \qquad f_k(x) = \begin{cases} 1 & \text{if } x \in (-\infty, a) \\ k\left(a + \frac{1}{k} - x\right) & \text{if } x \in \left[a, a + \frac{1}{k}\right] \\ 0 & \text{else} \end{cases}.$$

Then

$$g_k \leq 1_{(-\infty, a]} \leq f_k. \tag{3.3}$$

Also, note that $(g_k - f_k)(x) \to 0$, as $k \to \infty$, everywhere except for $x = a$, but since a is a continuity point of μ, it follows that $g_k - f_k \to 0$, as $k \to \infty$, μ-almost surely. This sequence is bounded by the constant 1, hence by the Lebesgue dominated convergence theorem, $\int (g_k - f_k) \; d\mu \to 0$, as $k \to \infty$. So, given $\epsilon > 0$, we can then choose k large enough so that

$$\left| \int g_k \; d\mu - \int f_k \; d\mu \right| < \epsilon. \tag{3.4}$$

Let k be fixed, such that (3.4) holds. We also have for all $n \in \mathbb{N}$ (by using (3.3)) that

$$\int g_k \; d\mu_n \leq \mu_n(-\infty, a] \leq \int f_k \; d\mu_n. \tag{3.5}$$

And similarly we have that

$$\int g_k \; d\mu \leq \mu(-\infty, a] \leq \int f_k \; d\mu. \tag{3.6}$$

Since $g_k, f_k \in C_b(\mathbb{R})$, we know by hypothesis that $\int g_k\ d\mu_n \to \int g_k\ d\mu$, as $n \to \infty$, and $\int f_k\ d\mu_n \to \int f_k\ d\mu$, as $n \to \infty$. So choose $N_1, N_2 \in \mathbb{N}$ such that

$$\begin{aligned}
\left| \int g_k\ d\mu_n - \int g_k\ d\mu \right| &< \epsilon \quad \text{for all } n \geq N_1 \\
\left| \int f_k\ d\mu_n - \int f_k\ d\mu \right| &< \epsilon \quad \text{for all } n \geq N_2.
\end{aligned}$$

Set $N = \max\{N_1, N_2\}$. By using using (3.5) and (3.6), and then using (3.4), we deduce that for all $n \geq N$

$$\begin{aligned}
|\mu_n(-\infty, a] - \mu(-\infty, a]| &\leq \left| \int f_k\ d\mu_n - \int g_k\ d\mu \right| + \left| \int f_k\ d\mu - \int g_k\ d\mu_n \right| \\
&\leq \left| \int f_k\ d\mu_n - \int f_k\ d\mu \right| + \left| \int g_k\ d\mu_n - \int g_k\ d\mu \right| + \epsilon + \epsilon < 4\epsilon.
\end{aligned}$$

Since ϵ was arbitrary, we conclude that $\mu_n(-\infty, a] \to \mu(-\infty, a]$, as $n \to \infty$. This proves that $\mu_n \xrightarrow{w} \mu$, as $n \to \infty$.

For the converse implication, let C_μ denote the set of continuity points of μ. Pick $f \in C_b(\mathbb{R})$, let $M > 0$ denote the bound and let $\epsilon > 0$. We have that (3.2) holds for indicator functions $1_{(-\infty,a]}$ for all $a \in C_\mu$. By additivity of integrals, (3.2) also holds for indicator functions $1_{(a,b]}$ where $a, b \in C_\mu$ and $a < b$. By linearity, (3.2) also holds for all finite sums of such indicator functions, which we call simple functions. By the approximation Theorem A.7 in Appendix A we can find a sequence of such simple functions $(s_k)_{k \in \mathbb{N}}$ such that

$$\sup_{x \in (-k,k]} |f(x) - s_k(x)| < \epsilon. \tag{3.7}$$

Now choose k large enough so that $1 - \mu(-k, k] < \epsilon$. Furthermore, choose N_1 large enough to ensure that for all $n \geq N_1$

$$\left| \int s_k\ d\mu_n - \int s_k\ d\mu \right| < \epsilon. \tag{3.8}$$

By hypothesis

$$\mu_n(-k, k] = \mu_n(-\infty, k] - \mu_n(-\infty, -k] \to \mu(-\infty, k] - \mu(-\infty, -k] = \mu(-k, k],$$

as $n \to \infty$. So choose N_2 large enough to ensure that for all $n \geq N_2$

$$|\mu_n(-k, k] - \mu(-k, k]| < \epsilon. \tag{3.9}$$

Then combining (3.8), (3.7) and (3.9), we obtain that for all $n \geq \max\{N_1, N_2\}$

$$\begin{aligned}
\left|\int f\, d\mu_n - \int f\, d\mu\right| &\leq \epsilon + \left|\int f\, d\mu_n - \int f\, d\mu - \int s_k\, d\mu_n + \int s_k\, d\mu\right| \\
&\leq \epsilon + \int |f - s_k|\, d\mu + \int |f - s_k|\, d\mu_n \\
&= \epsilon + \int_{(-k,k]} |f - s_k|\, d\mu + \int_{(-k,k]^c} |f - s_k|\; d\mu + \int_{(-k,k]} |f - s_k|\, d\mu_n \\
&\quad + \int_{(-k,k]^c} |f - s_k|\, d\mu_n \\
&\leq 3\epsilon + M\mu_n((-k,k]^c) + M\mu((-k,k]^c) \\
&\leq 3\epsilon + M\left(2 - \mu(-k,k] + \epsilon - \mu(-k,k]\right) \leq 3\epsilon + 3M\epsilon = 3\epsilon(M+1).
\end{aligned}$$

Since $\epsilon > 0$ was arbitrary, we conclude that (3.2) holds. □

3.2 Convergence of Random Variables

We will focus on a few types of convergence of random variables in this section.

Definition 3.4. *Let $X, (X_n)_{n\in\mathbb{N}}$ be real-valued random variables. We say that X_n converge in distribution to X, as $n \to \infty$, and write $X_n \xrightarrow{\mathcal{D}} X$, as $n \to \infty$, if the laws of X_n converge weakly to the law of X, i.e., $X_n(P_n) \xrightarrow{w} X(P)$, as $n \to \infty$.*

Note that we do not require the random variables to be defined on the same probability space. We illustrate this idea by an example, which follows section 25 in [2]:

Example 3.5. Let $\lambda > 0$. For each $n \in \mathbb{N}$ with $n > \lambda$, let Ω_n denote the space of n-tuples of 0's and 1's. Consider the σ-algebras $\mathbb{F}_n = \mathbb{P}(\Omega_n)$, the power set of Ω_n, and for $\omega \in \Omega$ let the probability measures P_n assign probability

$$P_n(\omega) = \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

Where k denotes the number of 1's in ω. Let X_n be the random variable on $(\Omega_n, \mathbb{F}_n, P_n)$, which represents the number of successes in n Bernoulli trials, with probability λ/n as success parameter. The probability mass is concentrated on a countable set. So if we know how the point masses behave in the limit, we know how the entire distribution behaves, hence we know the weak convergence properties of $(X_n)_{n\in\mathbb{N},\lambda<n}$. We see that

$$\begin{aligned}
X_n(P_n)(\{k\}) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n!}{k!(n-k)!}\frac{\lambda^k}{n^k}\left(1 - \frac{\lambda}{n}\right)^n \frac{1}{\left(1 - \frac{\lambda}{n}\right)^k} \\
&= \frac{\lambda^k \left(1 - \frac{\lambda}{n}\right)^n}{k!} \frac{1}{\left(1 - \frac{\lambda}{n}\right)^k} \prod_{i=0}^{k-1}\left(1 - \frac{i}{n}\right) \to \frac{\lambda^k e^{-\lambda}}{k!}, \quad \text{as } n \to \infty.
\end{aligned}$$

Hence X_n converge in distribution to the Poisson distribution.

We also want to point out why it is too restrictive to require (3.1) to hold for points of discontinuity. The basic argument is that it allows us to bring the weak law of large numbers under the theory of weak convergence. It is illustrated in the next example.

Example 3.6. Let $X_1, X_2, ...$ be independent random variables where $P(X_k = 1) = 1/2 = P(X_k = -1)$. For $n \in \mathbb{N}$, set $S_n = \sum_{k=1}^n X_k$. Then the weak law of large numbers asserts that for every $\epsilon > 0$,

$$P\left(\left|\frac{1}{n}S_n\right| > \epsilon\right) \to 0, \quad \text{as } n \to \infty. \tag{3.10}$$

If F_n is the distribution function of $n^{-1}S_n$ and μ_n denotes its law, then (3.10) gives us that

$$F_n(x) \;=\; 1 - P\left(\frac{1}{n}S_n > x\right) \to 1, \quad \text{as } n \to \infty, \quad \text{if } x > 0 \tag{3.11}$$

$$F_n(x) \;\leq\; P\left(\left|\frac{1}{n}S_n\right| \geq |x|\right) \to 0, \quad \text{as } n \to \infty, \quad \text{if } x < 0. \tag{3.12}$$

So if we let ε_0 denote the one-point measure at 0, then $\mu_n \xrightarrow{w} \varepsilon_0$, as $n \to \infty$.

However, we also see that if $\mu_n \xrightarrow{w} \varepsilon_0$, as $n \to \infty$, then (3.11) and (3.12) holds. So (3.10) is equivalent with $\mu_n \xrightarrow{w} \varepsilon_0$, as $n \to \infty$. Now we see why requiring (3.1) to hold in all points $a \in \mathbb{R}$ is too restrictive. Letting n go through the odd numbers, then $S_n = 0$ is impossible. By symmetry we must have $P(S_n \leq 0) = 1/2 = P(S_n \geq 0)$, hence $F_n(0) = 1/2$. Thus $\mu_n(-\infty, 0] = 1/2$ fails at converging to $\varepsilon_0(-\infty, 0] = 1$, as $n \to \infty$.

We introduce two other familiar types of convergence of real valued random variables. The definition requires all of the random variables to be defined on the same probability space, in contrast with the case of convergence in distribution.

Definition 3.7. *Let $X, (X_n)_{n\in\mathbb{N}}$ be real-valued random variables on the same probability space $(\Omega, \mathbb{F}, P)$. We say that X_n converge almost surely (or a.s.) to X, as $n \to \infty$ if*

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1.$$

In this case we write $X_n \xrightarrow{a.s.} X$, as $n \to \infty$.

Definition 3.8. *Let $X, (X_n)_{n\in\mathbb{N}}$ be real-valued random variables on the same probability space $(\Omega, \mathbb{F}, P)$. We say that X_n converge in probability to X, as $n \to \infty$ and write $X_n \xrightarrow{P} X$, as $n \to \infty$, if for every $\epsilon > 0$*

$$P(|X_n - X| > \epsilon) \to 0, \quad \textit{as } n \to \infty.$$

Theorem 3.9. *Let $X, (X_n)_{n\in\mathbb{N}}$ be real-valued random variables on the same probability space $(\Omega, \mathbb{F}, P)$. Then*

1. *If $X_n \xrightarrow{a.s.} X$, as $n \to \infty$, then $X_n \xrightarrow{P} X$, as $n \to \infty$.*
2. *If $X_n \xrightarrow{P} X$, as $n \to \infty$, then $X_n \xrightarrow{\mathcal{D}} X$, as $n \to \infty$.*

Moreover, all implications are strict.

Proof. Assume $X_n \xrightarrow{a.s.} X$, as $n \to \infty$, and let $\epsilon > 0$. Consider the set $\Omega_0 = \{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) \neq X(\omega)\}$. We have that $\Omega_0 \in \mathbb{F}$ and it has P-measure 0. Now,

$$\left(\bigcup_{n=m}^{\infty} \{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| > \epsilon\} \right)_{m\in\mathbb{N}}$$

is a decreasing sequence of $\mathbb{F}$-sets. By downward continuity of P,

$$\lim_{m\to\infty} P\left(\bigcup_{n=m}^{\infty} \{|X_n - X| > \epsilon\} \right) = P\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{|X_n - X| > \epsilon\} \right). \tag{3.13}$$

Note that

$$\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| > \epsilon\} \subseteq \Omega_0,$$

because if ω lies in the left-hand side then $|X_n(\omega) - X(\omega)| > \epsilon$ happens infinitly often, hence $X_n(\omega)$ cannot converge. But then the right-hand side in (3.13) is 0. This proves the first implication. Next we find a counterexample confirming that the implication is strict. Let P be the uniform distribution on $(0, 1]$. Let $X \equiv 0$ and set

$$X_1 = 1_{(0,1/2]} \quad X_2 = 1_{(1/2,1]}$$

$$X_3 = 1_{(0,1/4]} \quad X_4 = 1_{(1/4,1/2]} \quad X_5 = 1_{(1/2,3/4]} \quad X_6 = 1_{(3/4,1]},$$

and so on we divide each interval in the preceeding block in halfs. Then the probability of X_n and X being different will decrease towards 0, as we move along the blocks, so X_n converges in probability to X, as $n \to \infty$. However $X_n(\omega) = 1$ infinitely often for all $\omega \in (0, 1]$, hence X_n fails to converge to X on a set with P-measure 1.

Now assume that X_n converges in probability to X, as $n \to \infty$. Let $\epsilon > 0$ and $f \in C_b(\mathbb{R})$ be given, and let M denote the bound on f. Since f is continuous, there exists $\delta > 0$ such that $|X_n(\omega) - X(\omega)| \leq \delta$ implies $|f(X_n(\omega)) - f(X(\omega))| \leq \epsilon$. Now we see that

$$\begin{aligned}
\left| \int f \, dX_n(P) - \int f \, dX(P) \right| &= \left| \int f(X_n) \, dP - \int f(X) \, dP \right| \leq \int |f(X_n) - f(X)| \, dP \\
&= \int_{|X_n - X| \leq \delta} |f(X_n) - f(X)| \, dP \\
&+ \int_{|X_n - X| > \delta} |f(X_n) - f(X)| \, dP \\
&\leq \epsilon + 2MP(|X_n - X| > \delta) \to \epsilon, \quad \text{as } n \to \infty.
\end{aligned}$$

Since ϵ was arbitrary, we conclude that $X_n \xrightarrow{\mathcal{D}} X$ by Theorem 3.3.

Now let X and Y be independent random variables, each attaining the values 0 and 1 with probability 1/2. Set $X_n = Y$, for all $n \geq 1$. Then clearly X_n converge in distribution to X, as $n \to \infty$. However, because of independence, we have $P(|X - Y| = 1) = 1/2$. So X_n fails to converge in probability to X, thus proving the converse does not hold. $\square$

Remark 3.10. If the limit is a constant P-a.s. then convergence in distribution implies convergence in probability. Let $X = a$ P-a.s. and $\epsilon > 0$. Then

$$\begin{aligned} P\left(|X_n - a| > \epsilon\right) &\leq P\left(X_n \leq a - \epsilon\right) + P\left(X_n \geq a + \epsilon\right) \\ &= P\left(X_n \leq a - \epsilon\right) + 1 - P\left(X_n \leq a + \epsilon\right) \to 0, \quad \text{as } n \to \infty. \end{aligned}$$

The following theorem, due to Skorohod, is very useful in proving some of the following results.

Theorem 3.11 (Skorohod's Theorem)**.** *Let $\mu, (\mu_n)_{n\in\mathbb{N}}$ be probability measures on $(\mathbb{R}, \mathbb{B})$ and assume that $\mu_n \xrightarrow{w} \mu$, as $n \to \infty$. Then there exist real-valued random variables $Y, (Y_n)_{n\in\mathbb{N}}$ defined on the same probability space $(\Omega, \mathbb{F}, P)$ such that $Y, (Y_n)_{n\in\mathbb{N}}$ have distribution $\mu, (\mu_n)_{n\in\mathbb{N}}$, respectively, and $Y_n(\omega) \to Y(\omega)$, as $n \to \infty$ for all $\omega \in \Omega$.*

Proof. Set the background space to $\left((0,1), \mathbb{B}|_{(0,1)}, m|_{\mathbb{B}|_{(0,1)}}\right)$. Let F be the distribution function corresponding to μ. Then define

$$Y(\omega) = \inf\{x \in \mathbb{R} \mid \omega \leq F(x)\}\ , \quad \omega \in (0,1).$$

Consider the set $\{x \in \mathbb{R} \mid \omega \leq F(x)\}$. It is an interval going to ∞, because F is non-decreasing. Furthermore, this interval is closed, due to right-continuity of F. Therefore $\{x \in \mathbb{R} \mid \omega \leq F(x)\} = [Y(\omega), \infty)$. So for $\omega \in (0,1)$, we have that $\omega \leq F(x)$ **if and only if** $Y(\omega) \leq x$. This observation will be used several times in the proof. First, we get that

$$P\left(\omega \in \Omega \mid Y(\omega) \leq x\right) = P\left(\omega \in \Omega \mid \omega \leq F(x)\right) = F(x), \quad x \in \mathbb{R}.$$

So Y has distribution μ. The construction and above argument also hold for Y_n, $n \in \mathbb{N}$.

We only need to show the pointwise convergence. Let $\omega \in (0,1)$. Given $\epsilon > 0$, find $x \in \mathbb{R}$ such that $Y(\omega) - \epsilon < x < Y(\omega)$ and that $\mu(\{x\}) = 0$ (recall that there are at most countably many discontinuity points of μ). So $F(x) < \omega$. Since $F_n(x) \to F(x)$, as $n \to \infty$, this implies $F_n(x) < \omega$ for n large enough. Hence $Y(\omega) - \epsilon < x < Y_n(\omega)$ for n large enough, therefore we have that

$$\liminf_{n\to\infty} Y_n(\omega) \geq Y(\omega). \tag{3.14}$$

Next, if $\omega < \omega'$ choose x' such that $Y(\omega') < x' < Y(\omega') + \epsilon$ and that $\mu(\{x'\}) = 0$. Then we have that $\omega < \omega' \leq F(x')$. So for n large enough, $\omega \leq F_n(x')$ holds. Therefore we have that $Y_n(\omega) \leq x' < Y(\omega') + \epsilon$ for n large enough. Hence for all $\omega' > \omega$

$$\limsup_{n\to\infty} Y_n(\omega) \leq Y(\omega').$$

If ω is a point of continuity for Y, then $\limsup_{n\to\infty} Y_n(\omega) \leq Y(\omega)$. This combined with (3.14) gives us that $Y_n(\omega) \to Y(\omega)$, as $n \to \infty$, for all continuity points ω of Y. Since Y is non-decreasing, there can be at most countably many of these. Changing $Y_n(\omega) = Y(\omega) = 0$ for such points reveals pointwise convergence for all $\omega \in (0,1)$. Changing Y, Y_n on a set with Lebesgue measure 0 does not change the fact that Y and Y_n have distributions μ and μ_n, respectively, and the proof is complete. □

Theorem 3.12. *Let $h : \mathbb{R} \to \mathbb{R}$ be $\mathbb{B}$-$\mathbb{B}$-measurable. If $\mu_n \xrightarrow{w} \mu$, as $n \to \infty$ and $\mu(D_h) = 0$, where D_h is the set of discontinuity points of h, then $h(\mu_n) \xrightarrow{w} h(\mu)$, as $n \to \infty$.*

Proof. First, note that D_h is measurable, so the assumption makes sense. We show this by looking at the complement. If h is continuous at x then it is also continuous on an ϵ-neighbourhood of x, let $B(x, \epsilon_x)$ denote it. So

$$D_h^c = \bigcup_{x \in D_h^c} B(x, \epsilon_x),$$

hence D_h^c is open, hence D_h is measurable.

Let Y, Y_n denote the random variables having distribution μ and μ_n, respectively, according to Skorohod's Theorem. So if $Y(\omega) \notin D_h$ then $h(Y_n(\omega)) \to h(Y(\omega))$, as $n \to \infty$. Since $P(\omega \in \Omega \mid Y(\omega) \in D_h) = \mu(D_h) = 0$, the convergence happens with probability 1. So $h(Y_n)$ converges to $h(Y)$ in probability, as $n \to \infty$. Using Theorem 3.9 reveals that $h(\mu_n) \xrightarrow{w} h(\mu)$, as $n \to \infty$. □

3.3 Tightness and Weak Convergence of Measures

Using *Helly's selection principle* allows us to establish additional results about weak convergence. This will be illustrated in the next theorem.

Theorem 3.13. *For every sequence of distribution functions $(F_n)_{n \in \mathbb{N}}$ there exists a subsequence $(F_{n_k})_{k \in \mathbb{N}}$ and a non-decreasing right-continuous function F such that $F_{n_k}(x) \to F(x)$, as $k \to \infty$, for all continuity points x of F.*

Proof. Let D be a countable dense subset of $\mathbb{R}$. Let $\{r_k \mid k \in \mathbb{N}\}$ be an enumeration of it. Then the sequence of numbers $\{F_n(r_1) \mid n \in \mathbb{N}\}$ is bounded. Hence there exists a subsequence $(F_{1k})_{k \in \mathbb{N}}$ of $(F_n)_{n \in \mathbb{N}}$, that converges at r_1. Let l_1 denote the limit. Clearly $l_1 \in [0, 1]$. Now this new sequence of numbers $\{F_{1k}(r_2) \mid k \in \mathbb{N}\}$ is also bounded. So again there exists a further subsequence $(F_{2k})_{k \in \mathbb{N}}$, that converges at r_2. Let l_2 denote the limit and we have again that $l_2 \in [0, 1]$. In this way, we recursively define subsequences $(F_{jk})_{k \in \mathbb{N}}$ for each $j \in \mathbb{N}$, that converges at r_j. Let l_j denote the limit, we have that $l_j \in [0, 1]$.

Next choose the diagonal elements $(F_{kk})_{k \in \mathbb{N}}$. Pick $r_j \in D$, then from a certain step on $(F_{kk})_{k \in \mathbb{N}}$ is a subsequence of $(F_{jk})_{k \in \mathbb{N}}$, so $(F_{kk})_{k \in \mathbb{N}}$ converges at r_j, as $k \to \infty$. Thus we have found a subsequence $(F_{n_k})_{k \in \mathbb{N}}$, that converges at all points in D. We can then define the following function

$$G(r) = \lim_{k \to \infty} F_{n_k}(r), \quad \text{for all } r \in D.$$

Note that G is non-decreasing, because the F_n's are so. Define

$$F(x) = \inf\{G(r) \mid r \in D,\ x < r\}, \quad \text{for all } x \in \mathbb{R}.$$

By construction, F is non-decreasing. Given $x_0 \in \mathbb{R}$, we show right-continuity of F at this point. Let $\epsilon > 0$. There exists $r_0 \in D$ such that $x_0 < r_0$ and $G(r_0) - \epsilon \leq F(x_0) \leq G(r_0)$. If $r \in D$ and $x_0 < r < r_0$ then

$$0 \leq G(r) - F(x_0) \leq G(r_0) - F(x_0) \leq \epsilon,$$

but then by passing to infimum (and using the definition of F) we obtain

$$0 \le F(x) - F(x_0) \le \epsilon, \qquad \text{for all } x_0 < x < r_0.$$

This shows that F is right-continuous.

We then only need to show that $F_{n_k}(x)$ converges pointwise to $F(x)$ for all continuity points of F, as $k \to \infty$. Let x_0 be a continuity point of F and let $\epsilon > 0$. There exist $r, r', r'' \in D$ such that $r < r' < x_0 < r''$ and by continuity we can also assume that $F(r'') - F(r) \le \epsilon$. Then we have

$$F(r) \le G(r') \le F(x_0) \le G(r'') \le F(r) + \epsilon, \tag{3.15}$$

and, respectively, for all $k \in \mathbb{N}$,

$$F_{n_k}(r') \le F_{n_k}(x_0) \le F_{n_k}(r'').$$

Letting $k \to \infty$ combined with (3.15) gives us that for k large enough

$$|F_{n_k}(x_0) - F(x_0)| \le \epsilon.$$

Hence $F_{n_k}(x)$ converges pointwise to $F(x)$, as $k \to \infty$, for all continuity points x of F. $\square$

In accordance with Theorem B.5, we see that there exists a unique measure μ on $(\mathbb{R}, \mathbb{B})$, such that for all $a, b \in \mathbb{R}$, $a < b$

$$\mu(a, b] = F(b) - F(a) \quad \forall a, b \in \mathbb{R},\ a < b.$$

However, we cannot be sure that μ is a probability measure. We do see from the above proof that $F(x) \le 1$, for all $x \in \mathbb{R}$, but nothing is preventing mass from escaping at infinity. The following example shows that this can, indeed, happen.

Example 3.14. Consider, for each $n \in \mathbb{N}$, the uniform distributions on $(n, n+1)$, μ_n. Then for all $a, b \in \mathbb{R}$, $\mu_n(a, b] \to 0$, as $n \to \infty$. Hence every subsequence of $(\mu_n)_{n\in\mathbb{N}}$ must converge weakly to the 0-measure.

We therefore introduce a new concept which will play an important role for weak convergence.

Definition 3.15. *A sequence of probability measures $(\mu_n)_{n\in\mathbb{N}}$ on $(\mathbb{R}, \mathbb{B})$ is said to be tight, if for each $\epsilon > 0$ there exists a finite interval $(a, b]$ such that for all $n \in \mathbb{N}$*

$$\mu_n(a, b] > 1 - \epsilon.$$

Example 3.16. Let $(x_n)_{n\in\mathbb{N}}$ denote a sequence of real numbers. For each $n \ge 1$, let ε_{x_n} be the corresponding one-point-measure. Then

$$\varepsilon_{x_n}(a, b] = \begin{cases} 1 & \text{if } x_n \in (a, b] \\ 0 & \text{else.} \end{cases}$$

So the family of one-point-measures corresponding to the sequence $(x_n)_{n\in\mathbb{N}}$ is tight if and only if the sequence is bounded.

As it turns out, the concept of tightness is precisely what prevents mass from escaping at infinity.

Theorem 3.17. *Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $(\mathbb{R},\mathbb{B})$. Tightness is a necessary and sufficient condition that for every subsequence $(\mu_{n_k})_{k\in\mathbb{N}}$ there exists a further subsequence $(\mu_{n_{k(j)}})_{j\in\mathbb{N}}$ and a **probability** measure μ such that $\mu_{n_{k(j)}} \xrightarrow{w} \mu$, as $j\to\infty$.*

Proof. Let $(\mu_n)_{n\in\mathbb{N}}$ be tight. For each $n\geq 1$, let F_n denote the distribution function of μ_n. Applying Theorem 3.13 to the sequence of distribution functions reveals a new function F. Where F_{n_k} converges pointwise to F, as $k\to\infty$, in every continuity point of F, for some subsequence $(F_{n_k})_{k\in\mathbb{N}}$ of $(F_n)_{n\in\mathbb{N}}$. Now, by applying Theorem B.5, we can find a measure μ on $(\mathbb{R},\mathbb{B})$, such that $\mu(a,b]=F(b)-F(a)$ for every $a,b\in\mathbb{R}$ and $a<b$. Given $\epsilon>0$, choose $a<b$ according to tightness so that

$$\mu_n(a,b] > 1-\epsilon \quad \forall n\in\mathbb{N}$$

By increasing b and a ensures it holds in continuity points of F. Then we obtain that

$$\mu(a,b] = F(b)-F(a) = \lim_{n\to\infty}(F_n(b)-F_n(a)) = \lim_{n\to\infty}\mu_n(a,b] \geq 1-\epsilon,$$

hence μ is a probability measure (we knew from the proof of Theorem 3.13 that $\mu(\mathbb{R})\leq 1$). Hence F is indeed a distribution function and F_n converges pointwise to F, in all of the continuity points of F, along a subsequence. As discussed in the section after Definition 3.2, this is equivalent with the measures μ_n converging weakly to μ, along a subsequence. Hence the convergence also holds along any further subsequence.

For the other implication, assume that $(\mu_n)_{n\in\mathbb{N}}$ is not tight. Then there exists $\epsilon>0$ so that for every finite interval $(a,b]$, $\mu_n(a,b]\leq 1-\epsilon$ for some n. For each $k\in\mathbb{N}$, choose n_k such that $\mu_{n_k}(-k,k]\leq 1-\epsilon$. Assume that some subsequence $(\mu_{n_{k(j)}})_{j\in\mathbb{N}}$ of $(\mu_{n_k})_{k\in\mathbb{N}}$ converges weakly to some probability measure μ. Let $(a,b]$ be any finite interval, where $a<b$ and $\mu(\{a\})=\mu(\{b\})=0$. For j large enough $(a,b]\subseteq(-k(j),k(j)]$ and so

$$1-\epsilon \geq \mu_{n_{k(j)}}(-k(j),k(j)] \geq \mu_{n_{k(j)}}(a,b] \to \mu(a,b], \quad \text{as } j\to\infty.$$

So $\mu(a,b]\leq 1-\epsilon$ for every finite inteval. Hence μ cannot be a probability measure, which is a contradiction. ☐

Corollary 3.18. *Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $(\mathbb{R},\mathbb{B})$. If $(\mu_n)_{n\in\mathbb{N}}$ is tight and if every weakly convergent subsequence $(\mu_{n_k})_{k\in\mathbb{N}}$ has the same limit μ, then $\mu_n \xrightarrow{w} \mu$, as $n\to\infty$.*

Proof. Note that μ is a probability measure due to the previous theorem. According to the theorem every sequence $(\mu_{n_k})_{k\in\mathbb{N}}$ contains a further subsequence $(\mu_{n_{k(j)}})_{j\in\mathbb{N}}$ converging weakly to some limit. This limit must be μ by hypothesis. So every subsequence $(\mu_{n_k})_{k\in\mathbb{N}}$ contains a further subsequence $(\mu_{n_{k(j)}})_{j\in\mathbb{N}}$ converging weakly to μ.

Now suppose $(\mu_n)_{n\in\mathbb{N}}$ fails to converge weakly to μ, as $n\to\infty$. So for some continuity point a of μ, $\mu_n(-\infty,a]\nrightarrow\mu(-\infty,a]$, as $n\to\infty$. Hence there exists $\epsilon>0$ so that

$$|\mu_{n_k}(-\infty,a]-\mu(-\infty,a]|\geq\epsilon,$$

for an infinite sequence $(\mu_{n_k})_{k\in\mathbb{N}}$. No further subsequence of such can converge weakly to μ, contradicting what we just concluded. $\square$

Tightness also has a strong connection to characteristic functions.

Theorem 3.19. *Let $(\mu_n)_{n\in\mathbb{N}}$ be probability measures on $(\mathbb{R},\mathbb{B})$ and let $(\phi_n)_{n\in\mathbb{N}}$ be the respective characteristic functions. If for all $\theta\in\mathbb{R}$*

$$\phi_n(\theta)\to\phi(\theta),\quad \text{as } n\to\infty,$$

where ϕ is a function that is continuous at $\theta=0$, then $(\mu_n)_{n\in\mathbb{N}}$ is tight.

Proof. Observe that for every $u>0$ we have

$$\begin{aligned}\frac{1}{u}\int_{-u}^{u}1-e^{i\theta x}\,d\theta &= \frac{1}{u}\left[\theta-\frac{1}{ix}e^{i\theta x}\right]_{\theta=-u}^{\theta=u}=\frac{1}{u}\left(u-\frac{1}{ix}e^{iux}+u+\frac{1}{ix}e^{-iux}\right)\\ &= 2-\frac{2}{ux2i}\left(e^{iux}-e^{-iux}\right)=2-\frac{2}{ux}\sin(ux).\end{aligned}$$

Furthermore we also have that

$$\int_{-u}^{u}|1-\phi_n(\theta)|\;d\theta\le\int_{-u}^{u}2\;d\theta=4u.$$

Therefore we can apply Fubini's theorem. Using the definition of the characteristic function gives

$$\begin{aligned}\frac{1}{u}\int_{-u}^{u}1-\phi_n(\theta)\;d\theta &= \int\frac{1}{u}\int_{-u}^{u}1-e^{i\theta x}\;d\theta\;d\mu_n(x)\\ &= 2\int 1-\frac{\sin(ux)}{ux}\;d\mu_n(x)\ge 2\int_{|x|\ge 2/u}1-\frac{1}{|ux|}\;d\mu_n(x)\\ &\ge \mu_n\left\{x\in\mathbb{R}\mid |x|\ge\frac{2}{u}\right\}.\end{aligned}\tag{3.16}$$

The function ϕ is continuous at $\theta=0$ and $\phi(0)=\lim_{n\to\infty}\phi_n(0)=1$, hence it is also continuous in a neighbourhood $(-u_0,u_0)$. So for every $\epsilon>0$ we find a $u>0$ such that $u<u_0$ and

$$\frac{1}{u}\int_{-u}^{u}1-\phi(\theta)\;d\theta<\epsilon.$$

We have that ϕ_n converges to ϕ bounded by 1, the Lebesgue dominated convergence theorem gives us that there exists $n_0\in\mathbb{N}$ such that

$$\frac{1}{u}\int_{-u}^{u}1-\phi_n(\theta)\;d\theta<2\epsilon,\quad\text{for all } n\ge n_0.$$

Setting $a=2/u$ in (3.16) we get that

$$\mu_n\left\{x\in\mathbb{R}\mid |x|\ge a\right\}<2\epsilon,\quad n\ge n_0.$$

Letting a increase insures that it also holds for the first finitely many n. Hence $(\mu_n)_{n\in\mathbb{N}}$ is tight. $\square$

This leads us to a connection between weak convergence and characteristic functions.

Theorem 3.20 (The Continuity Theorem). *Let $(\mu_n)_{n\in\mathbb{N}}$ be probability measures on $(\mathbb{R}, \mathbb{B})$ and let $(\phi_n)_{n\in\mathbb{N}}$ be the respective characteristic functions. Assume that ϕ_n converges pointwise to some function ϕ, as $n \to \infty$. The following are equivalent:*

1. *$(\mu_n)_{n\in\mathbb{N}}$ is tight.*
2. *$(\mu_n)_{n\in\mathbb{N}}$ converges weakly to some probability measure μ.*
3. *ϕ is the characteristic function of some probability measure μ.*
4. *ϕ is continuous.*
5. *ϕ is continuous at $\theta = 0$.*

Where μ in 2. and 3. refers to the same probability measure.

Proof. (2. implies 3.) Assume that $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to some probability measure μ, let ϕ^* denote the characteristic function of μ. For every $\theta \in \mathbb{R}$, the function $x \mapsto e^{i\theta x}$ has bounded modulus and it is continuous. Applying Theorem 3.3 to both imaginary and real part of $e^{i\theta x}$ reveals that $\phi_n(\theta) \to \phi^*(\theta)$, as $n \to \infty$ for every $\theta \in \mathbb{R}$. So we must have that $\phi = \phi^*$. Hence ϕ is the characteristic function of μ.

(1. implies 2.) By Corollary 3.18, we only need to check that every weakly convergent subsequence of $(\mu_n)_{n\in\mathbb{N}}$ has the same limit. Assume that $\mu_{n_k} \xrightarrow{w} \nu$, as $k \to \infty$ and let ϕ^* denote the characteristic function of ν. By the proof of the preceeding implication and since $(\mu_{n_k})_{k\in\mathbb{N}}$ is a subsequence of $(\mu_n)_{n\in\mathbb{N}}$, we must have that $\phi = \phi^*$. So the limit of every weakly convergent subsequence of $(\mu_n)_{n\in\mathbb{N}}$ has the same characteristic function, hence by Theorem 1.10, we conclude that the limits are the same. Corollary 3.18 gives us, that $\mu_n \xrightarrow{w} \mu$, as $n \to \infty$.

(3. implies 4.) By Lemma 1.12 every characteristic function is continuous.

(4. implies 5.) Trivially true.

(5. implies 1.) Follows simply by Theorem 3.19. □

With this result at hand, we can prove the following theorem, which is a version of the Central Limit Theorem. It illustrates how useful the Continuity theorem is.

Theorem 3.21. *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of independent, identically distributed real-valued random variables. Assume that the second moment exists and let $E(X_i) = \mu$ and let the variance be $V(X_i) = E((X_i - E(X_i))^2) = \sigma^2$. Set, for each $n \geq 1$, $S_n = \sum_{i=1}^n X_i$. Then*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} X, \quad \text{as } n \to \infty,$$

where X is standard normal distributed.

Proof. Set $Y_i = (X_i - \mu)/\sigma$, $i \geq 1$. Then $E(Y_i) = 0$, $V(Y_i) = 1$ and the sequence $(Y_i)_{i\in\mathbb{N}}$ is independent and identically distributed. Also note that

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^n Y_i}{\sqrt{n}}.$$

So if we prove the claim for $\mu = 0$ and $\sigma^2 = 1$, then we have also proven it in general.

We calculate the characteristic function ϕ of X,

$$\begin{aligned}\phi(\theta) = \int e^{i\theta X}\, dP &= \int e^{i\theta x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dm(x) = \int \frac{1}{\sqrt{2\pi}} e^{-\frac{-2i\theta x + x^2}{2}}\, dm(x) \\ &= \int \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-i\theta)^2}{2}} e^{\frac{(i\theta)^2}{2}}\, dm(x) = e^{-\frac{\theta^2}{2}} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dm(y) = e^{-\frac{\theta^2}{2}}.\end{aligned}$$

In view of the Continuity Theorem, we only need to prove that

$$\phi_{\frac{S_n}{\sqrt{n}}}(\theta) \to e^{-\frac{\theta^2}{2}}, \quad \text{as } n \to \infty,$$

for all $\theta \in \mathbb{R}$. We write out, using Lemma 1.8, Lemma 1.9 and Theorem 1.17

$$\phi_{\frac{S_n}{\sqrt{n}}}(\theta) = \phi_{S_n}\left(\frac{\theta}{\sqrt{n}}\right) = \phi_{X_1}\left(\frac{\theta}{\sqrt{n}}\right)^n = \left(1 - \frac{\theta^2}{2n} + o\left(\frac{\theta^2}{n}\right)\right)^n.$$

We want to show that the last expression converges to $e^{-\theta^2/2}$, as $n \to \infty$. Let $\theta \in \mathbb{R}$ be fixed and let $\epsilon > 0$. Then we can find $N \in \mathbb{N}$, such that for all $n \geq N$ we have that $\left|o(\theta^2/n)\right| / (\theta^2/n) < \epsilon/(2\theta^2)$. Hence for all $n \geq N$ we have that $2n\left|o(\theta^2/n)\right| < \epsilon$. So we have, for n large enough, that

$$\left(1 - \frac{\theta^2 + \epsilon}{2n}\right)^n \leq \left(1 - \frac{\theta^2}{2n} + o\left(\frac{\theta^2}{n}\right)\right)^n \leq \left(1 - \frac{\theta^2 - \epsilon}{2n}\right)^n.$$

From the above we can then obtain that

$$\limsup_{n\to\infty} \left(1 - \frac{\theta^2}{2n} + o\left(\frac{\theta^2}{n}\right)\right)^n \leq e^{-\frac{\theta^2-\epsilon}{2}}, \quad \text{and} \quad \liminf_{n\to\infty} \left(1 - \frac{\theta^2}{2n} + o\left(\frac{\theta^2}{n}\right)\right)^n \geq e^{-\frac{\theta^2+\epsilon}{2}}.$$

Since ϵ was arbitrary, we conclude that

$$\limsup_{n\to\infty} \left(1 - \frac{\theta^2}{2n} + o\left(\frac{\theta^2}{n}\right)\right)^n \leq e^{-\frac{\theta^2}{2}}, \quad \text{and} \quad \liminf_{n\to\infty} \left(1 - \frac{\theta^2}{2n} + o\left(\frac{\theta^2}{n}\right)\right)^n \geq e^{-\frac{\theta^2}{2}}.$$

Hence

$$\phi_{\frac{S_n}{\sqrt{n}}}(\theta) = \left(1 - \frac{\theta^2}{2n} + o\left(\frac{\theta^2}{n}\right)\right)^n \to \phi(\theta), \quad \text{as } n \to \infty.$$

By the Continuity Theorem, the proof is complete. □

As an additional consequence of the Continuity theorem, we have an approach to determine if a function is a characteristic function of some probability measure. Next we will prove a result that gives necessary and sufficient conditions for a function to be a characteristic function of some probability measure. But first we will need a definition and a few lemmas.

Definition 3.22. *A function $f : \mathbb{R} \to \mathbb{C}$ is positive definite if for any finite set of numbers z_j, t_j, where $1 \leq j \leq n$ and $t_j \in \mathbb{R}$ and $z_j \in \mathbb{C}$, we have that*

$$\sum_{j=1}^{n}\sum_{k=1}^{n} f(t_j - t_k) z_j \overline{z_k} \geq 0. \tag{3.17}$$

We have the following properties of such functions:

Lemma 3.23. *If $f : \mathbb{R} \to \mathbb{C}$ is positive definite, then for each $t \in \mathbb{R}$*

$$f(-t) = \overline{f(t)}, \qquad |f(t)| \leq f(0).$$

If f is continuous at $t = 0$, then f is uniformly continuous. If so, then for every $g : \mathbb{R} \to \mathbb{C}$ continuous we have that

$$\int_0^T \int_0^T f(s-t)g(s)\overline{g(t)}\, dsdt \geq 0, \quad \textit{for all } T > 0. \tag{3.18}$$

Proof. Setting $n = 1$, $t_1 = 0$ and $z_1 = 1$ at (3.17) reveals that $f(0) \geq 0$. Then we set $n = 2$, $t_1 = 0$, $t_2 = t$ and $z_1 = 1 = z_2$ and get that $2f(0) + f(t) + f(-t) \geq 0$, hence $f(t) + f(-t)$ is real (since $f(0)$ was). Now we set $n = 2$, $t_1 = 0$, $t_2 = t$, $z_1 = 1$ and $z_2 = i$, and so we get $2f(0) - if(-t) + if(t) \geq 0$, hence $f(t) - f(-t)$ is purely imaginary. We therefore conclude that $f(-t) = \overline{f(t)}$. Let $t_1 = 0$, $t_2 = t$, $z_1 = f(t)$, $z_2 = -|f(t)|$, then we get that

$$f(0)|f(t)|^2 + f(0)|f(t)|^2 + f(t)(-|f(t)|)\overline{f(t)} + f(-t)f(t)(-|f(t)|) = 2f(0)|f(t)|^2 - 2|f(t)|^3,$$

using what we just proved. The left-hand side is non-negative, hence we conclude that $f(0) \geq |f(t)|$. We see that if $f(0) = 0$, then f is constantly 0, in this case (3.18) holds. If $f(0) \neq 0$ then we can assume without loss of generality that $f(0) = 1$, just replace $f(t)$ with $f(t)/f(0)$. Note that (3.17) can be written as a matrix product, so we see that f is positive definite if and only if the matrix $A = (f(t_j - t_k))_{j,k=1,2,\ldots,n}$ is positive semi-definite for all choices of $n \in \mathbb{N}$ and $t_j \in \mathbb{R}$. We know that positive semi-definite matrices have non-negative determinants. Hence setting $t_1 = 0$, $t_2 = t$ and $t_3 = t + h$ we have that

$$\begin{vmatrix} f(0) & f(t) & f(t+h) \\ f(-t) & f(0) & f(h) \\ f(-t-h) & f(-h) & f(0) \end{vmatrix} = 1 - |f(t)|^2 - |f(h)|^2 - |f(t+h)|^2 + 2\mathrm{Re}(f(t)f(h)\overline{f(t+h)}).$$

Where the determinant is non-negative. So we see that

$$\begin{aligned} |f(t) - f(t+h)|^2 &= |f(t)|^2 + |f(t+h)|^2 - 2\mathrm{Re}(f(t)\overline{f(t+h)}) \\ &\leq 1 - |f(h)|^2 + 2\mathrm{Re}(f(t)f(h)\overline{f(t+h)}) - 2\mathrm{Re}(f(t)\overline{f(t+h)}) \\ &= 1 - |f(h)|^2 + 2\mathrm{Re}(f(t)f(h)\overline{f(t+h)}(f(h) - 1)) \\ &\leq 1 - |f(h)|^2 + 2|1 - f(h)| \to 0, \quad \text{as } h \to 0. \end{aligned}$$

Hence f is uniformly continuous, when f is continuous at $t = 0$. Therefore we see that the integral $\int_0^T \int_0^T f(s-t)g(s)\overline{g(t)}\, dsdt$ is well-defined, both as a double Riemann-integral and as a double Lebesgue integral. It is therefore equal to the limit of Riemann sums. Each sum is non-negative due to (3.17), hence the limit is also non-negative. We conclude that $\int_0^T \int_0^T f(s-t)g(s)\overline{g(t)}\, dsdt \geq 0$, for each $T > 0$. $\square$

Lemma 3.24. *Let $h : [0, \infty) \to \mathbb{R}$ be a measurable function, which is bounded from below, and assume that $h(\beta)$ converges, as $\beta \to \infty$ (possibly to ∞). If*

$$\frac{1}{\alpha} \int_0^\alpha h(\beta)\ d\beta \to c, \quad \text{as } \alpha \to \infty, \qquad \text{for some } c \in \mathbb{R},$$

then $\lim_{\beta\to\infty} h(\beta) = c$.

Proof. Assume that $\lim_{\beta\to\infty} h(\beta) \neq c$. If the limit is larger, then we can find $\epsilon > 0$ and $K > 0$ such that $h(\beta) \geq c + \epsilon$ for all $\beta \geq K$. Hence we have that

$$\frac{1}{\alpha} \int_0^\alpha h(\beta)\ d\beta = \frac{1}{\alpha} \int_0^K h(\beta)\ d\beta + \frac{1}{\alpha} \int_K^\alpha h(\beta)\ d\beta \geq \frac{K}{\alpha} \inf_{\beta\in[0,K]} h(\beta) + \frac{\alpha - K}{\alpha}(c + \epsilon)$$

This converges to $(c + \epsilon)$, as $\alpha \to \infty$, contradicting the assumption $\lim_{\beta\to\infty} h(\beta) \neq c$. If the limit is assumed to be smaller than c, the argument is completely analogous. □

Theorem 3.25 (Bochner)**.** *Let $f : \mathbb{R} \to \mathbb{C}$ be a function. f is a characteristic function of some probability measure if and only if f is positive definite, $f(0) = 1$ and f is continuous at $t = 0$.*

Proof. Let f be the characteristic function of the probability measure μ. Then $f(0) = 1$ is clear, and f is continuous due to Lemma 1.12. We see that:

$$\begin{aligned}
\sum_{j=1}^n \sum_{k=1}^n f(t_j - t_k) z_j \bar{z}_k &= \sum_{j=1}^n \sum_{k=1}^n \int e^{i(t_j - t_k)x}\ d\mu(x) z_j \overline{z_k} = \int \sum_{j=1}^n \sum_{k=1}^n e^{it_jx} z_j \overline{e^{it_kx} z_k}\ d\mu(x) \\
&= \int \left| \sum_{j=1}^n e^{it_jx} z_j \right|^2\ d\mu(x) \geq 0
\end{aligned}$$

Hence f is positive definite.

Conversely, let f have the above properties. By the Continuity theorem, we only need to construct a sequence of probability measures, such that the corresponding characteristic functions converge to f. Set for each $T > 0$

$$f_T(t) = \begin{cases} \left(1 - \frac{|t|}{T}\right) f(t) & \text{if } t \in [-T, T] \\ 0 & \text{else.} \end{cases}$$

$$p_T(x) = \frac{1}{2\pi} \int_{-T}^T \left(1 - \frac{|y|}{T}\right) f(y) e^{-iyx}\ dy, \quad x \in \mathbb{R}.$$

Then $f_T \to f$ pointwise, as $T \to \infty$. So we only need to verify, that for each $T > 0$, p_T is a probability density with respect to the Lebesgue measure and it has f_T as its characteristic function. We quickly realize that p_T is measurable for all $T > 0$. Also p_T only attains real-values, since the integrand is conjugate symmetric (since f is due to Lemma 3.23), so the integral over the interval $[-T, T]$ is real.

By use of Lemma 3.23 with the function $g(t) = e^{-itx}$, we obtain

$$\begin{aligned} 0 &\leq \frac{1}{2\pi T}\int_0^T\int_0^T f(s-t)e^{-i(s-t)x}\ dsdt = \frac{1}{2\pi T}\int f(s-t)e^{-i(s-t)x}\ d(1_{[0,T]^2}\cdot m_2)(s,t) \\ &= \frac{1}{2\pi T}\int f(y)e^{-iyx}\ dh(1_{[0,T]^2}\cdot m_2)(y) = \frac{1}{2\pi T}\int_{-T}^T (T-|y|)f(y)e^{-iyx}\ dy = p_T(x). \end{aligned}$$

Here we have used Fubini's Theorem (we integrate over bounded intervals, so the integrability conditions are met). Then we used the Change of variable theorem, with the transformation $h(s,t) = s-t$. We considered the image measure $h(\nu)$, where $\nu = 1_{[0,T]^2}\cdot m_2$. It can be verified that $h(\nu)$ has the density $y \mapsto (T-|y|)1_{[-T,T]}(y)$ with respect to the Lebesgue measure m.

So we have now concluded that p_T is non-negative. Next we verify that p_T integrates to 1 for all $T > 0$. Consider, for $\alpha > 0$ and $y \neq 0$, the integral

$$\frac{1}{\alpha}\int_0^\alpha\int_{-\beta}^\beta e^{iyx}\ dxd\beta = \frac{1}{\alpha}\int_0^\alpha \frac{2\sin(\beta y)}{y}\ d\beta = \frac{2(1-\cos(\alpha y))}{\alpha y^2}. \tag{3.19}$$

And for $y = 0$

$$\frac{1}{\alpha}\int_0^\alpha\int_{-\beta}^\beta 1\ dxd\beta = \alpha = \lim_{y\to 0}\frac{2(1-\cos(\alpha y))}{\alpha y^2}.$$

So for $\alpha > 0$ the following holds

$$\begin{aligned} \frac{1}{\alpha}\int_0^\alpha\int_{-\beta}^\beta p_T(x)\ dxd\beta &= \frac{1}{\alpha}\int_0^\alpha\int_{-\beta}^\beta \frac{1}{2\pi}\int_{-T}^T\left(1-\frac{|y|}{T}\right)f(y)e^{-iyx}\ dydxd\beta \\ &= \frac{1}{\pi}\int_{-T}^T\frac{1}{2}\left(1-\frac{|y|}{T}\right)f(y)\frac{1}{\alpha}\int_0^\alpha\int_{-\beta}^\beta e^{-iyx}\ dxd\beta dy \\ &= \frac{1}{\pi}\int_{-T}^T\left(1-\frac{|y|}{T}\right)f(y)\frac{1-\cos(\alpha y)}{\alpha y^2}\ dy \\ &= \frac{1}{\pi}\int f_T(y)\frac{1-\cos(\alpha y)}{\alpha y^2}\ dm(y) \\ &= \frac{1}{\pi}\int f_T\left(\frac{t}{\alpha}\right)\frac{1-\cos(t)}{t^2}\ dm(t). \end{aligned} \tag{3.20}$$

Where we used Fubini's theorem and the transformation $t = \alpha y$. Observe that the last integrand is bounded by the map $t \mapsto (1-\cos(t))/t^2$ (recall that Lemma 3.23 ensures that $|f(t)| \leq f(0) = 1$). The map is integrable and integrates to π, hence we can apply the Lebesgue dominated convergence theorem and obtain

$$\begin{aligned} \lim_{\alpha\to\infty}\frac{1}{\alpha}\int_0^\alpha\int_{-\beta}^\beta p_T(x)\ dxd\beta &= \lim_{\alpha\to\infty}\frac{1}{\pi}\int f_T\left(\frac{t}{\alpha}\right)\frac{1-\cos(t)}{t^2}\ dy \\ &= \frac{1}{\pi}\int f_T(0)\frac{1-\cos(t)}{t^2}\ dy = 1 \end{aligned}$$

Since f (so f_T as well) is continuous at $t = 0$. Consider the function $h(\beta) = \int_{-\beta}^\beta p_T(x)\ dx$, it is increasing (since $p_T(x) \geq 0$), hence it converges, as $\beta \to \infty$. By Lemma 3.24 we conclude that $h(\beta) \to 1$, as $\beta \to \infty$. So p_T is a probability density with respect to the Lebesgue measure, for each $T > 0$.

Let $\theta \in \mathbb{R}$. By substitution in (3.19), we get that

$$\frac{1}{\alpha} \int_0^\alpha \int_{-\beta}^\beta e^{i\theta x} e^{iyx} \, dx d\beta = \frac{2\sin(\alpha(\theta - y))}{\alpha(\theta - y)^2}.$$

Hence by substitution in (3.20) we get that, for $\alpha > 0$

$$\begin{aligned} \frac{1}{\alpha} \int_0^\alpha \int_{-\beta}^\beta e^{i\theta x} p_T(x) \, dx d\beta &= \frac{1}{\pi} \int f_T(y) \frac{1 - \cos(\alpha(\theta - y))}{\alpha(\theta - y)^2} \, dm(y) \\ &= \frac{1}{\pi} \int f_T\left(\theta - \frac{t}{\alpha}\right) \frac{1 - \cos(t)}{t^2} \, dm(t) \to f_T(\theta), \quad \text{as } \alpha \to \infty. \end{aligned}$$

Where we used the transformation $t = \alpha(\theta - y)$, the Lebesgue dominated convergence theorem and that f (so also f_T) is continuous at $t = \theta$ by Lemma 3.23. Now consider the function $h(\beta) = \int_{-\beta}^\beta e^{i\theta x} p_T(x) \, dx$. Due to conjugate symmetry of the integrand, it is real-valued. The integrand is bounded by the integrable function p_T, hence by the Lebesgue dominated convergence theorem $h(\beta)$ converges to the characteristic function of $p_T \cdot m$, as $\beta \to \infty$. By Lemma 3.24, $h(\beta)$ also converges to $f_T(\theta)$. So now we conclude that, for each $T > 0$, $\mu_T = p_T \cdot m$ is a probability measure, and that f_T is the corresponding characteristic function. By the Continuity theorem, we conclude that f is the characteristic function of some probability measure. □

3.4 Convergence of Moments

In this section we will establish a method of deducing weak convergence from convergence of moments. First we consider yet another type of convergence of random variables:

Definition 3.26. *Let $p > 0$ and $X, (X_n)_{n \in \mathbb{N}}$ be real-valued random variables defined on a common probability space $(\Omega, \mathbb{F}, P)$. Then X_n converge in L^p to X if*

$$E\,|X_n - X|^p \to 0, \quad \textit{as } n \to \infty.$$

It is written $X_n \xrightarrow{L^p} X$, as $n \to \infty$.

Theorem 3.27. *Let $p > 0$ and $X, (X_n)_{n \in \mathbb{N}}$ be real-valued random variables defined on a common probability space $(\Omega, \mathbb{F}, P)$. If $X_n \xrightarrow{L^p} X$, as $n \to \infty$, then $X_n \xrightarrow{P} X$, as $n \to \infty$.*

Proof. It follows from Chebyshev's inequality, that for every $\epsilon > 0$

$$P\left(|X_n - X| > \epsilon\right) \le \frac{E\,|X_n - X|^p}{\epsilon^p} \to 0, \quad \text{as } n \to \infty.$$

Rearranging this yields the result. □

Remark 3.28. The converse is not necessarily true. For each n let X_n attain two values 0 and n and let

$$P(X_n = 0) = 1 - \frac{1}{n^p}\ , \quad P(X_n = n) = \frac{1}{n^p}.$$

Then clealy $X_n \xrightarrow{P} 0$, as $n \to \infty$, but

$$E\,|X_n - 0|^p = \int |X_n|^p \ dP = n^p \frac{1}{n^p} = 1 \nrightarrow 0, \quad \text{as } n \to \infty.$$

So X_n fails to converge to 0 in L^p, as $n \to \infty$.

So we have checked for general relations between all the types of convergence, except between L^p-convergence and almost surely convergence. As we will see next, none of them implies the other.

Let X_n be uniformly distributed on $(n, n+1)$, then

$$X_n(\omega) \to 0, \quad \text{as } n \to \infty, \quad \text{for all } \omega \in \Omega.$$

Hence $X_n \xrightarrow{a.s.} 0$, as $n \to \infty$. But

$$E\,|X_n - 0|^p = \int |X_n|^p \ dP \geq \int_{X_n \in (n,n+1)} |X_n|^p \ dP \geq n^p \to \infty, \quad \text{as } n \to \infty.$$

So X_n fails to converge to 0 in L^p for every $p > 0$.

For proving that the implication also does not hold the other way, consider the counter example in Theorem 3.9. Here X_n where indicator functions on smaller and smaller blocks of $(0, 1]$. We saw that X_n failed to converge to 0 almost surely, but

$$E\,|X_n - 0|^p = \int |X_n|^p \ dP = \int_{\{X_n = 1\}} |X_n|^p \ dP \to 0, \quad \text{as } n \to \infty.$$

So $X_n \xrightarrow{L^p} 0$, as $n \to \infty$, for every $p > 0$

Consider a sequence of random variables, which converges in distribution to some limit random variable. We want to consider if the moments of the limit random variable exist.

Lemma 3.29. *Let $X, (X_n)_{n\in\mathbb{N}}$ be real-valued random variables, not necessarily defined on the same probability space. If $X_n \xrightarrow{\mathcal{D}} X$, as $n \to \infty$, then*

$$E\,|X| \leq \liminf_{n\to\infty} E_n\,|X_n|\,.$$

Proof. Because the random variables are not necessarily defined on the same probability space, the means are dependent of the probability space. This is emphazised by using the index. However, using Skorohod's Theorem makes it possible to construct random variables $Y, (Y_n)_{n\in\mathbb{N}}$ such that X and Y have the same distribution, and similarly for X_n and Y_n. These new variables are defined on the same background space, therefore using $Y = \lim_{n\to\infty} Y_n$ almost surely combined with Fatou's Lemma gives us

$$E\,|X| = E\,|Y| = E\left|\liminf_{n\to\infty} Y_n\right| \leq \liminf_{n\to\infty} E\,|Y_n| = \liminf_{n\to\infty} E\,|X_n|\,.$$

This completes the proof. $\square$

The following concept will be used in order to establish connection with convergence of moments and convergence in L^p.

Definition 3.30. *A sequence of random variables $(X_n)_{n\in\mathbb{N}}$, not necessarily defined on a common probability space, is uniformly integrable if*

$$\lim_{\alpha\to\infty} \sup_{n\in\mathbb{N}} \int_{|X_n|\geq\alpha} |X_n| \ dP_n = 0. \tag{3.21}$$

Theorem 3.31. *Let $X, (X_n)_{n\in\mathbb{N}}$ be real-valued random variables defined on a common background space. If $X_n \xrightarrow{a.s.} X$, as $n\to\infty$, then the following two statements hold:*

1. *If $(X_n)_{n\in\mathbb{N}}$ is uniformly integrable, then $(X_n)_{n\in\mathbb{N}}$ as well as X have first moment and $E(X_n)\to E(X)$, as $n\to\infty$.*
2. *If $X, (X_n)_{n\in\mathbb{N}}$ are non-negative and have first moment and $E(X_n)\to E(X)$, as $n\to\infty$, then $(X_n)_{n\in\mathbb{N}}$ is uniformly integrable.*

Proof. (1) If (3.21) holds, then let α be large enough so the supremum is less than 1. This gives us that

$$E|X_n| = \int |X_n| \ dP = \int_{|X_n|<\alpha} |X_n| \ dP + \int_{|X_n|\geq\alpha} |X_n| \ dP \leq \alpha+1.$$

So the moments exists for $(X_n)_{n\in\mathbb{N}}$, as well for X by applying Lemma 3.29. Now define a new set of random variables for each $\alpha\in\mathbb{R}$

$$X_n^\alpha = 1_{\{|X_n|<\alpha\}}X_n \ , \quad X^\alpha = 1_{\{|X|<\alpha\}}X.$$

The set of discontinuity points of X is at most countable, so pick a sequence of continuity points $(\alpha_k)_{k\in\mathbb{N}}$ which goes to ∞, as $k\to\infty$. Since $P(|X|=\alpha_k)=0$, then $X_n^{\alpha_k} \xrightarrow{a.s.} X^{\alpha_k}$, as $n\to\infty$, bounded by α_k. Hence by the Lebesgue dominated convergence theorem

$$E(X_n^{\alpha_k}) = \int X_n^{\alpha_k} \ dP \to \int X^{\alpha_k} \ dP = E(X^{\alpha_k}), \quad \text{as } n\to\infty.$$

So let $\epsilon>0$, then choose n large enough, so that $|E(X_n^{\alpha_k}) - E(X^{\alpha_k})| < \epsilon$. Then we have for n large enough

$$\begin{aligned}
|E(X_n)-E(X)| &\leq |E(X_n)-E(X_n^{\alpha_k})| + |E(X^{\alpha_k})-E(X)| + \epsilon \\
&= \left|\int X_n - X_n^{\alpha_k} \ dP\right| + \left|\int X^{\alpha_k} - X \ dP\right| + \epsilon \\
&= \left|\int_{|X_n|\geq\alpha_k} X_n \ dP\right| + \left|\int_{|X|\geq\alpha_k} X \ dP\right| + \epsilon \\
&\leq \left|\sup_{n\in\mathbb{N}} \int_{|X_n|\geq\alpha_k} X_n \ dP\right| + \left|\int_{|X|\geq\alpha_k} X \ dP\right| + \epsilon \to \epsilon, \quad \text{as } k\to\infty.
\end{aligned}$$

It holds for every ϵ as long as n is large enough, hence $E(X_n)\to E(X)$, as $n\to\infty$.

(2) We still have that

$$\int X_n^{\alpha_k}\ dP = E(X_n^{\alpha_k}) \to E(X^{\alpha_k}) = \int X^{\alpha_k}\ dP, \quad \text{as } n \to \infty.$$

Thus leading to

$$\begin{aligned} \left| \int_{X_n \geq \alpha_k} X_n\ dP - \int_{X \geq \alpha_k} X\ dP \right| &= \left| \int X_n - X_n^{\alpha_k}\ dP - \int X - X^{\alpha_k}\ dP \right| \\ &= \left| \int X_n - X\ dP \right| + \left| \int X_n^{\alpha_k} - X^{\alpha_k}\ dP \right|. \end{aligned}$$

Which tends to 0, as $n \to \infty$. It follows that

$$\int_{X_n \geq \alpha_k} X_n\ dP \to \int_{X \geq \alpha_k} X\ dP, \quad \text{as } n \to \infty,$$

for every α_k. For each $\epsilon > 0$ choose α_k so large that $\int_{X \geq \alpha_k} X\ dP < \epsilon$. Then there exists n_0 so that for all $n \geq n_0$,

$$\int_{X_n \geq \alpha_k} X_n\ dP \leq \epsilon.$$

By possibly increasing α_k, the above holds for every n. Since X_n are assumed to be non-negative, we can replace X_n with $|X_n|$. Thus $(X_n)_{n\in\mathbb{N}}$ is uniformly integrable. $\square$

Theorem 3.32. *Let $X, (X_n)_{n\in\mathbb{N}}$ be real-valued random variables not necessarily defined on a common probability space. If $X_n \xrightarrow{\mathcal{D}} X$, as $n \to \infty$, and $(X_n)_{n\in\mathbb{N}}$ is uniformly integrable, then X has first moment and*

$$E_n(X_n) \to E(X), \quad \text{as } n \to \infty.$$

Proof. Construct new random variables defined on a common probability space $Y, (Y_n)_{n\in\mathbb{N}}$, as in the proof of Lemma 3.29. Since $Y_n \xrightarrow{a.s.} Y$, as $n \to \infty$, and $(Y_n)_{n\in\mathbb{N}}$ is uniformly integrable the result follows from Theorem 3.31. $\square$

Corollary 3.33. *Let $X, (X_n)_{n\in\mathbb{N}}$ be real-valued random variables, not necessarily defined on the same probability space. Assume that $X_n \xrightarrow{\mathcal{D}} X$, as $n \to \infty$, and $\sup_{n\in\mathbb{N}} E\,|X_n|^p = M < \infty$ for some $p > 1$. If $r < p$ is a positive integer, then*

$$E_n(X_n^r) \to E(X^r), \quad \text{as } n \to \infty.$$

Proof. $E(X^r), E(X_n^r)$ are welldefined, since $|X_n|^r \leq |X_n|^p + 1$, so $E\,|X_n|^r \leq M + 1$. By Lemma 3.29, $E\,|X|^r \leq \liminf_{n\to\infty} E\,|X|^r \leq M + 1$.

If $|X_n| \geq \alpha$, then $(|X_n|/\alpha)^r \leq (|X_n|/\alpha)^p$. Therefore we get that

$$\int_{|X_n|\geq\alpha} |X_n|^r\ dP_n \leq \alpha^{r-p} \int_{|X_n|\geq\alpha} |X_n|^p\ dP_n \leq \alpha^{r-p} M.$$

Hence

$$\sup_{n\in\mathbb{N}} \int_{|X_n|\geq\alpha} |X_n|^r \; dP_n \leq \alpha^{r-p} M \to 0, \quad \text{as } \alpha \to \infty.$$

So $(X_n^r)_{n\in\mathbb{N}}$ is uniformly integrable for every positive integer $r < p$. By Theorem 3.12, $X_n^r \xrightarrow{\mathcal{D}} X^r$, as $n \to \infty$. So now $\lim_{n\to\infty} E_n(X_n^r) = E(X^r)$ follows from Theorem 3.32. $\square$

With this result available, we can now prove Stirling's Formula:

Theorem 3.34 (Stirling's Formula)**.** *The following holds*

$$\lim_{n\to\infty} \left(\frac{n}{e}\right)^n \frac{\sqrt{2n\pi}}{n!} = 1.$$

Proof. Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of real-valued, independent, identically distributed random variables. Let X_i follow a gamma distribution, with shape and scale parameter 1, so the mean and variance are 1. For each $n \geq 1$, set $S_n = \sum_{i=1}^n X_i$, then S_n is gamma distributed with shape parameter n and scale parameter 1. Then for each $n \geq 1$, set $U_n = (S_n - n)/\sqrt{n}$. According to Theorem 3.21 $U_n \xrightarrow{\mathcal{D}} X$, as $n \to \infty$, where X is standard normal distributed. By Theorem 3.12, $|U_n| \xrightarrow{\mathcal{D}} |X|$, as $n \to \infty$.

Observe that

$$E(U_n^2) = E\left(\left(\frac{S_n - n}{\sqrt{n}}\right)^2\right) = \frac{E(S_n^2) + n^2 - 2nE(S_n)}{n} = \frac{n(n+1) + n^2 - 2n^2}{n} = 1.$$

Hence $\sup_{n\in\mathbb{N}} E(|U_n|^2) = 1 < \infty$. By Corollary 3.33, we conclude that

$$E\,|U_n| \to E\,|X| = \sqrt{\frac{2}{\pi}}, \qquad \text{as } n \to \infty. \tag{3.22}$$

Now we calculate $E\,|U_n|$, using that S_n are gamma distributed,

$$E\,|U_n| = \int_0^\infty \frac{|x-n|}{\sqrt{n}} \frac{x^{n-1}e^{-x}}{\Gamma(n)} \; dx = \frac{\sqrt{n}}{n!} \int_0^\infty |x-n|\, x^{n-1} e^{-x} \; dx.$$

Consider the last integral. By the Monotone convergence Theorem (cf. Theorem 16.2 in [2]),

$$\int_0^\infty |x-n|\, x^{n-1} e^{-x} \; dx = \lim_{k\to\infty, k\geq n} \int_0^k |x-n|\, x^{n-1} e^{-x} \; dx$$

We split up the last integral and use partial integration

$$\begin{aligned}
\int_0^k |x-n|\, x^{n-1} e^{-x}\ dx &= \int_0^n (n-x) x^{n-1} e^{-x}\ dx + \int_n^k (x-n) x^{n-1} e^{-x}\ dx \\
&= n\int_0^n x^{n-1} e^{-x}\ dx - \int_0^n x^n e^{-x}\ dx + \int_n^k x^n e^{-x}\ dx \\
&\quad -n\int_n^k x^{n-1} e^{-x}\ dx \\
&= n\int_0^n x^{n-1} e^{-x}\ dx - \left[-x^n e^{-x}\right]_0^n - n\int_0^n x^{n-1} e^{-x}\ dx \\
&\quad + \left[-x^n e^{-x}\right]_n^k + n\int_n^k x^{n-1} e^{-x}\ dx - n\int_n^k x^{n-1} e^{-x}\ dx \\
&= \left[-x^n e^{-x}\right]_n^k - \left[-x^n e^{-x}\right]_0^n \to 2\left(\frac{n}{e}\right)^n, \quad \text{as } k \to \infty.
\end{aligned}$$

So the above calculations gives us that

$$E\,|U_n| = \frac{2\sqrt{n}}{n!}\left(\frac{n}{e}\right)^n.$$

This combined with (3.22) yields the result. □

With Corollary 3.33 at hand, we are also able to prove the next theorem, which gives a method of deducing weak convergence from convergence of moments, under some additional assumptions.

Theorem 3.35 (The Method of Moments)**.** *Let $X, (X_n)_{n\in\mathbb{N}}$ be real-valued random variables not necessarily defined on a common probability space, but all of them have all moments. Suppose that X is uniquely determined by its moments (see Chapter 2). If for every positive integer r*

$$E_n(X_n^r) \to E(X^r), \quad \textit{as } n \to \infty,$$

then $X_n \xrightarrow{\mathcal{D}} X$, as $n \to \infty$.

Proof. Let $\mu, (\mu_n)_{n\in\mathbb{N}}$ denote the distributions of $X, (X_n)_{n\in\mathbb{N}}$, respectively. Since $E(X_n^2)$ converges, it is bounded. Let K denote the bound. Chebyshev's inequality gives us, for each $n \in \mathbb{N}$, that

$$P_n(|X_n| \geq x) \leq \frac{K}{x^2}.$$

So in terms of μ_n this means that

$$\mu_n(-x, x) \geq 1 - \frac{K}{x^2},$$

for all $n \in \mathbb{N}$, hence $(\mu_n)_{n\in\mathbb{N}}$ is tight. So by Theorem 3.17 there exists a weakly convergent subsequence. Now let $(\mu_{n_k})_{k\in\mathbb{N}}$ denote any weakly convergent subsequence of $(\mu_n)_{n\in\mathbb{N}}$

(such exists as we just saw). Let ν denote the limit, we want to show that $\nu = \mu$. For any even positive integer p we have, by hypothesis

$$\int x^p \ d\mu_{n_k}(x) = E_{n_k}(X_{n_k}^p) \to E(X^p), \quad \text{as } k \to \infty,$$

since $(\mu_{n_k})_{k\in\mathbb{N}}$ is a subsequence of $(\mu_n)_{n\in\mathbb{N}}$. So $\left(\int x^p \ d\mu_{n_k}(x)\right)_{k\in\mathbb{N}}$ is bounded. Since p is even, $\int x^p \ d\mu_{n_k}(x) = \int |x|^p \ d\mu_{n_k}(x)$. Therefore $\left(\int |x|^p \ d\mu_{n_k}(x)\right)_{k\in\mathbb{N}}$ is bounded, hence by Corollary 3.33 we have for all positive integers $r < p$

$$\int x^r \ d\mu_{n_k}(x) \to \int x^r \ d\nu(x), \quad \text{as } k \to \infty.$$

But it also converges to $\int x^r \ d\mu(x)$. So μ and ν have the same moments, since $p \in \mathbb{N}$ was arbitrary. Now, μ is uniquely determined by its moments by hypothesis, so we have that $\mu = \nu$. Now we have seen that every weakly convergent subsequence converges to μ weakly, hence by Corollary 3.18, we conclude that $\mu_n \xrightarrow{w} \mu$, as $n \to \infty$. □

To end this chapter we will discuss an application of the above theorem. More precisely, we apply the Method of moments to prove yet another more general version of the Central Limit Theorem for the case of so-called **triangular arrays** of random variables. Triangular arrays are described as follows: Let $(k_n)_{n\in\mathbb{N}}$ be a sequence of positive integers tending to ∞, as $n \to \infty$. Let for each $n \in \mathbb{N}$, $\{X_{n,1}, X_{n,2}, ..., X_{n,n_k}\}$ denote independent real-valued random variables defined on a common probability space $(\Omega_n, \mathbb{F}_n, P_n)$. The probability spaces are allowed to vary with n. Let $S_n = \sum_{k=1}^{k_n} X_{n,k}$, for $n \geq 1$, denote the row sums.

Theorem 3.36. *Let $(\{X_{n,1}, X_{n,2}, ..., X_{n,n_k}\})_{n\in\mathbb{N}}$ be a triangular array of real-valued random variables as described above. Assume that for each $n \in \mathbb{N}$, there exists $M_n > 0$ such that $|X_{n,k}| \leq M_n$ for each $k = 1, 2, ..., k_n$. Furthermore, assume that $E(X_{n,k}) = 0$ for all $n \in \mathbb{N}$ and $k = 1, 2, ..., n_k$, and let $E(X_{n,k}^2) = \sigma_{n,k}^2$ denote the 2^{nd} moments. Let $s_n = \sum_{k=1}^{k_n} \sigma_{n,k}^2$, $n \geq 1$. If*

$$\frac{M_n}{s_n} \to 0, \quad \text{as } n \to \infty,$$

then $S_n/s_n \xrightarrow{\mathcal{D}} X$, as $n \to \infty$, where X is standard normal distributed.

Proof. Note that the bounds ensure that the moments of all the random variables $X_{n,k}$ exist. Hence the moments of S_n exist. Using the multinomial formula we obtain

$$(S_n)^r = \sum \frac{r!}{r_1!r_2!...r_{n_k}!} X_{n,1}^{r_1} X_{n,2}^{r_2} ... X_{n,n_k}^{r_{n_k}}, \quad r \in \mathbb{N}, \tag{3.23}$$

where the sum $\sum$ extends over all k_n-tuples of non-negative integers, which sum up to r. Since $r < k_n$ eventually for n large enough, some factors will be $X_{n,k}^{r_k} = 1$ (for $r_k = 0$). To avoid this, we consider another way of writing up the above sum. Let u denote the number of r_k's different from 0, then $1 \leq u \leq r$. Summing over these, we get:

$$(S_n)^r = \sum_{u=1}^{r} \sum{}' \frac{r!}{r_1!r_2!...r_u!} \frac{1}{u!} \sum{}'' X_{n,i_1}^{r_1} X_{n,i_2}^{r_2} ... X_{n,i_u}^{r_u}, \quad \text{for n large enough so that } r \leq k_n.$$

Here the sum $\sum'$ extend over all u-tuples $(r_1, r_2, ..., r_u)$ of *positive* integers, which sum up to r, and $\sum''$ extends over u-tuples $(i_1, i_2, ..., i_u)$ of *distinct* integers in the range $1 \leq i_\alpha \leq k_n$. The sum $\sum''$ is introduced to ensure that all the k_n-tuples in (3.23), which are represented by the same u-tuple in the sum $\sum'$, are accounted for. However $\sum'$ and $\sum''$ both accommodate for any permutation of the index $(i_\alpha)_{1\leq i\leq u}$, hence $1/(u!)$ is introduced.

By independence we have that

$$E\left(\left(\frac{S_n}{s_n}\right)^r\right) = \sum_{u=1}^{r} \sum{}' \frac{r!}{r_1!r_2!...r_u!}\frac{1}{u!}A_n(r_1, r_2, ..., r_u), \quad \text{for n large enough so that } r \leq k_n, \tag{3.24}$$

where

$$A_n(r_1, r_2, ..., r_u) = \sum{}'' \frac{1}{s_n^{\sum_{\alpha=1}^{u} r_\alpha}} E(X_{n,i_1}^{r_1})E(X_{n,i_2}^{r_2})...E(X_{n,i_u}^{r_u}).$$

We want to show that

$$\lim_{n\to\infty} A_n(r_1, r_2, ..., r_u) = \begin{cases} 1 & \text{if } r_1 = r_2 = ... = r_u = 2 \\ 0 & \text{else.} \end{cases} \tag{3.25}$$

For the case $r_1 = r_2 = ... = r_u = 2$, observe that we can rewrite the sum as

$$A_n(2, 2, ..., 2) = \sum{}^* \frac{1}{s_n^{2u}}\sigma_{n,i_1}^2\sigma_{n,i_2}^2...\sigma_{n,i_u}^2 - \sum{}^{**} \frac{1}{s_n^{2u}}\sigma_{n,i_1}^2\sigma_{n,i_2}^2...\sigma_{n,i_u}^2, \tag{3.26}$$

where $\sum^*$ extends over all u-tuples $(i_1, i_2, ..., i_u)$ in the range $1 \leq i_\alpha \leq k_n$. And $\sum^{**}$ extends over all u-tuples $(i_1, i_2, ..., i_u)$ in the range $1 \leq i_\alpha \leq k_n$, with at least two index equal. Now observe that

$$s_n^{2u} = \sum_{k=1}^{k_n} \left(\sigma_{n,i}^2 s_n^{2u-1}\right) = \sum_{k=1}^{k_n} \left(\sigma_{n,k}^2 \sum_{j=1}^{k_n} \left(\sigma_{n,j}^2 s_n^{2u-2}\right)\right) = ... = \sum{}^* \sigma_{n,i_1}^2\sigma_{n,i_2}^2...\sigma_{n,i_u}^2.$$

Hence (3.26) is reduced to

$$A_n(2, 2, ..., 2) = 1 - \sum{}^{**} \frac{1}{s_n^{2u}}\sigma_{n,i_1}^2\sigma_{n,i_2}^2...\sigma_{n,i_u}^2.$$

Each term in $\sum^{**}$ has at least one σ_{n,i_α}^2 repeated. For each such term, replace one of these repeated factors with the bound M_n^2. Then

$$0 \leq \sum{}^{**} \frac{1}{s_n^{2u}}\sigma_{n,i_1}^2\sigma_{n,i_2}^2...\sigma_{n,i_u}^2 \leq \left(\frac{M_n}{s_n}\right)^2 \to 0, \quad \text{as } n \to \infty.$$

We here used that the rest of the sum is bounded by the sum over all $(u-1)$-tuples $(i_1, i_2, ..., i_{u-1})$ in the range $1 \leq i_\alpha \leq k_n$, which we just saw was equal to s_n^{2u-2}. So we conclude that (3.25) holds for $r_1 = r_2 = ... = r_u = 2$.

If $r_\alpha = 1$ for at least one α, then $E(X_{n,i_\alpha}^{r_\alpha}) = 0$, so (3.25) holds. Now assume that $r_\alpha \geq 2$ for all α and $r_\alpha > 2$ for at least one α. Then $r = \sum_{\alpha=1}^{u} r_\alpha > 2u$. We now have $E|X_{n,i_\alpha}^{r_\alpha}| \leq M_n^{r_\alpha - 2}\sigma_{n,i_\alpha}^2$, so

$$\begin{aligned} A_n(r_1, r_2, ..., r_u) &\leq \sum{}'' \frac{1}{s_n^r} M_n^{r-2u} \sigma_{n,i_1}^2 \sigma_{n,i_2}^2 ... \sigma_{n,i_u}^2 \\ &= \left(\frac{M_n}{s_n}\right)^{r-2u} \sum{}'' \frac{1}{s_n^{2u}} \sigma_{n,i_1}^2 \sigma_{n,i_2}^2 ... \sigma_{n,i_u}^2 \\ &= \left(\frac{M_n}{s_n}\right)^{r-2u} A_n(2, 2, ..., 2) \to 0, \qquad \text{as } n \to \infty. \end{aligned}$$

Thus we conclude (3.25) holds.

Now we return to (3.24). If r is odd, then all terms in the sum will go to 0, as $n \to \infty$, because of (3.25). This agrees with the fact that $E(X^r) = 0$, when X is standard normal distributed. Similarly if r is even, then all terms will also go to 0, as $n \to \infty$, except for the case $u = r/2$ and $r_1 = r_2 = ... = r_u = 2$. So in this case (3.24) will converge to

$$\frac{r!}{2!2!...2!} \frac{1}{(r/2)!} = 1 \cdot 3 \cdot 5 \cdot ... \cdot (r-1) = E(X^r),$$

where X is standard normal distributed. Since the normal distribution is uniquely determined by its moments (Corollary 2.5), we conclude by the Method of moments (Theorem 3.35), that S_n/s_n converge in distribution to the standard normal distribution, as $n \to \infty$. □

Appendix A

Approximation Results

A.1 Real Polynomials

Theorem A.1 (The Weierstrass Approximation Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous. There exists a sequence of real polynomials $(q_n)_{n \in \mathbb{N}}$, so that*

$$\sup_{x \in [a,b]} |f(x) - q_n(x)| \to 0, \quad \text{as } n \to \infty.$$

The idea of the proof is due to Bernstein (cf. section 2.6 in [7])

Proof. We prove it first for continuous functions defined on $[0, 1]$. Consider the Bernstein polynomials: $p_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k}$, they are defined for $x \in [0, 1]$. Observe that $p_n(0) = f(0)$ and $p_n(1) = f(1)$. So we need to check that for any $\epsilon > 0$, we can choose $n \in \mathbb{N}$, such that $\sup_{x \in (0,1)} |f(x) - p_n(x)| < \epsilon$.

Since f is continous on a compact interval, it is also uniformly continous. Choose δ such that, for any $x, y \in [0, 1]$, where $|x - y| < \delta$ implies that $|f(x) - f(y)| < \frac{\epsilon}{2}$. Now choose $n \in \mathbb{N}$ large enough, such that $\|f\| / (2n\delta^2) < \epsilon$. Here $\|f\| = \sup_{x \in [a,b]} |f(x)|$ is the uniform norm, which is finite since f is continous on the compact interval $[0, 1]$. Now let $x \in (0, 1)$, we see that

$$\begin{aligned} |f(x) - p_n(x)| &= \left| f(x) \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} - \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} \right| \\ &= \sum_{k=0}^{n} \left| f(x) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} x^k (1-x)^{n-k} \\ &= \sum_{\left|\frac{k}{n} - x\right| < \delta}^{n} \left| f(x) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} x^k (1-x)^{n-k} \\ &\quad + \sum_{\left|\frac{k}{n} - x\right| \geq \delta}^{n} \left| f(x) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} x^k (1-x)^{n-k} \end{aligned} \tag{A.1}$$

Where we used that $\sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = (1-x+x)^n = 1$. For the first sum we will use uniform continuity and in the second we will use the general inequality $|f(x) - f(k/n)| \leq$

$2\|f\|$. Now let X be a binomial distributed random variable, with size n and succes parameter x, and let P denote the corresponding probability measure. Then we recognize

$$\sum_{|\frac{k}{n}-x|\geq\delta}^{n} \binom{n}{k} x^k(1-x)^{n-k} = P(|X-E(X)|\geq \delta n) \leq \frac{V(X)}{(\delta n)^2} = \frac{nx(1-x)}{\delta^2 n^2} \leq \frac{1}{4\delta^2 n},$$

by use of Chebyshev's inequality and that $x(1-x) \leq 1/4$. So in (A.1) we get

$$|f(x)-p_n(x)| \leq \frac{\epsilon}{2} + 2\|f\| \sum_{|\frac{k}{n}-x|\geq\delta}^{n} \binom{n}{k} x^k(1-x)^{n-k} \leq \frac{\epsilon}{2} + 2\|f\|\frac{1}{4\delta^2 n} \leq \frac{\epsilon}{2}+\frac{\epsilon}{2} = \epsilon.$$

Thus we have justified that

$$\sup_{x\in[0,1]} |f(x)-p_n(x)| \leq \epsilon.$$

For establishing the result for arbitrary compact intervals $[a,b]$, with $a<b$, we use the affine transformation $t(x)=(x-a)/(b-a)$. It maps $[a,b]$ bijectively to $[0,1]$. So given a continuous function f on $[a,b]$, the function $\tilde{f} = f\circ t$ is continous on $[0,1]$. Let $(\tilde{p}_n)_{n\in\mathbb{N}}$ denote the approximating polynomials, then $(p_n)_{n\in\mathbb{N}} = (\tilde{p}_n \circ t^{-1})_{n\in\mathbb{N}}$ approximates f, since

$$\sup_{x\in[a,b]} |f(x)-p_n(x)| = \sup_{y\in[0,1]} |f\circ t(y) - p_n\circ t(y)| = \sup_{y\in[0,1]} \left|\tilde{f}(y)-\tilde{p}_n(y)\right|.$$

□

The above proof can easily be generalized to the multidimensional case, only the notation is changed. Let $\mathbf{x}=(x_1,x_2,...,x_k)$ and $\mathbf{i}=(i_1,i_2,...,i_k)$. Recall that a polynomial of several variables $x_1,x_2,...,x_k$ is a function $p:\mathbb{R}^k\to\mathbb{R}$, of the form

$$p(\mathbf{x}) = \sum_{\mathbf{i}\in\Delta_N} \alpha_{\mathbf{i}} \prod_{j=1}^{k} x_j^{i_j} \quad \mathbf{x}\in\mathbb{R}^k,$$

for some $N\in\mathbb{N}$, where $\alpha_{\mathbf{i}}\in\mathbb{R}$. The summation extends over the set of all k-tuples, that sums to N, $\Delta_N = \{(i_1,i_2,...,i_k)\in\mathbb{N}_0^k \mid \sum_{j=1}^k i_j = N\}$.

Theorem A.2. *Let $f:\prod_{j=1}^k [a_j,b_j]\to\mathbb{R}$ be continuous, where $(a_j)_{j\in\{1,2,...,k\}}, (b_j)_{j\in\{1,2,...,k\}}$ are real numbers such that $a_j<b_j$ for each j. There exists a sequence of real polynomials $(p_n(\mathbf{x}))_{n\in\mathbb{N}}$ such that*

$$\sup_{x\in\prod_{j=1}^k [a_j,b_j]} |f(\mathbf{x})-p_n(\mathbf{x})| \to 0, \quad \textit{as } n\to\infty.$$

As a consequence, we now obtain the following result which will be very useful in the next section.

Corollary A.3. *Let $f : C \to \mathbb{R}$ be continuous, where C is a compact subset of $\mathbb{R}^k$. There exists a sequence of polynomials $(p_n(\mathbf{x}))_{n\in\mathbb{N}}$, such that*

$$\sup_{\mathbf{x}\in C} |f(\mathbf{x}) - p_n(\mathbf{x})| \to 0, \quad \text{as } n \to \infty.$$

Proof. Since $\mathbb{R}^k$ is a normal space and C is closed, Tietze's extension theorem (cf. Theorem 4.16 in [5]) gives us that f can be extended in to a continuous map $\tilde{f}$ defined on $\mathbb{R}^k$. We therefore have $f(\mathbf{x}) = \tilde{f}(\mathbf{x})$, for $\mathbf{x} \in C$. Since C is bounded, choose $K > 0$ large enough, so that $C \subseteq [-K, K]^k$. Then the previous theorem gives us the sequence of polynomials and

$$\sup_{\mathbf{x}\in C} |f(\mathbf{x}) - p_n(\mathbf{x})| \le \sup_{\mathbf{x}\in[-K,K]^k} |f(\mathbf{x}) - p_n(\mathbf{x})| \to 0, \quad \text{as } n \to \infty.$$

□

A.2 Trigonometric Polynomials

Definition A.4. *A real trigonometric polynomial is a function $p : \mathbb{R} \to \mathbb{R}$ of the form*

$$p(\theta) = c + \sum_{k=1}^{n} a_k \cos(k\theta) + b_k \sin(k\theta), \quad \theta \in \mathbb{R}, \tag{A.2}$$

where $c, a_k, b_k \in \mathbb{R}$ for all $k = 1, 2, ..., n$.

A complex trigonometric polynomial is a function $p : \mathbb{R} \to \mathbb{C}$ of the form

$$p(\theta) = \sum_{k=-n}^{n} c_k e^{ik\theta}, \quad \theta \in \mathbb{R}, \tag{A.3}$$

where $c_k \in \mathbb{C}$ for all $k = -n, ..., -1, 0, 1, ..., n$.

A generalized trigonometric polynomial is a function $p : \mathbb{R} \to \mathbb{C}$ of the form

$$p(\theta) = \sum_{k=1}^{n} a_k e^{i\alpha_k\theta}, \quad \theta \in \mathbb{R}, \tag{A.4}$$

where $a_k \in \mathbb{C}$ and $\alpha_k \in \mathbb{R}$ for all $k = 1, 2, ..., n$. We call it a real generalized trigonometric polynomial, if the imaginary part is 0.

We see that the product of two complex trigonometric polynomials is again a complex trigonometric polynomials, using the identity $e^{ik\theta}e^{im\theta} = e^{i(k+m)\theta}$. Using translations between the complex exponential function and the trigonometric functions, reveals that the product of two real trigonometric polynomials are again a real trigonometric polynomial. We also see that every complex trigonometric polynomial, p, can be written on the form $p = p_1 + ip_2$, for propriate real trigonometric polynomials p_1, p_2.

Theorem A.5. *Let $f : [-\pi, \pi] \to \mathbb{R}$ be a function where $f(-\pi) = f(\pi)$. If f is continuous, then there exists a sequence of real trigonometric polynomials $(p_n)_{n\in\mathbb{N}}$, such that*

$$\sup_{x\in[-\pi,\pi]} |f(x) - p_n(x)| \to 0, \quad \text{as } n \to \infty.$$

Proof. The periodicity of f secures that we can find a function $F : S^1 \to \mathbb{R}$, where S^1 denotes the unit circle, such that

$$F(\cos(\theta), \sin(\theta)) = f(\theta), \qquad \forall \theta \in [-\pi, \pi].$$

F is continuous, since f is so. So by Corollary A.3, there exists a sequence of polynomials of two variables $(q_n)_{n \in \mathbb{N}}$, such that

$$\sup_{(x,y) \in S^1} |F(x,y) - q_n(x,y)| \to 0, \quad \text{as } n \to \infty.$$

Where $q_n(x,y) = \sum_{k,l=0}^{n} a_{k,l} x^k y^l$ for some $a_{k,l} \in \mathbb{R}$. Then

$$p_n(\theta) = q_n(\cos(\theta), \sin(\theta)) = \sum_{k,l=0}^{n} a_{k,l} \cos^k(\theta) \sin^l(\theta),$$

so p_n is a linear combination of products of real trigonometric polynomials, which is again a real trigonometric polynomial. Using that the map, $\theta \mapsto (\cos(\theta), \sin(\theta))$, is a bijection of $[-\pi, \pi]$ to S^1, gives us that

$$\sup_{\theta \in [-\pi,\pi]} |f(\theta) - p_n(\theta)| = \sup_{(x,y) \in S^1} |F(x,y) - q_n(x,y)| \to 0, \quad \text{as } n \to \infty.$$

$\square$

Corollary A.6. *Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and bounded. For every $K > 0$ there exists a sequence of real generalized trigonometric polynomials $(p_n)_{n \in \mathbb{N}}$ such that*

$$\sup_{\theta \in [-K,K]} |f(\theta) - p_n(\theta)| \to 0, \quad \text{as } n \to \infty,$$

and for all $n \in \mathbb{N}$

$$\|p_n\| \leq \|f\| + 1.$$

Proof. Let $g : \mathbb{R} \to \mathbb{R}$ be a function given by

$$g(x) = \begin{cases} 1 & \text{if } x \in [-K, K] \\ (x + K + 1) & \text{if } x \in (-K-1, -K) \\ (K + 1 - x) & \text{if } x \in (K, K+1) \\ 0 & \text{else.} \end{cases}$$

and define $\tilde{f} : \mathbb{R} \to \mathbb{R}$ by

$$\tilde{f}(\theta) = g(\theta) f(\theta), \qquad \theta \in \mathbb{R}.$$

Then $\tilde{f}$ is continuous and equal to 0 on $(-K-1, K+1)^c$ (so especially $\tilde{f}(-K-1) = \tilde{f}(K+1)$). Also $\tilde{f}(\theta) = f(\theta)$ for $\theta \in [-K, K]$ and $\left\|\tilde{f}\right\| \leq \|f\|$. By the previous theorem, we can find a sequence of real trigonometric polynomials $(q_n)_{n \in \mathbb{N}}$, such that

$$\sup_{\theta \in [-\pi,\pi]} \left| \tilde{f}\left(\frac{(K+1)\theta}{\pi}\right) - q_n(\theta) \right| \to 0, \quad \text{as } n \to \infty.$$

With the possibility of removing the first finitely many n's, we can assume that

$$\|q_n\| \leq \left\|\tilde{f}\right\| + \left\|\tilde{f} - q_n\right\| \leq \left\|\tilde{f}\right\| + 1 \leq \|f\| + 1.$$

Where periodicity of q_n ensures that we only need to consider the interval $[-\pi, \pi]$. Then set

$$p_n(\theta) = q_n\left(\frac{\pi\theta}{K+1}\right), \qquad \theta \in \mathbb{R}.$$

These are real generalized trigonometric polynomials. We have that

$$\begin{aligned}
\sup_{\theta\in[-K,K]} |f(\theta) - p_n(\theta)| &= \sup_{\theta\in[-K,K]} \left|\tilde{f}(\theta) - p_n(\theta)\right| \leq \sup_{\theta\in[-K-1,K+1]} \left|\tilde{f}(\theta) - p_n(\theta)\right| \\
&= \sup_{\theta\in[-\pi,\pi]} \left|\tilde{f}\left(\frac{(K+1)\theta}{\pi}\right) - q_n(\theta)\right| \to 0, \quad \text{as } n \to \infty.
\end{aligned}$$

□

A.3 Simple Functions

Theorem A.7. *Let $f : \mathbb{R} \to \mathbb{R}$ be continuous, let $a, b \in \mathbb{R}$ where $a < b$ and let $\epsilon > 0$. There exists a simple function of the form $s(x) = \sum_{i=1}^{n} 1_{(a_i,b_i]}c_i$, such that*

$$\sup_{x\in(a,b]} |f(x) - s(x)| < \epsilon.$$

Proof. f is uniformly continuous on $[a, b]$. So given ϵ choose $\delta > 0$ such that

$$\forall x, y \in [a, b] : \quad |x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon.$$

Divide $[a, b]$ into δ-pieces, i.e., find $a_1, a_2, ..., a_n$, so that $a = a_1 < a_2 < ... < a_n = b$ and

$$|a_i - a_{i+1}| < \delta \qquad \text{for } i = 1, 2, ..., n - 1.$$

Then $s(x) = \sum_{i=1}^{n-1} 1_{(a_i,a_{i+1}]}f(a_i)$ has the wanted properties. □

Appendix B

The Carathéodory Theorem

We will briefly introduce a very powerful mathematical construction in order to prove that certain real-valued functions uniquely determines a measure on $(\mathbb{R}, \mathbb{B})$. The idea is to prove existence of measures from certain set-functions that only have some of the properties of a measure. Let $\mathbb{P}(X)$ denote the power set of a set X.

Definition B.1. *An outer measure μ^* on a set X is a function $\mu^* : \mathbb{P}(X) \to [0, \infty]$ such that*

1. $\mu^*(\emptyset) = 0$.
2. $\mu^*(A) \leq \mu^*(B)$, *if* $A \subseteq B$.
3. $\mu^* \left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$ *for any sequence* $(A_n)_{n \in \mathbb{N}}$ *of subsets of* X

We then define another concept:

Definition B.2. *Let μ^* be an outer measure on X. A subset E of X is called μ^*-measurable if*

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c)$$

holds for all $A \in \mathbb{P}(X)$.

The definition of outer measures automatically gives us one inequality

$$\mu^*(A) = \mu^*((A \cap E) \cup (A \cap E^c)) \leq \mu^*(A \cap E) + \mu^*(A \cap E^c).$$

Lemma B.3. *Let $\mathcal{A}$ be an algebra on X. If $\mathcal{A}$ is stable under countable disjoint union, then $\mathcal{A}$ is a σ-algebra.*

Proof. We only need to check that for given any sequence $(A_n)_{n \in \mathbb{N}}$ of $\mathcal{A}$-sets, the union will still be in $\mathcal{A}$. Set $B_1 = A_1$ and inductively define $B_n = A_n \setminus \bigcup_{j=1}^{n-1} B_j$, for $n \geq 2$. Then $(B_n)_{n \in \mathbb{N}}$ are disjoint and therefore $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n \in \mathcal{A}$. □

The proof of the Carathéodory theorem follows [5], section 1.4.

Theorem B.4 (The Carathéodory Theorem). *Let μ^* be an outer measure on X. Let $\mathbb{E}$ denote the set of μ^*-measurable subsets of X. Let μ be the restriction of μ^* to $\mathbb{E}$. Then $(X, \mathbb{E}, \mu)$ is a measure space.*

Proof. First we show that $\mathbb{E}$ is an algebra. By definition $\emptyset \in \mathbb{E}$ and $\mathbb{E}$ is stable under complement. Let $E, F \in \mathbb{E}$ and let $A \subseteq X$ be arbitrary, then

$$\begin{aligned}
\mu^*(A) &= \mu^*(A \cap E) + \mu^*(A \cap E^c) \\
&= \mu^*(A \cap E \cap F) + \mu^*(A \cap E \cap F^c) + \mu^*(A \cap E^c \cap F) + \mu^*(A \cap E^c \cap F^c) \\
&\geq \mu^*((A \cap E \cap F) \cup (A \cap E \cap F^c) \cup (A \cap E^c \cap F)) + \mu^*(A \cap (E \cup F)^c) \\
&= \mu^*(A \cap (E \cup F)) + \mu^*(A \cap (E \cup F)^c),
\end{aligned}$$

so $E \cup F \in \mathbb{E}$, so $\mathbb{E}$ is an algebra. Now we show that $\mathbb{E}$ is stable under disjoint countable union. Let $(E_n)_{n \in \mathbb{N}}$ be a disjoint sequence of $\mathbb{E}$-sets. Set $F_n = \bigcup_{j=1}^n E_j \in \mathbb{E}$ for each $n \in \mathbb{N}$, and set $F = \bigcup_{n=1}^\infty E_n = \bigcup_{n=1}^\infty F_n$. We want to show that $F \in \mathbb{E}$. Let $A \subseteq X$ be arbitrary. We first show that $\mu^*(A \cap F_n) = \sum_{j=1}^n \mu^*(A \cap E_j)$, by induction on n. It clearly holds for $n = 1$. Let it be true for $n-1$, $n \geq 2$. Then we have that

$$\mu^*(A \cap F_n) = \mu^*(A \cap F_n \cap E_n) + \mu^*(A \cap F_n \cap E_n^c) = \mu^*(A \cap E_n) + \mu^*(A \cap F_{n-1}) = \sum_{j=1}^n \mu^*(A \cap E_j).$$

So now we have that

$$\begin{aligned}
\mu^*(A) &= \mu^*(A \cap F_n) + \mu^*(A \cap F_n^c) = \sum_{j=1}^n \mu^*(A \cap E_j) + \mu^*(A \cap F_n^c) \\
&\geq \sum_{j=1}^n \mu^*(A \cap E_j) + \mu^*(A \cap F^c).
\end{aligned}$$

Hence it also holds in the limit. That combined with the fact that $A = (A \cap F) \cup (A \cap F^c) = \left(\bigcup_{j=1}^\infty A \cap E_j \right) \cup (A \cap F^c)$, gives us that

$$\mu^*(A) \geq \sum_{j=1}^\infty \mu^*(A \cap E_j) + \mu^*(A \cap F^c) \geq \mu^*(A \cap F) + \mu^*(A \cap F^c) \geq \mu^*(A). \tag{B.1}$$

So $F \in \mathbb{E}$, hence $\mathbb{E}$ is stable under countable disjoint union, hence by Lemma B.3, $\mathbb{E}$ is a σ-algebra. By setting $A = \bigcup_{n=1}^\infty E_n$ in (B.1) also gives us that μ^* is a measure when restricted to $\mathbb{E}$. □

Theorem B.5. *Let $F : \mathbb{R} \to \mathbb{R}$ be a function. Assume that F is non-decreasing and right-continuous. Then there exists a unique measure μ on $(\mathbb{R}, \mathbb{B})$, with the property that for all $a, b \in \mathbb{R}$, where $a < b$*

$$\mu(a, b] = F(b) - F(a).$$

Proof. (*Uniqueness*) Let λ be another measure on $(\mathbb{R}, \mathbb{B})$ with the above property. The measure(s) having this property must be σ-finite, i.e., there exists a sequence of $\mathbb{B}$-sets $(K_n)_{n\in\mathbb{N}}$ such that $\lambda(K_n) < \infty$ and $\bigcup_{n=1}^{\infty} K_n = \mathbb{R}$. It follows by using the sequence $((-n,n])_{n\in\mathbb{N}}$. We see that μ and λ must agree on the collection $\{(a,b] \mid a,b \in \mathbb{R},\ a \le b\}$, which is an intersection stable generator of $\mathbb{B}$. By the uniqueness theorem of σ-finite measures (cf. Theorem 10.3 in [2]), we conclude that $\mu = \lambda$.

(*Existence*) Let $L = \lim_{x\to-\infty} F(x)$ denote the limit to the left, it may be $-\infty$. The limit exists since F is non-decreasing. Let $\mathbb{I}$ denote the collection

$$\mathbb{I} = \{(-\infty, a] \mid a \in \mathbb{R}\} \cup \{(a,b] \mid a,b \in \mathbb{R}\ a < b\} \cup \{\emptyset\}\,.$$

Define $\nu : \mathbb{I} \to [0,\infty]$ by $\nu(\emptyset) = 0$, $\nu(-\infty, a] = F(a) - L$ and $\nu(a,b] = F(b) - F(a)$. Now let

$$\mu^*(A) = \inf\left\{\sum_{j=1}^{\infty} \nu(I_j) \mid I_j \in \mathbb{I},\ A \subseteq \bigcup_{j=1}^{\infty} I_j\right\}, \tag{B.2}$$

which is well-defined for all subsets $A \subseteq \mathbb{R}$. By Carathéodory's Theorem the proof is complete, if we show the following:

1. μ^* is an outer-measure.
2. $\mathbb{B} \subseteq \mathbb{E}$, where $\mathbb{E}$ denotes the set of μ^*-measurable subsets of $\mathbb{R}$.
3. $\nu(a,b] = \mu^*(a,b]$ for all $a,b \in \mathbb{R}$, $a < b$.

1. $\mu^*(\emptyset) = 0$ and monotonicity of μ^* follows from the definition of μ^*. We only need to check, for every sequence $(A_n)_{n\in\mathbb{N}}$ of subsets of X, that

$$\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty} \mu^*(A_n). \tag{B.3}$$

First (B.3) holds whenever $\mu^*(A_n) = \infty$ for at least one n, so now we assume they are all finite. Let $\epsilon > 0$ be given. For each A_n choose a sequence $(I_{n,j})_{j\in\mathbb{N}}$ in $\mathbb{I}$, so that $A_n \subseteq \bigcup_{j=1}^{\infty} I_{n,j}$ and $\mu^*(A_n) + \epsilon 2^{-n} \ge \sum_{j=1}^{\infty} \nu(I_{n,j})$. Then $\bigcup_{n=1}^{\infty} A_n \subseteq \bigcup_{n,j=1}^{\infty} I_{n,j}$, so

$$\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n,j=1}^{\infty} I_{n,j} = \sum_{n=1}^{\infty} \mu^*(A_n) + \epsilon 2^{-n} = \epsilon + \sum_{n=1}^{\infty} \mu^*(A_n).$$

Since ϵ was arbitrary, (B.3) holds.

2. Since $\mathbb{I}$ generates $\mathbb{B}$, and $\mathbb{E}$ is a σ-algebra by Carathéodory's Theorem, it is enough to show that $\mathbb{I} \subseteq \mathbb{E}$. Let $I \in \mathbb{I}$, according to the definition of an outer measure and the remark following it, we only need to show that

$$\mu^*(A) \ge \mu^*(A \cap I) + \mu^*(A \cap I^c),$$

for all subsets $A \subseteq \mathbb{R}$. Let $A \subseteq \mathbb{R}$ be arbitrary and $(I_n)_{n\in\mathbb{N}}$ be any sequence of $\mathbb{I}$-sets, such that $A \subseteq \bigcup_{n=1}^{\infty} I_n$. Observe that the complement of any $\mathbb{I}$-set is a disjoint union of two other $\mathbb{I}$-sets. So $I = I' \cup I''$, where $I', I'' \in \mathbb{I}$ are disjoint. Now set for each $n \in \mathbb{N}$

$$J_n := I \cap I_n\ , \qquad J_n' := I' \cap I_n\ , \qquad J_n'' := I'' \cap I_n.$$

For each $n \in \mathbb{N}$, J_n, J_n', J_n'' are $\mathbb{I}$-sets and they form a disjoint union of I_n. Furthermore $\nu(J_n) + \nu(J_n') + \nu(J_n'') = \nu(I_n)$. This combined with the fact that

$$A \cap I \subseteq \bigcup_{n=1}^{\infty} J_n \ , \qquad A \cap I' \subseteq \bigcup_{n=1}^{\infty} J_n' \ , \qquad A \cap I'' \subseteq \bigcup_{n=1}^{\infty} J_n'',$$

and that μ^* was an outer measure gives us

$$\begin{aligned}
\sum_{n=1}^{\infty} \nu(I_n) &= \sum_{n=1}^{\infty} \nu(J_n) + \sum_{n=1}^{\infty} \nu(J_n') + \sum_{n=1}^{\infty} \nu(J_n'') \\
&\geq \mu^*(A \cap I) + \mu^*(A \cap I') + \mu^*(A \cap I'') \\
&\geq \mu^*(A \cap I) + \mu^*(A \cap (I' \cup I'')) = \mu^*(A \cap I) + \mu^*(A \cap I^c).
\end{aligned}$$

Since $(I_n)_{n \in \mathbb{N}}$ was arbitrary we must have that

$$\mu^*(A) \geq \mu^*(A \cap I) + \mu^*(A \cap I^c).$$

3. Let $I = (a, b]$ be given. $\mu^*(I) \leq \nu(I)$ by definition. In order to show the other inequality we need to establish the following:

$$(a, b] \subseteq \bigcup_{n=1}^{N} (a_n, b_n] \Rightarrow \nu(a, b] \leq \sum_{n=1}^{N} \nu(a_n, b_n]. \tag{B.4}$$

We prove it by complete induction by the number N. The induction start holds because F is non-decreasing. So now let $N \geq 2$ and assume that (B.4) holds for all covers of strictly less than N intervals. Find $n_0 \in \{1, 2, .., N\}$, such that $b \in (a_{n_0}, b_{n_0}]$. If $a_{n_0} \leq a$, then we are done. So now assume $a < a_{n_0}$, then

$$(a, a_{n_0}] = (a, b] \setminus (a_{n_0}, b_{n_0}] \subseteq \bigcup_{n \neq n_0}^{N} (a_n, b_n].$$

Then the induction assumption gives us that $\nu(a, a_{n_0}] \leq \sum_{n \neq n_0}^{N} \nu(a_n, b_n]$. So

$$\begin{aligned}
\nu(a, b] &= F(b) - F(a) \leq F(b_{n_0}) - F(a_{n_0}) + F(a_{n_0}) - F(a) \\
&\leq \nu(a_{n_0}, b_{n_0}] + \nu(a, a_{n_0}] = \nu(a_{n_0}, b_{n_0}] + \sum_{n \neq n_0}^{N} \nu(a_n, b_n] = \sum_{n=1}^{N} \nu(a_n, b_n].
\end{aligned}$$

Thus proving (B.4). Now choose any sequence $(I_n)_{n \in \mathbb{N}}$ of $\mathbb{I}$-sets, such that $I \subseteq \bigcup_{n=1}^{\infty} I_n$. Then $\mu^*(I) \geq \nu(I)$ will follow if

$$\nu(I) \leq \sum_{n=1}^{\infty} \nu(I_n). \tag{B.5}$$

It is enough to show it, when I_n's are of the form $(a_n, b_n]$. If $I_n = \emptyset$ we can disregard it, if $I_n = (-\infty, a]$ then replace it with $(c, a]$ for a propriate c. We will still have that

$I \subseteq \bigcup_{n=1}^{\infty} I_n$ and monotonicity of ν (since F is non-decreasing) will still ensure (B.5). Since F is right-continuous we have that

$$\nu(I) = \sup\left\{F(b') - F(a') \mid a', b' \in I,\ a' < b'\right\}. \tag{B.6}$$

Let $\epsilon > 0$ and $a', b' \in I$ be arbitrary, such that $a' < b'$. For each n choose $b_n < b'_n$, such that $F(b'_n) \leq F(b_n) + \epsilon 2^{-n}$ according to right-continuity in b_n. Then $I_n = (a_n, b_n] \subseteq (a_n, b'_n)$ and then $[a', b'] \subseteq I \subseteq \bigcup_{n=1}^{\infty} (a_n, b'_n)$. In accordance with compactness of $[a', b']$, choose a finite subcover such that

$$(a', b'] \subseteq [a', b'] \subseteq \bigcup_{n=1}^{N} (a_n, b'_n) \subseteq \bigcup_{n=1}^{N} (a_n, b'_n].$$

By (B.4) we obtain

$$\begin{aligned} F(b') - F(a') &= \nu(a', b'] \leq \sum_{n=1}^{N} \nu(a'_n, b'_n] = \sum_{n=1}^{N} F(b'_n) - F(a_n) \\ &\leq \sum_{n=1}^{\infty} F(b_n) - F(a_n) + \epsilon 2^{-n} = \epsilon + \sum_{n=1}^{\infty} \nu(I_n). \end{aligned}$$

Since ϵ was arbitrary, $\nu(a', b'] \leq \sum_{n=1}^{\infty} \nu(I_n)$. Now since a', b' were arbitrary (B.6) gives us that $\nu(a, b] \leq \sum_{n=1}^{\infty} \nu(I_n)$. Since $(I_n)_{n \in \mathbb{N}}$ was an arbitrary cover of I, we conclude that $\mu^*(I) \geq \nu(I)$. Hence we have proved the existence of a measure with the wanted properties. $\square$

Setting $F(x) = x$ proves the existence of a measure on $(\mathbb{R}, \mathbb{B})$ which, for every interval of the form $(a, b]$, assigns a value equal to the length of $(a, b]$. As a further remark, if F also has the property that $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$, then the unique measure corresponding to F is a probability measure.

Bibliography

[1] C. Berg, J. P. R. Christensen, P. Ressel: *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions*, Springer Textbook, Springer-Verlag New York Inc., 1984.

[2] P. Billingsley: *Probability and Measure*, Third Edition, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1995.

[3] K. L. Chung: *A Course in Probability Theory*, Second Edition, Probability and Mathematical Statistics, A Series of Monographs and Textbooks, Academic Press, Inc., San Diego, 1974.

[4] R. Durrett: *Probability: Theory and Examples*, Second Edition, Duxbury Press, Wadsworth Publishing Company, Belmont, 1996.

[5] G. B. Folland: *Real Analysis: Modern Techniques and Their Applications*, Second Edition, Pure and Applied Mathematics: A Wiley-Interscience Series of Texts, Monographs and Tracts, John Wiley & Sons, Inc., New York, 1999.

[6] A. Gut: *Probability: A Graduate Course*, Springer Texts in Statistics, Springer, New York, 2005.

[7] Y. G. Sinai: *Probability Theory: An Introductory Course*, Springer Textbook, Springer-Verlag, Berlin, 1992.

[8] J. Stoyanov: *Krein condition in probabilistic moment problems*, Bernoulli Vol. 6, No. 5 (Oct., 2000), pp. 939-949, International Statistical Institute (ISI) and Bernoulli Society for Mathematical Statistics and Probability.